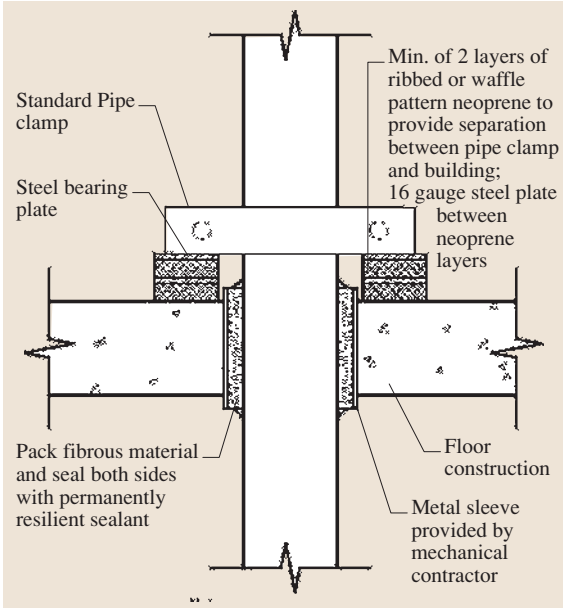


Fig. 11.62 General considerations for sealing pipe penetrations



Fig. 11.63 Examples of sealed pipe and duct penetrations

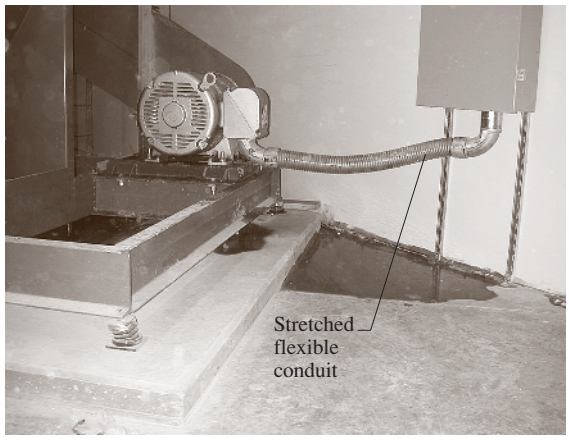


Fig. 11.64 An example of stretched flexible conduit



Fig. 11.65 An example of separating an electrical box from mechanical equipment to isolate vibrations

Figure 11.66 shows an enclosure for a rooftop HVAC unit and Fig. 11.67 shows an enclosure for a rooftop chiller unit.

Note in Fig. 11.66 the relief hoods that provide ventilation while directing the sound away from the community. When designing these types of enclosures, it would be prudent to face any ventilation openings away from noise-sensitive locations, including skylights on the roof of the building supporting the unit. Figure 11.68 shows a chiller enclosure at ground level next to a building near a residential community.

Amazon
EX1013 Page 432

11.6 Acoustical Privacy in Buildings

Privacy issues can be a concern for spaces in which speech is a predominant source. A partial list of these areas includes offices, dwelling units, classrooms, health care facilities, libraries, hotels, and motels. The office environment provides a good example of privacy concerns and how they are handled. These solutions can be used for sound privacy in most situations.

11.6.1 Office Acoustics Concerns

The most common concerns related to office acoustics are privacy and distractions. Distractions can affect productivity. The concern for privacy has recently received much broader attention due to the passage of numerous privacy-protection laws by the US federal government and the European Parliament. Current legislation in the USA requires that the transfer of health records, in whatever format, conforms with requirements for privacy (through the Health Insurance Portability and Accounting Act of 1996). This includes privacy for digital transfer of health records between health care providers and presumably aural privacy as well. Similar concerns are also entering the realm of potential legislation for the financial world (through the Gramm Leach Bliley Financial Services Modernization Act of 1999, relating to privacy of financial transactions).



Fig. 11.66 An example of a rooftop HVAC unit enclosure



Fig. 11.67 An example of a rooftop chiller enclosure

11.6.2 Metrics for Speech Privacy

Many metrics are available for rating office environments, particularly those related to speech privacy. Listed below are the most popular.

Articulation Index

Speech intelligibility and acoustics in open-plan offices have been rated in terms of the articulation index (AI) until recently (through ANSI S3.5), but AI is still being used by many consultants. AI is a measure of the ratio between a voice level and steady background noise (background noise from mechanical equipment, traffic, or electronic sound masking). AI was originally developed to evaluate communication systems, and has been widely used to assess conditions for speech intelligibility in rooms. AI values range from near 0 (low speech level and relatively high background noise resulting in poor intelligibility and good speech privacy) to 1.0 (high voice level and low background noise resulting in excellent communication and no speech privacy).



Fig. 11.68 An example of a ground-level chiller enclosure

Amazon
EX1013 Page 433

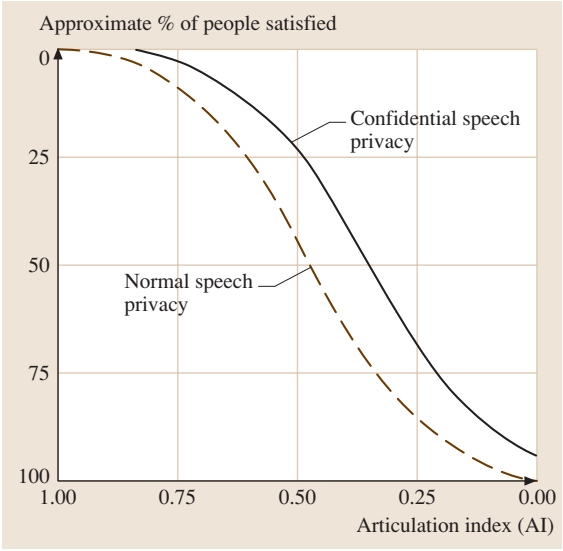


Fig. 11.69 Correlation between articulation index and speech privacy (adapted from [11.1])

When privacy is desired, it is necessary to have a low AI. When communication is desired, it is necessary to have a high AI so people can clearly understand speech.

Figure 11.69 shows how AI relates to speech privacy. This graph provides general guidelines from experience. Speech privacy goals are divided into three categories – minimal distraction, normal speech privacy, and confidential speech privacy. Minimal distraction corresponds to an AI of 0.35 or less. Normal speech privacy, in which the average person can work without distraction although occasional parts of outside conversations can be heard, corresponds to an AI of 0.20 or less. Confidential privacy, where the average person can carry on discussions with assurance that he or she will not be understood by neighbors, corresponds to an AI of 0.05 or less. There is no privacy when AI exceeds 0.40.

Speech Intelligibility Index

AI has been replaced by the speech intelligibility index (SII) in recent versions of ANSI S3.5. SII is still, like AI, a weighted speech-to-noise ratio. However, it is somewhat more complex to calculate than AI and includes revised frequency weightings and the masking effect of one frequency band on nearby frequency bands. Like AI, it has values that range between 0 and 1, but for the same conditions SII values are slightly larger than AI values.

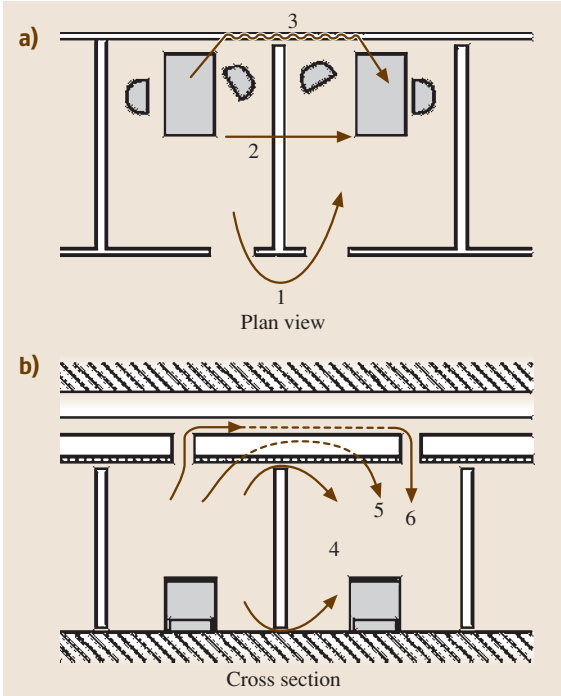


Fig. 11.70a,b The most common paths for sound leakage between closed offices. **(a)** Plan view; **(b)** cross section (1-diffraction between openings, 2-transmission through partitions, 3-structure-borne transmission, 4-transmission through air gaps in wall seals, 5-diffraction through plenum when partition does not seal to structural deck, and 6-transmission through ductwork)

Privacy Index

Low AI values indicate a higher degree of privacy. Since this may be confusing to laypersons who want to focus on better privacy conditions, a metric call privacy index (PI) has been published in ASTM E 1130. The Privacy Index (PI) is defined as:

$$\mathrm{PI} = (1 - \mathrm{AI}) \times 100 \,, \tag{11.9}$$

presented as a percentage. Higher PI values indicate more privacy; lower PI values indicate less privacy. For example, an AI of 0.10 corresponds to a PI of 90%.

Generally accepted practice for the design of open-plan offices refers to two levels of speech privacy. *Confidential privacy* is defined as a condition in which speech may be detected but not understood. *Normal privacy* allows modest amounts of intelligibility, but normal work patterns are usually not interrupted. These terms are standardized in ASTM E 1374.

Amazon
EX1013 Page 434

Table 11.9 AI, SII and PI for open plan offices

AI	SII	PI	Privacy condition	Office environment
> 0.65	> 0.75	< 35%	Good communication	Necessary when communication is desirable (conference rooms, classrooms, auditoriums, etc.)
> 0.40	> 0.45	< 60%	No privacy	Clear intelligibility of conversations and distraction
0.35	0.45	65%	Freedom from distraction	Reasonable work conditions not requiring heavy concentration or speech privacy; can hear and understand neighboring conversations
0.20	0.27	80%	Normal speech privacy	Only occasional intelligibility from a neighbor's conversation; work patterns not interrupted
< 0.05	< 0.10	> 95%	Confidential speech privacy	Aware of neighbor's conversation but it is not intelligible

Table 11.9 compares AI, SII, and PI for the same types of communication and privacy conditions.

Articulation Class

In an open-plan office, the effectiveness of the ceiling in determining noise reduction is partially related to the height of the barrier that separates two workstations, the distance between a talker and listener, and the absorption characteristics. Therefore, another metric has been developed to describe a ceiling's contribution to noise reduction between typical work areas. This is the articulation class (AC), measured in an office mockup environment in accordance with ASTM E 1110 and E 1111. The AC is the sum of the weighted sound attenuation numbers in a series of 15 test bands for a carefully specified office layout, barrier height, and ceiling height. AC values usually exceed 100 and typically range from about 150 to 220, with higher values indicating more privacy and less speech intelligibility. AC only considers the effects of noise reduction while AI, SII, and PI also consider the effects of speech level and spectrum characteristics, and background level and spectrum characteristics. Subjective ratings of AC for levels of privacy (like those listed in Table 11.9) have not been analyzed at this time.

Speech Interference Level

The international community uses the speech interference level (SIL), as defined in ISO 9921-1, to rate speech communication environments. SIL is defined as the arithmetic average of sound pressure levels in the ambient environment at the 500, 1000, 2000, and 4000 Hz octave-band frequencies. SIL can also be approximated by subtracting 8 dB from the overall dBA ambient level. Speech intelligibility is rated by the difference between the SIL and the A-weighted sound pressure level of speech at the listener's location. If this signal-to-noise ratio exceeds 10 dB, that indicates satisfactory speech communication. Table 11.10 shows general characteristics of assessing speech intelligibility against this signal-to-noise ratio.

Speech Transmission Index

While the metrics above are typically used for office environments to rate privacy between spaces, there is another metric that is widely used to describe speech intelligibility within performance and lecture spaces. This metric, known as the speech transmission index (STI), can be measured with commercially available instruments, often in terms of the metric known as the rapid speech transmission index (RASTI) to streamline the monitoring effort. STI measurements consider octave-

Table 11.10 ISO 9921-1 speech intelligibility ratings based on SIL

Signal-to-noise ratio at listener's position (dBA-SIL)	Speech intelligibility rating
< −6	Insufficient
−6 to −3	Unsatisfactory
−3 to 0	Sufficient
0 to 6	Satisfactory
6 to 12	Good
12 to 18	Very good
> 18	Excellent

Amazon
EX1013 Page 435

band levels between 125 and 8000 Hz while RASTI measurements only consider octave-band levels at 500 and 2000 Hz. STI measurement procedures are standardized in the International Electrotechnical Commission (IEC) standard IEC 60268-16. STI, as AI and SII, ranges from 0 to 1 and, unlike AI and SII, STI takes room reverberation into account in its calculations. The meanings of STI ratings are generally similar to those for SII, with higher values being good for speech intelligibility and bad for privacy, and low values representing the opposite. While STI can be measured directly with instruments, AI and SII cannot.

Ceiling Attenuation Class

With current lightweight construction, walls often do not extend to the structure. In these cases, the path for sound through the ceiling plenum is the weakest sound path between offices. This sound path is rated with a ceiling attenuation class (CAC) that is analogous to an STC rating. The CAC value is measured in accordance with ASTM E 1414 and measures the sound transfer from one standard-sized office through an acoustic tile ceiling, then into a standard plenum, and then back into the neighboring office again through an acoustic tile ceiling. Mineral fiber acoustic tiles typically have a rating of CAC 35–39. Fiberglass tiles have significantly lower CAC values. In cases where there is a return air grille in the ceiling so air can be exhausted through the plenum, the field performance will also be significantly degraded.

11.6.3 Fully Enclosed Offices

If confidential privacy is required, employees must be located in closed offices with sealed doors and walls that extend to the structural ceiling. This point needs to be stressed because many closed offices are designed with unsealed doors and walls extending to suspended, nonstructural ceilings, allowing space for HVAC and electrical equipment. Typical suspended ceilings provide acoustic absorption for reflected sound but provide little in the way of transmission loss. This means that sound will travel through suspended ceiling panels and over walls as if the walls are barriers, thus limiting the acoustic privacy of these offices almost to the open-plan condition. Figure 11.70 shows the most common sound leakage paths between closed offices.

Speech privacy between two rooms, in general, is a function of two key parameters – noise reduction and background noise level. One must not only reduce the sound level of the unwanted source, but one must also generate sufficient background noise to mask the unwanted source, making it less intelligible and less distracting. Confidential speech privacy can be achieved in closed offices by using walls having a high transmission loss. If these types of walls are used, the background sound levels can remain low. However, for normal speech privacy in an open-plan office, one must strike a balance between a lower transmission loss (because of the limited effectiveness of barriers) and a higher background noise level (that can often be set using an electronic masking system).

For typical contemporary offices, the range of construction options is often limited to just the number of layers of gypsum board on the wall, the type of stud framing, and whether or not there will be insulation in the cavity of the wall. A single metal stud with a single layer of gypsum board on each side of the stud and no insulation in the cavity will have a performance in the range of STC 40. If the layers of gypsum board are doubled (that is, two layers on each side), then the overall STC value will increase by about 5 STC points. There is negligible difference between the performance of the wall with gypsum board that is 1.2 cm thick or 1.6 cm thick. Insulation in the cavity of a wall with a single stud may provide an additional 3–5 STC points performance. Wood studs instead of metal studs can provide a more rigid path for the transfer of sound (in a single-wall design), and may degrade the performance by 3–5 STC points. Further improvements above STC 50 require the use of resilient channels or staggered stud framing, as discussed in Sect. 11.5.1.

11.6.4 Open-Plan Offices

The open-plan office design has become very popular through the latter half of the 20th century and into the 21st. This design scheme saves money, promotes teamwork, and improves flexibility for future renovations. Many employees, however, view this design as a series of compromises in terms of space, prestige, and (most of all) privacy. As employees consider changing from the traditional closed office to open-plan cubicles, they often have concerns about their abilities to work productively in what they anticipate to be a noisier, more distracting workplace. The overwhelming complaint about open-plan office design is the lack of acoustic privacy.

The first step in the acoustic design of an office is to determine the needs of the employees in each area. Employees in some companies may need to communicate freely as part of their jobs. These workers would not need any acoustic privacy. Some may need visual privacy (using barriers that block their line of sight with others) but

Amazon
EX1013 Page 436

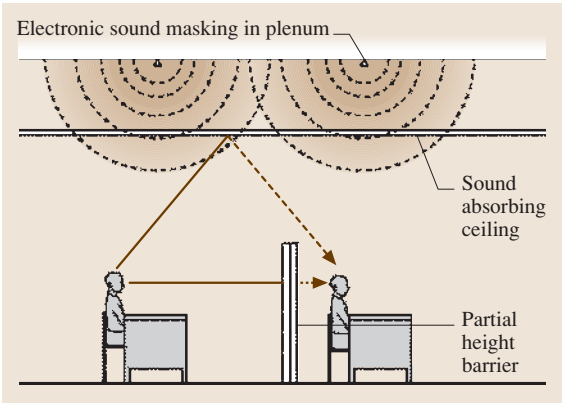


Fig. 11.71 Acoustic considerations for open-plan office design in cross section

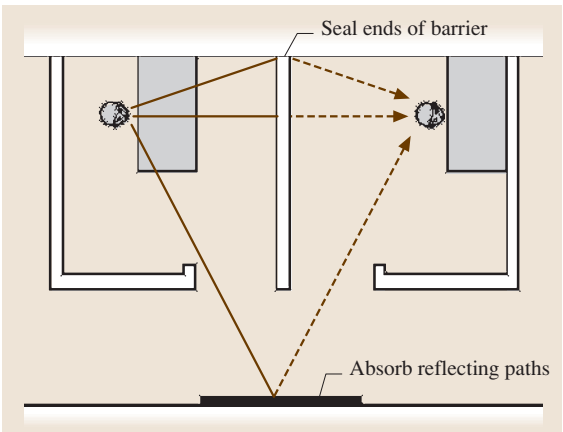


Fig. 11.72 Acoustic considerations for open-plan office design in plan view

need only a minimal amount of speech privacy. Some may require an environment free from distractions for the performance of detailed work. It is often this last category of employees that feels most compromised by being moved into cubicles. This category of employee needs normal speech privacy, in which conversations in adjacent areas can be understood but are not distracting to concentration on tasks. Normal speech privacy is attainable for open-plan offices with the proper acoustic design.

Confidential speech privacy, in which no part of a conversation can be understood from an adjacent space, cannot be expected from open-plan office designs.

Assuming a goal of normal speech privacy, there are three acoustic design principles that must be addressed to achieve that goal in an open-plan office. These three principles are to absorb, block and cover speech in nearby workstations. These three principles are briefly discussed below.

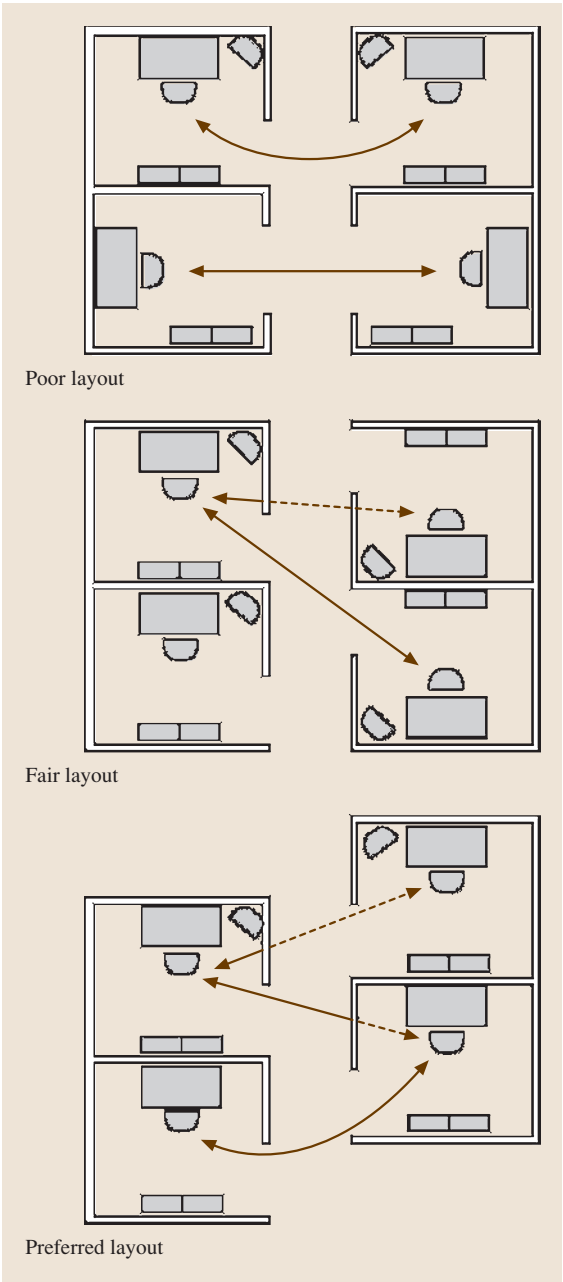


Fig. 11.73 Acoustical considerations for open-plan office layout

Amazon
EX1013 Page 437

Absorption

It is critical to have as much absorption as possible in rooms designed for open-plan offices. This will stop sound from reflecting off room surfaces to travel appreciable distances and from causing remote distractions. If one has to choose surfaces for absorptive treatment, the ceiling is the most critical surface to start with. This is because reflections off ceilings significantly compromise the already limited noise reduction capabilities of barriers. Ceiling finishes should have NRC ratings of at least 0.90 or an AC of at least 180. Ceiling treatments can be part of suspended ceilings or surface-applied. Hard, sound-reflecting lighting fixtures can degrade the performance of the ceiling finish, so it is ideal to use indirect floor lighting to keep the ceiling as absorbing as possible. If ceiling fixtures must be used, parabolic fixtures are more appropriate than others because they diffuse the sound more evenly. Figure 11.71 shows ceiling considerations. Absorptive finishes are also important on the interior of workstation panels and on any walls that may reflect sound from one cubicle to another. Figure 11.72 shows how to place absorptive materials strategically on walls to eliminate reflected paths between cubicles. Carpeting is important for floors to control impact noise (such as footfalls, dropped objects and furniture movement) as well as to absorb reflected sound energy.

Blocking

Sound is blocked in open-plan offices by cubicle barriers. To be most effective, these barriers must be designed to minimize the compromises caused by diffraction. In this regard, barriers should be at least 1.5 m tall and should have STC ratings of at least 25. The amount by which STC ratings exceed 25 will not make a difference in barrier performance (because of diffraction). They should be arranged to block the line of sight between workers (as is shown in Fig. 11.73) and should be flush with floors, window sills, and side walls.

Masking

A typical contemporary office, with sealed windows and properly maintained HVAC systems, often has too low of a background noise level to allow for normal speech privacy. The most practical way to improve privacy under these conditions is to increase the background noise level by adding an electronic sound masking system. These systems can include loudspeakers distributed throughout ceiling plenums, hanging unobtrusively from ceilings, or in each cubicle. These systems, when set up properly, sound like typical HVAC noise to employees. Figure 11.21 shows a typical spectrum for sound masking systems.

11.7 Relevant Standards

The following standards are referenced in this chapter:

US Standards

ANSI (American National Standards Institute)

S3.5	Standard Methods for Calculation of the Speech Intelligibility Index
S12.2	Standard Criteria for Evaluating Room Noise

ASTM (American Society of Testing and Materials)

C 423	Standard Test Method for Sound Absorption and Sound Absorption Coefficients by the Reverberation Room Method
E 90	Standard Test Method for Laboratory Measurement of Airborne Sound Transmission Loss of Building Partitions and Elements
E 413	Classification for Rating Sound Insulation
E 492	Standard Test Method for Laboratory Measurement of Impact Sound Transmission Through Floor-Ceiling Assemblies Using the Tapping Machine
E 989	Standard Classification for Determination of Impact Insulation Class (IIC)
E 1041	Standard Guide for Measurement of Masking Sound in Open Offices
E 1110	Standard Classification for Determination of Articulation Class
E 1111	Standard Test Method for Measuring the Interzone Attenuation of Ceiling Systems
E 1130	Standard Test Method for Objective Measurement of Speech Privacy in Open Offices Using Articulation Index

Amazon
EX1013 Page 438

E 1179 Standard Specification for Sound Sources Used for Testing Open Office Components and Systems

E 1332 Standard Classification for Determination of Outdoor–Indoor Transmission Class

E 1374 Standard Guide for Open Office Acoustics and Applicable ASTM Standards

E 1375 Standard Test Method for Measuring the Interzone Attenuation of Furniture Panels Used as Acoustical Barriers

E 1376 Standard Test Method for Measuring the Interzone Attenuation of Sound Reflected by Wall Finishes and Furniture Panels

E 1414 Standard Test Method for Airborne Sound Attenuation Between Rooms Sharing a Common Ceiling Plenum

International Standards

International Electrotechnical Commission (IEC) and International Organization for Standardization (ISO)

IEC 60268-16 Objective Rating Of Speech Intelligibility By Speech Transmission Index

ISO 9921-1 Speech Interference Level And Communication Distances For Persons With Normal Hearing Capacity In Direct Communication (SIL method)

References

11.1 W.J. Cavanaugh, W.R. Farrell, P.W. Hirtle, B.G. Watters: Speech privacy in buildings, J. Acoust. Soc. Am. **34**(4), 475–492 (1962)

Amazon
EX1013 Page 439

Part D Hearing

Part D Hearing and Signal Processing

12 Physiological Acoustics
Eric D. Young, Baltimore, USA

13 Psychoacoustics
Brian C. J. Moore, Cambridge, UK

14 Acoustic Signal Processing
William M. Hartmann, East Lansing, USA

Amazon
EX1013 Page 440

12. Physiological Acoustics

The analysis of sound in the peripheral auditory system solves three important problems. First, sound energy impinging on the head must be captured and presented to the transduction apparatus in the ear as a suitable mechanical signal; second, this mechanical signal needs to be transduced into a neural representation that can be used by the brain; third, the resulting neural representation needs to be analyzed by central neurons to extract information useful to the animal. This chapter provides an overview of some aspects of the first two of these processes. The description is entirely focused on the mammalian auditory system, primarily on human hearing and on the hearing of a few commonly used laboratory animals (mainly rodents and carnivores). Useful summaries of non-mammalian hearing are available [12.1]. Because of the large size of the literature, review papers are referenced wherever possible.

12.1 **The External and Middle Ear** 429
12.1.1 External Ear 429
12.1.2 Middle Ear 432
12.2 **Cochlea** 434
12.2.1 Anatomy of the Cochlea 434
12.2.2 Basilar-Membrane Vibration and Frequency Analysis in the Cochlea 436
12.2.3 Representation of Sound in the Auditory Nerve 441
12.2.4 Hair Cells 443
12.3 **Auditory Nerve and Central Nervous System** 449
12.3.1 AN Responses to Complex Stimuli 449
12.3.2 Tasks of the Central Auditory System 451
12.4 **Summary** 452
References 453

12.1 The External and Middle Ear

The external and middle ears capture sound energy and couple it efficiently to the cochlea. The problem that must be solved here is that sound in air is not efficiently absorbed by fluid-filled structures such as the inner ear. Most sound energy is reflected from an interface between air and water, and terrestrial animals gain up to 30 dB in auditory sensitivity by virtue of their external and middle ears.

Comprehensive reviews of the external and middle ears are available [12.2, 3]. Figure 12.1 shows an anatomical picture of the ear (Fig. 12.1a) and a schematic showing the important signals in this part of the system (Fig. 12.1b). The external ear collects sound from the ambient field (p_F) and conducts it to the eardrum where sound pressure fluctuations (p_T) are converted to mechanical motion of the middle-ear bones, the malleus (with velocity v_M), the incus, and the stapes; the velocity of the stapes (v_S) is the input signal to the cochlea, producing a pressure p_V in the scala vestibuli. The following sections describe the properties of the transfer function from p_F to p_V.

12.1.1 External Ear

The acoustical properties of the external ear transform the sound pressure (p_F) in the free field to the sound pressure at the eardrum (p_T). Two main aspects of this transformation important for hearing are:

1. efficient capture of sound impinging on the head; and
2. the directional sensitivity of the external ear, which provides a cue for localization of sound sources.

Figures 12.2a and 12.2b show external-ear transfer functions for the human and cat ear, These are the dB ratio of the sound pressure near the eardrum p_T (see the caption) to the sound in free field (p_F). The free-field sound p_F is the sound pressure that would exist at the approximate location of the eardrum if the subject were removed from the sound field completely. Thus the transformation from p_F to p_T contains all the effects of putting the subject in the sound field, including those due to refraction and reflection of sound from the

Amazon
EX1013 Page 441

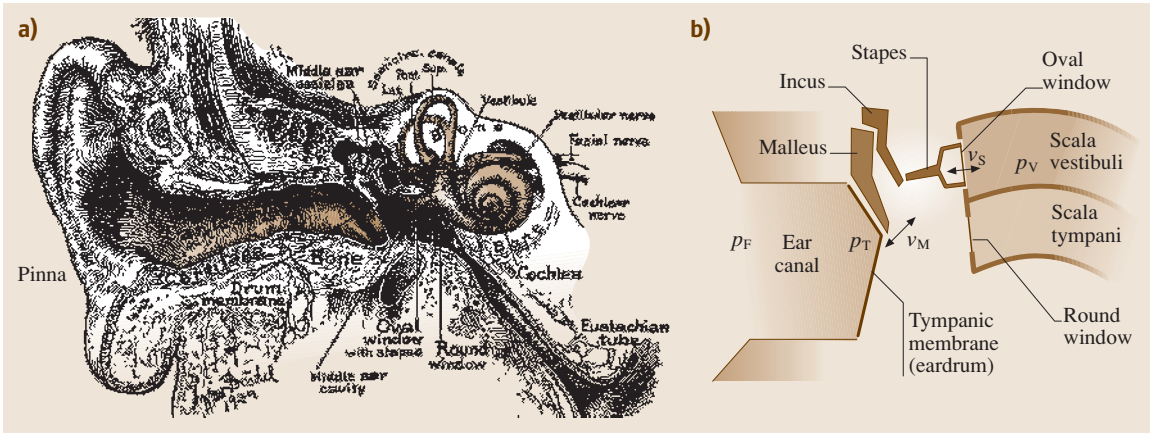


Fig. 12.1 **(a)** Drawing of a cross section of the human ear, showing the pinna, the ear canal, eardrum (drum membrane), middle-ear bones (ossicles), and the labyrinth. The latter contains the vestibular organs of balance and the cochlea, or inner ear. (After [12.8]) **(b)** Schematic of the mammalian external and middle ears showing the signals referred to in the text. The middle-ear bones are the malleus, incus, and stapes. The stapes terminates on the cochlea, on a flexible membrane called the oval window. The velocity of the stapes (v_S) is the input signal to the cochlea; it produces a pressure variation p_V in the fluids of the scala vestibuli, one of the chambers making up the cochlea. The structure of the cochlea and definitions of the scala vestibuli and scala tympani are shown in a later figure. The function of the external and middle ears is to capture the energy in the external sound field p_F and transfer it to stapes motion v_S or equivalently, sound pressure p_V in the scala vestibuli

surfaces of the body, the head, and the ear and the effects of propagation down the ear canal. In each case several transfer functions are shown. The dotted lines show p_T/p_F for sound originating directly in front of the subject [12.4, 5]; these curves are averaged across a number of ears. The solid lines show transfer functions measured in one ear from different locations in space [12.6, 7]. These have a fixed azimuth (i. e., position along the horizon, with 0° being directly in front of the subject) and vary in elevation (again 0° is directly in front) in the median plane. Elevations are given in the legends. External-ear transfer functions differ in detail from animal to animal. In particular, the large fluctuations at high frequencies, above 3 kHz in the human ear and 5 kHz in the cat ear, differ from subject to subject. These fluctuations are not seen in the averaged ears because they are substantially reduced by averaging across subjects, although some aspects of them remain [12.7]. However, the general features illustrated here are typical.

Capture of Sound Energy by the External Ear

Over most of the range of frequencies, the sound pressure is higher at the eardrum than in the free field in both the human (Fig. 12.2a) and the cat (Fig. 12.2b) subjects. The amplification results from resonances in the external ear canal and in the cavities of the pinna; the resonances produce the broad peak between 2 and 5 kHz [12.2]. Although this seems to be an amplification of the sound field, the pressure gain is not by itself sufficient to specify fully the sound-collecting function of the external ear, because it is really the power collection that is important and sound power is the product of pressure and velocity. A useful summary measure of power collection is the effective cross-sectional area of the ear a_{ETM} [12.2, 9]. a_{ETM} measures the ear's ability to collect sound from a diffuse pressure field incident on the head, in the sense that the sound power into the middle ear is equal to a_{ETM} multiplied by the power density in the sound field. If a_{ETM} is larger (smaller) than the anatomical area of the eardrum, for example, it means that the external ear is more (less) efficient at collecting sound than a simple eardrum-sized membrane located on the surface of the head.

Figure 12.2c shows calculations of the magnitude of a_{ETM} for the cat and human ear [12.2]. For comparison, the horizontal line shows the cross-sectional area of the external opening of the cat's pinna. Around 2–3 kHz, a_{ETM} equals the pinna area for the cat. The dotted line marked "ideal" is the maximum possible value of a_{ETM}, based on the power incident on a sphere in a diffuse field [12.2, 9]. At frequencies above 2 kHz, the cat ear is close to this maximum and the human ear is some-

Amazon
EX1013 Page 442

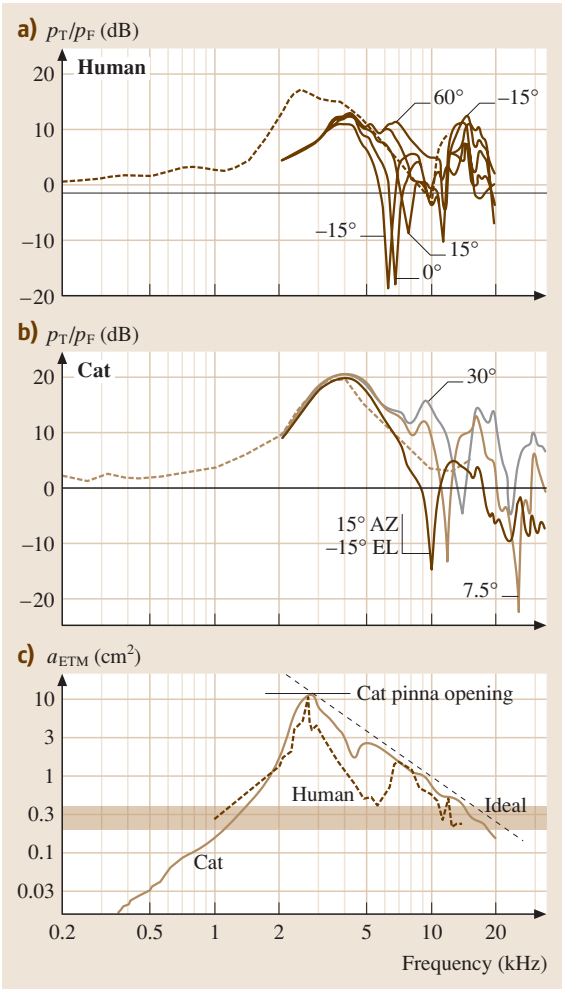


Fig. 12.2a–c Transformation of sound pressure by the external ear, considered as the ratio p_T/p_F for the human ear **(a)** and the cat ear **(b)**. p_T is the pressure measured by a probe microphone in the ear canal, within 2 mm of the eardrum for all cases except the *solid curves* in **(a)**, which show the pressure at the external end of the ear canal, with the ear canal blocked. The *dashed curves* are averaged across many ears, with the sound originating straight ahead, and the *solid curves* are data from one ear at various sound-source directions. The *numbers* next to the curves or in the legend identify the elevation of the sound source relative to the head. In both cases, the azimuth of the source is fixed. **(c)** Capture of sound energy by the external ear, considered as the effective cross-sectional area of the ear, mapped to the eardrum and plotted versus frequency. Curves for the cat and human ear are shown. The *shaded bar* shows the range of cross-sectional areas of the ear canal, for comparison. (Fig. 12.2a after [12.4, 7]; Fig. 12.2b after [12.5, 6]; Fig. 12.2c after [12.2])

what lower. At these frequencies, the external ear is collecting the sound power in a diffuse field nearly as efficiently as is possible. At lower frequencies, the efficiency drops rapidly; *Shaw* [12.9] attributes the drop off to three effects:

1. loss of pressure gain below the canal resonance
2. increase in the wavelength of sound compared to the size of the ear
3. increase in the impedance of the middle ear at low frequencies, decreasing the ability of the ear to absorb sound.

Figure 12.2c shows that the external ear can be considered as an acoustical horn that captures sound and couples it to the eardrum. A feeling for the effectiveness of the external ear's sound collection can be gotten by comparing the values of a_{ETM} to the cross-sectional area of the ear canal, which is 0.2–0.4 cm^2 in cat, shown by the shaded bar [12.10].

Directional Sensitivity

Directional sensitivity means that the transfer function of the external ear depends on the direction to the sound source. As a result, the spectra of broadband sounds (like noise) will be different, depending on the direction in space from which they originate. Listeners, both humans and non-human mammals, are sensitive to these differences and are able to extract sound localization information from them. For example, human observers use these so-called *spectral cues* to localize sounds in elevation and to disambiguate binaural difference cues, helping to distinguish sounds in front and behind the listener [12.11]. Cats depend on spectral cues to localize sounds in the frontal plane, both in azimuth and elevation [12.12].

The directional sensitivity of the external ear is generated by sound reflecting off the structures of the pinna. In humans the pinna is the cartilaginous structure attached to the external ear canal on the side of the head, and includes the cavities that lead to the entrance of the ear canal [12.13, 14]. In the cat, the pinna is a collection of similar cartilaginous shapes and cavities leading to the ear canal, most of which are located on the surface of the head at the ear canal opening. The external auricle, which forms the movable part of the cat's ear, serves as an additional sound collector and a way for the cat to change the directionality of its ear under vol-

Amazon
EX1013 Page 443

untary control [12.15]. The nature of spectral cues in human and cat are shown in Figs. 12.2a,b. The most prominent spectral cue is the notch at mid frequencies (3–10 kHz in human and 7–20 kHz in cat). The notch is created as an interference pattern when sound propagates through the pinna. Essentially, there are multiple sound paths through the pinna due to reflections. These echoed sounds reach the ear canal after variable delays, depending on the path taken. At the ear canal, sounds can cancel at frequencies where the delays are approximately a half-cycle. The interference patterns so generated produce the notches shown in Fig. 12.2. The notches occur at relatively high frequencies, where the wavelength of sound is comparable to the lengths of the reflection paths in the pinna. In cats, the notch moves toward higher frequencies as the elevation (or the azimuth, not shown) increases [12.6, 16]. At higher frequencies, above 10 kHz in humans and 20 kHz in cat, there are additional complex changes in the transfer function of the external ear that are not easily summarized. Cats appear to be sensitive to spectral cues at all frequencies, but use the notch for localizing sounds and the higher-frequency characteristics for discriminating the locations of two sounds [12.17].

12.1.2 Middle Ear

The function of the middle ear is to transfer sound from the air to the fluids of the cochlea (e.g., from p_T to p_V in Fig. 12.1b). The process can be considered as an impedance transformation [12.18]. The specific acoustic impedance of a medium is the ratio of the sound pressure to the particle velocity of a plane wave propagating through the medium and is a property of the medium itself. When sound impinges on an interface between two media with different impedances, such as an air–water interface, energy is reflected from the boundary. In the case of air and water, the air is a low-impedance medium (low pressure, large velocity) and the water is high impedance medium (high pressure, low velocity). At the boundary, the pressure and velocity must be continuous, i. e. equal across the boundary. If the pressure is equal across the boundary, then the velocities must be different, larger in air than in water because of the differing impedances. A reflection occurs to allow both boundary conditions to be satisfied, i. e., the boundary conditions can only be satisfied by creating a third reflected wave at the boundary. The reflected wave, of course, takes energy away from the wave that propagates through the boundary. Maximum energy transfer across the boundary occurs when the impedance of the source medium equals that of the second medium (no reflection). In the middle ear, the challenge is to transform the low impedance of air to the high impedance of the cochlea, so as to couple as much energy as possible into the cochlea.

The function of the middle ear as a impedance transformer is schematized in Fig. 12.3a which shows a drawing of the eardrum and the ossicles in approximately the anatomically correct position. In this model, the eardrum and oval window are assumed to act as pistons and the ossicles are assumed to act as a lever system, rotating approximately in the plane of the figure around the axis shown by the white dot in the head of the malleus [12.3]. The area of the eardrum (A_{TM}) is larger than the area of the oval window (A_{OW}), which increases the pressure by the ratio $p_V/p_T = A_{TM}/A_{OW}$. The difference in the lengths of the malleus and incus further increases the pressure by the lever ratio L_M/L_I, and decreases the velocity by the same amount. Thus the net pressure ratio is $A_{TM}L_M/A_{OW}L_I$. The area ratio amounts to a factor of 10–40 and the lever ratio is about 1.2–2.5 in mammalian ears [12.3]. For the cat, the pressure ratio should be about 36 dB. The impedance change is the ratio of the pressure and velocity ratios, $A_{TM}/A_{OW}(L_M/L_I)^2$, which is about 29 for the human ear [12.18]. This is less than the ideal value of 3500 for an air–water interface or 135 estimated for an air–cochlea interface, but is still a significant improvement (about 15 dB).

In reality, the motion of middle-ear components is more complex than is assumed in this model [12.2]. First, the eardrum does not displace as a simple piston; second the malleus and incus undergo a more complex motion than the one-dimensional rotation assumed in Fig. 12.3a; and third, there are losses in the motion of the ossicles that are not included in the model. Figure 12.3b shows a comparison of the actual pressure transformation in the middle ear of a cat, given as p_V/p_T. The dashed line is the prediction of the transformer model (36 dB). The cat ear gives a smaller pressure p_V than the transformer model for the reasons listed above, along with the fact that additional impedances in the middle ear must be considered when doing this calculation, such as the middle-ear cavity behind the eardrum, whose air is compressed and expanded as the eardrum moves [12.2]. Figure 12.3c shows the transfer function as it is usually displayed, as a transfer admittance from the eardrum pressure p_T to the velocity in the cochlea v_S.

The overall function of the external and middle ears in collecting sound from a diffuse field in the environment and delivering it to the cochlea is shown in

Amazon
EX1013 Page 444

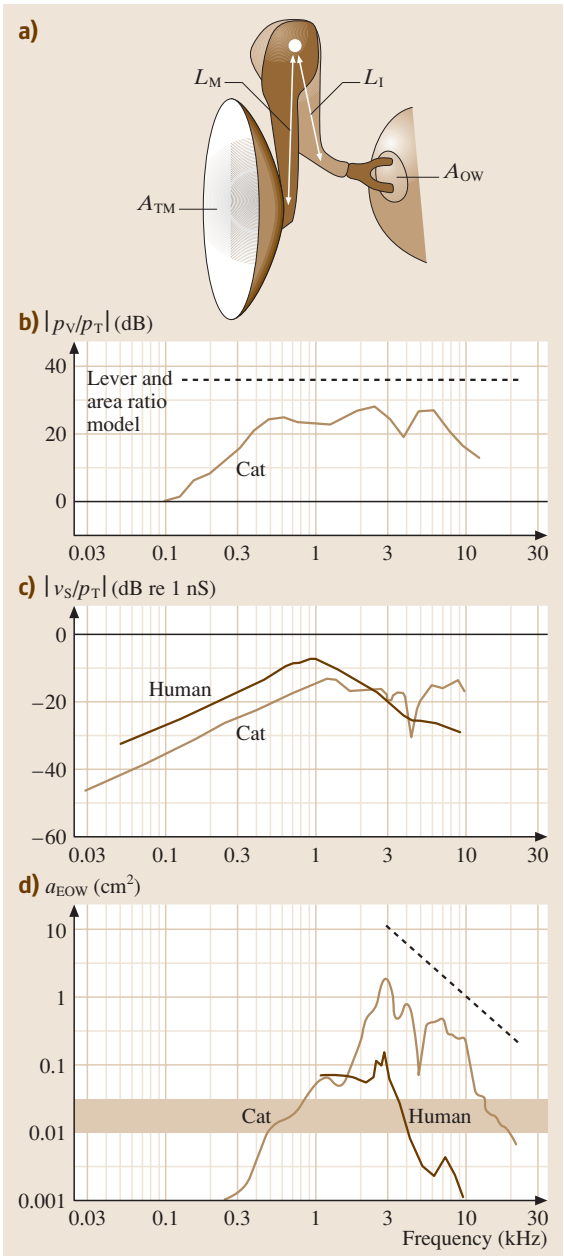


Fig. 12.3d [12.2]. This figure plots the effective area of the ear as a sound collector as in Fig. 12.2c, except now it refers to the sound power delivered to the oval window. Again, the dashed line shows the performance of an ideal spherical receiver and is the same line as in Fig. 12.2c. The effective area has a bandpass shape, as in Fig. 12.2c, with a maximum at 3 kHz. The sharp drop off in performance below 3 kHz was seen in the external ear analysis and occurs because energy is not absorbed at the eardrum at low frequencies. At higher frequencies the effective area tracks the ideal receiver, but is about 10–15 dB smaller than the performance at the eardrum, which approximates the ideal for the cat. This decrease reflects the losses in the middle ear discussed above. Although the external and middle ear do not approach ideal performance, they do serve to couple sound into the cochlea. As a comparison to the effective area in Fig. 12.3d, the cross-sectional area of the oval window is about 0.01–0.03 cm^2 in the cat and human (shaded bar). Thus the effective area is larger than the area of the oval window over the mid frequencies. Moreover, if there were no middle ear, the collecting cross section of the oval window would be smaller by about 15–30 dB because of the impedance mismatch between the air and cochlear fluids [12.19].

Fig. 12.3 **(a)** Schematic drawing of the mammalian middle ear from the eardrum (TM on the *left*) to the oval window (OW on the *right*). The malleus, incus, and stapes are shown in their approximate anatomical arrangement. The areas of the TM and OW are shown along with the lever arms of the malleus (L_M) and incus (L_I). These lever arms are drawn as if the malleus and incus rotate in the plane of the paper around an axis indicated by the *white dot*. In reality, the motion is more complex. **(b)** Ratio of the sound pressure in scala vestibuli (p_V) to the sound pressure at the eardrum (p_T) as a function of frequency, from measurements by *Décory* (unpublished doctoral thesis, 1989 [12.2]). The *dashed line* is the prediction of the model in **(a)** for typical dimensions of the cat middle ear. **(c)** Transfer admittance of the middle ear in human and cat, given as the velocity at the stapes (v_S) divided by the pressure at the eardrum (p_T). **(d)** Performance of the external and middle ears in sound collection plotted as the effective area of the ear, referenced to the oval window. This is the cross-sectional area across which the ear collects sound power in a diffuse sound field, plotted against frequency; the *shaded bar* shows the range of cross sectional areas of the oval window for comparison. Comparing with Fig. 12.2c shows the effect of the middle ear. (After [12.2])

Amazon
EX1013 Page 445

12.2 Cochlea

The cochlea contains the ear's transduction apparatus, by which sound energy is converted into electrical activity in nerve cells. The conversion occurs in transduction cells, called inner hair cells, and is transferred to neurons in the auditory nerve (AN) that connect to the brain. However the function of the cochlea is more than a transducer. The cochlea also does a frequency analysis, in which a complex sound such as speech or music is separated into its component frequencies, a process not unlike a Fourier or wavelet transform. Auditory perception is largely based on the frequency content of sounds; such features as the identity of sounds (which speech sound?), their pitches (which musical note?), and the extent to which they interact in the ear (e.g., to make it hard to hear one sound because of the presence of another) are determined by the mixture of frequencies making them up (Chap. 13).

In this section, the steps in the transduction process in the inner ear or cochlea are described. The problems solved in this process include frequency analysis, mentioned above, but also regulating the sensitivity of process. Transduction must be very sensitive at low sound levels, so that sounds can be detected near the limit imposed by Brownian motion of the components of the cochlea [12.20]. At the same time, the cochlea must function over the wide dynamic range of sound intensities (up to 100 dB or more) that we encounter in the world. Responses to this wide dynamic range must be maintained in neurons with much more limited dynamic ranges, 20–40 dB [12.21]. In part this is accomplished by compressing the sound in the transduction process, so that acoustic sound intensities varying over a range of 60–100 dB are mapped into neural excitation over a much more limited dynamic range. Thus cochlear sensitivity must be high at low sound levels to allow soft sounds to be heard and must be reduced at high sound levels to maintain responsiveness without saturation. Both frequency tuning and dynamic-range adjustment depend on the same cochlear element, the second set of transducer cells called outer hair cells. Whereas inner hair cells convey the representation of sound to the AN, the outer hair cells participate in cochlear transduction itself, making the cochlea more sensitive, increasing its frequency selectivity, and compressing its dynamic range. These processes are described in more detail below.

The remainder of this chapter assumes some knowledge of neural excitation and synaptic transmission. These subjects are too extensive to review here, but can be quickly grasped from an introductory text on neurophysiology or neuroscience.

12.2.1 Anatomy of the Cochlea

The inner ear consists of the cochlea and the AN. The cochlea contains the transduction apparatus; it is a coiled structure, as shown in Fig. 12.1a. A cross-sectional view of a cochlea is shown in Fig. 12.4a. This figure is a low-resolution sketch that is provided to show the locations of the major parts of the structure. The section is cut through the center of the cochlear coil and shows approximately 3.5 turns of the coil. The cochlea actually consists of three fluid-filled chambers (or scalae) that coil together. A cross section of one turn of the spiral showing the three chambers is in Fig. 12.4b. The scala tympani (ST) and scala media (SM) are separated by the basilar membrane, a complex structure consisting of connective tissue and several layers of epithelial cells. The scala vestibuli (SV) is separated from the SM by Reissner's membrane, a thin epithelial sheet consisting of two cell layers. Reissner's membrane is not important mechanically in the cochlea and is usually ignored when discussing cochlear function. In the following, the mechanical properties of the basilar membrane are discussed without reference to the Reissner's membrane and it is usual to speak of the basilar membrane as if it separated the SV and ST.

What Reissner's membrane does do is to separate the fluids of the SV from the SM. The fluids in the SV and ST are typical of the fluids filling extracellular spaces in the body; this fluid, called *perilymph* is basically a dilute NaCl solution, with a variety of other constituents but with low K^+ concentration. In contrast, the fluid in SM, called *endolymph*, has a high K^+ concentration with low Na^+ and Ca^{2+} concentrations. Such a solution is rare in extracellular spaces, but is commonly found bathing the apical surfaces of hair cells in sensory organs from insects to mammals, suggesting that the high potassium concentration is important for hair-cell function. The endolymph is generated in the stria vascularis, a specialized epithelium in the lateral wall of the SM by a complex multicellular active-transport system [12.22].

Mounted on the basilar membrane is the organ of Corti, which contains the actual cochlear transduction apparatus, shown in more detail in Fig. 12.4c. The organ of Corti consists of supporting cells and hair cells. There are two types of hair cells, called inner (IHC) and outer

Amazon
EX1013 Page 446

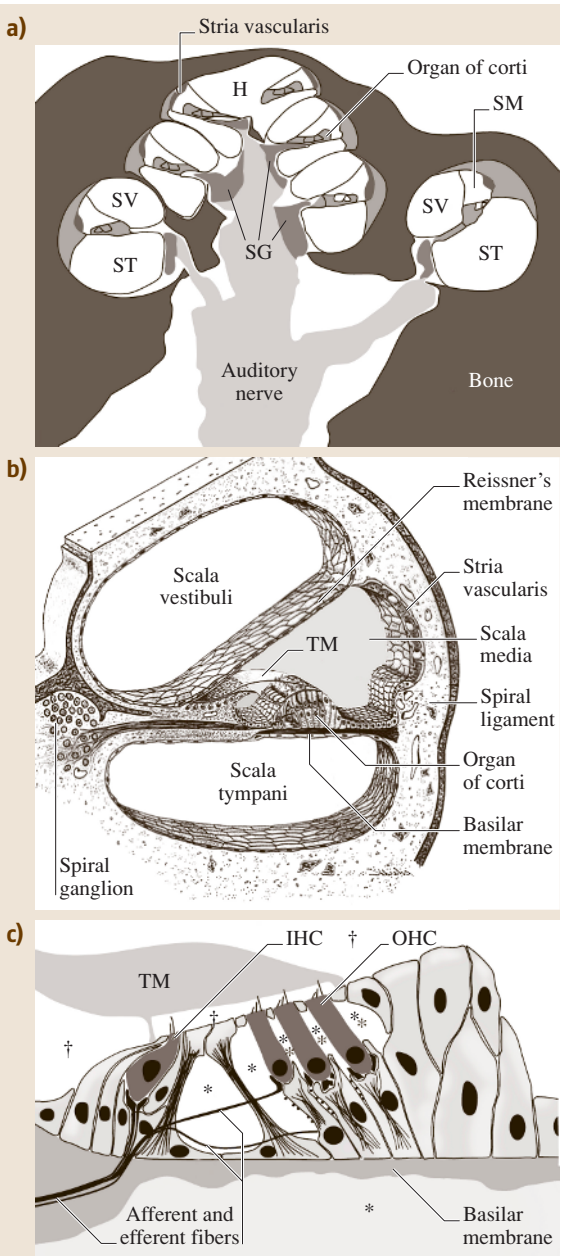


Fig. 12.4 **(a)** Sketch of a cross section of the cochlea, cut through the center of the coil. The cochlear spirals are cut in cross section, shown at higher resolution in **(b)**. The cochlea consists of three fluid-filled chambers that spiral together: ST – scala tympani, SV – scala vestibuli, and SM – scala media. H is the helicotrema, a connection between the SV and ST at the apex of the cochlea. The cochlear transducer has three essential functional parts:

1. the basilar membrane which separates the SM and ST;
2. the organ of Corti, which contains the hair cells, and
3. the stria vascularis which secretes the fluid in the SM and provides an energy source to the transducer.

The AN fibers have their cell bodies in the spiral ganglion (SG); their axons project to the brain in the AN. **(b)** Cross section of the cochlear spiral showing the three scalae. The scalae are fluid-filled; the SV and ST contain perilymph. The SM contains endolymph. **(c)** Schematic drawing of the organ of Corti showing the inner (IHC) and outer hair cells (OHC) and several kinds of supporting cells. The spaces containing perilymph are marked with *asterisks* and the spaces containing endolymph with *daggers* (†). TM is the tectorial membrane, which lies on the organ of Corti and is important in stimulating the hair cells. Nerve fibers enter the organ from the SG (off to the left). Both afferent and efferent fibers are present. Afferents are the SG neurons that are connected synaptically to hair cells, mainly IHCs, and carry information to the brain. Efferents are the axons of neurons located in the brain which connect to OHCs and to the afferents under the IHCs (Fig. 12.5). (After [12.23])

(OHC), because of their positions. The AN fibers and their cell bodies in the spiral ganglion (*SG* in Fig. 12.4a) occupy the center of the cochlear coil. Spiral ganglion cells have two processes: a distal process that invades the organ of Corti and innervates one or more hair cells, usually an IHC, and a central process that travels in the AN to the brain.

The arrangement of the nerve fibers in the cochlea is reviewed in detail by *Ryugo* [12.24] and by *Warr* [12.25] and is summarized in Fig. 12.5. Both afferent and efferent fibers are present. The afferents are the neurons mentioned above with their cell bodies in the SG which carry information from the hair cells to the brain. The efferents are the axons of neurons in the brain that terminate in the cochlea and allow central control of cochlear transduction.

Considering the afferents firsts, there are two groups of SG neurons, called type I and type II. Type I neurons make up about 90–95% the population; their distal processes travel directly to the IHC in the organ of Corti. Each type I neuron innervates one IHC, and each hair cell receives a number of type I fibers; the exact number varies with species and location in the cochlea, but is typically 10–20. The connection between the IHC and type I hair cells is a standard chemical synapse, by which the hair cell excites the AN fiber. Type I AN fibers project into the core areas of the cochlear nucleus, the

Amazon
EX1013 Page 447

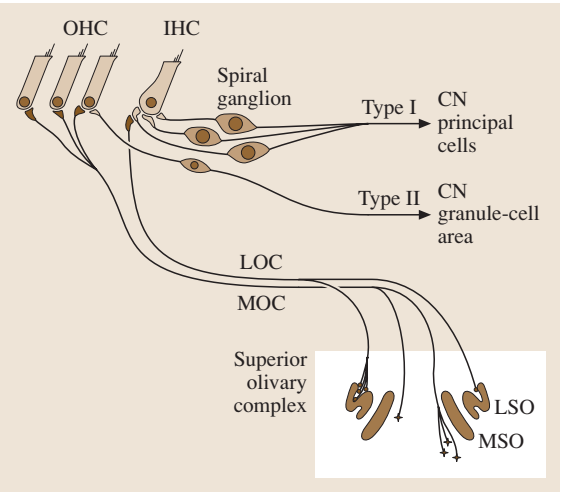


Fig. 12.5 A schematic of the wiring diagram of the mammalian cochlea. The hair cells are on the left. Afferent terminals on hair cells, those in which the hair cell excites an AN fiber, are shown *lightly shaded*; efferent terminals, in which an axon from a cell body in the central nervous system contacts a hair cell or another terminal in the organ of Corti, are shown *heavily shaded*. All the synapses in the cochlea are chemical. Inner hair cells (IHC) are contacted by the distal processes of AN fibers. These fibers have bipolar cell bodies in the spiral ganglion; their axons are myelinated and travel centrally (type I fibers) to innervate principal neurons in the cochlear nucleus (CN). Outer hair cells (OHC) are innervated by a small population of type II spiral ganglion cells whose axons terminate in granule cell regions in the CN. The efferent fibers to the cochlea are called olivocochlear neurons; they originate in the superior olivary complex (SOC, box at lower right), another part of the central auditory system. There are two kinds of efferents: the medial olivocochlear system (MOC), originates near the medial nucleus of the SOC and innervates OHCs. The lateral olivocochlear system (LOC) originates near the lateral nucleus of the SOC and innervates afferent terminals of type I afferent fibers under the IHCs. The MOC and LOC contain fibers from both sides of the brain, although most LOC fibers are ipsilateral and most MOC fibers are contralateral, as drawn. The exact ratios vary with the animal

first auditory structure in the brain; the type I fibers plus the neurons in the core of the cochlear nucleus make up the main pathway for auditory information entering the brain.

The remaining 5–10% of the spiral ganglion neurons, type II, innervate OHCs. The fibers cross the fluid spaces between the IHC and the OHC (shown by the asterisks in Fig. 12.4c) and spiral along the basilar membrane beneath the OHCs toward the base of the cochlea (toward the stapes, not shown) innervating a few OHCs along the way. Although the type II fibers are like type I fibers in that they connect hair cells (in this case OHC) to the cochlear nucleus, there are some important differences. The axons of type II SG neurons are unmyelinated, unlike the type I fibers; their central processes terminate in the granule-cell areas of the cochlear nucleus, where they contact interneurons, meaning neurons that participate in the internal circuitry of the nucleus but do not contribute their axons to the outputs of the nucleus. Finally, type II fibers do not seem to respond to sound [12.26–28], even though they do propagate action potentials [12.29, 30]. It is clear that the type I fibers are the main afferent auditory pathway, but the role of the type II fibers is unknown. In the remainder of this chapter, the term *AN fiber* refers to type I fibers only.

The efferent neurons have their cell bodies in an auditory structure in the brain called the superior olivary complex; for this reason, they are called the olivocochlear bundle (OCB). The anatomy and function of the OCB efferents are reviewed by *Warr* [12.25] and by *Guinan* [12.31]. Anatomically, there are two groups of OCB neurons. So-called lateral efferents (LOC) travel mainly to the ipsilateral cochlea and make synapses on the dendrites of type I afferents under the IHC. Thus they affect the afferent pathway directly. The second group, medial efferents (MOC), travel to both the ipsilateral and contralateral ears and make synapses on the OHCs. Their effect on the afferent pathway is thus indirect, in that they act through the OHC's effects on the transduction process in the cochlea.

12.2.2 Basilar-Membrane Vibration and Frequency Analysis in the Cochlea

An overview of the steps in cochlear transduction is shown in Fig. 12.6a. The sound pressure at the eardrum (p_T) is transduced into motion of the middle-ear ossicles, ultimately resulting in the stapes velocity (v_S), which couples sound energy into the SV. The transformations in this part of the system have mainly to do with acoustical mechanics and were described in the first section above. The stapes motion produces a sound pressure signal in the cochlea that results in vibration of the basilar membrane, which is the topic of this section. The vibration results in stimulation of hair cells, through opening and closing of transduction channels

Amazon
EX1013 Page 448

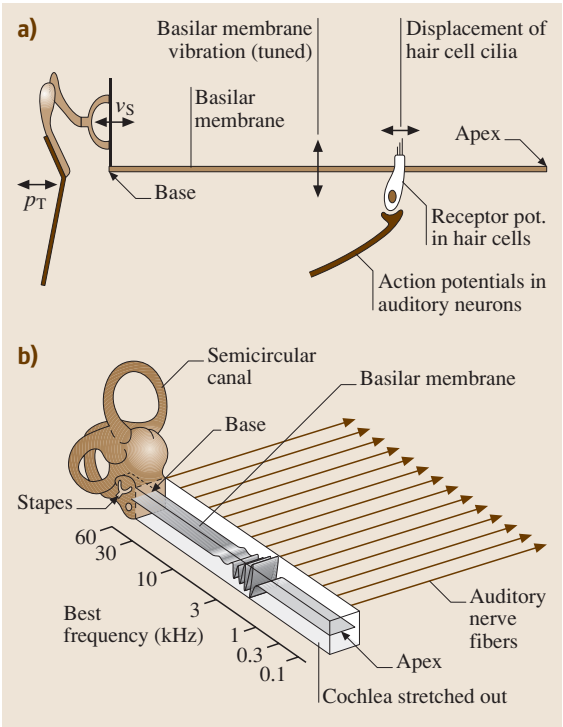


Fig. 12.6 **(a)** Summary of the steps in direct cochlear transduction. **(b)** Schematic diagram showing the basilar membrane in an unrolled cochlea and the nature of cochlear frequency analysis. A snapshot of basilar-membrane displacement (greatly magnified) by a tone with frequency near 3 kHz is shown. The frequency *left scale* shows the location of the maximum membrane displacement at various frequencies for a cat cochlea. For a human cochlea, the frequency scale runs from about 20 Hz to 15 kHz. The array of AN fibers is shown by the *parallel lines on the right*. Each fiber innervates a hair cell at one place on the basilar membrane, so the fiber's sensitivity is maximal at the frequency corresponding to that place, as shown by the *left scale*. In other words, the separation of frequencies done by the basilar membrane is preserved in the AN fiber array. (After [12.32] with permission)

built into the cilia that protrude from the top of the hair cell, described in a later section. The hair cells are synaptically coupled to AN fibers, so that basilar-membrane vibration is eventually coupled to activation of the nerve.

A key aspect of cochlear function is frequency analysis. The nature of cochlear frequency analysis is illustrated in Fig. 12.6b, which shows the cochlea uncoiled, with the basilar membrane stretching from the base, the end near the stapes, to the apex (the helicotrema). As described above, the basilar membrane lies between (separates) the ST and the SM/SV. The sound entering the cochlea through the vibration of the stapes is coupled into the fluids of the SV and SM at the base, above the basilar membrane in Fig. 12.6. The sound energy produces a pressure difference between the SV/SM and the ST, which causes vertical displacement of the basilar membrane. The displacement is tuned, in that the motion produced by sound of a particular frequency is maximum at a particular place along the cochlea. The resulting place map is shown by the frequency scale on the left in Fig. 12.6. Basilar-membrane displacement was first observed and described by *von Békésy* [12.33]; more recent data are reviewed by *Robles* and *Ruggero* [12.34].

Basilar-Membrane Motion

The displacement of any single point on the basilar membrane is an oscillation vertically (perpendicular to the surface of the membrane); the relative phase (or timing) of the oscillations of adjacent points is such that the overall displacement looks like a traveling wave, frequently described as similar to the waves propagating away from an object dropped into a pool of water. Essentially, the motion of the basilar membrane is delayed in time as the observation point moves from base to apex. Figure 12.7a shows an estimate of the cochlear traveling wave on the guinea-pig basilar membrane, based on electrical potentials recorded across the membrane [12.35]. The horizontal dimension in this figure is distance along the basilar membrane, with the base at left and the apex at right. The vertical dimension is the estimate of displacement, shown at a greatly expanded scale. The oscillations marked "2 kHz" show membrane deflections in response to a 2 kHz tone at five successive instants in time (labeled 1–5). The wave appears to travel rightward, from the base toward the apex; as it does so, its amplitude changes. When the stimulus is a tone, the displacement envelope (i. e., the maximum amplitude of the displacement at each point along the basilar membrane) has a single peak at a location that depends on the tone frequency (approximately at point 2 in this case). In Fig. 12.7a, displacement envelopes are shown by dashed lines for 0.5 kHz and 0.15 kHz tones; these envelopes peak at more apical locations, compared to 2 kHz. The locations of the envelope peaks are given by the cochlear frequency map for the cat cochlea, shown by the scale running along the basilar membrane in Fig. 12.6b. Notice that the scale is logarithmic: the position along the basilar membrane corresponds more

Amazon
EX1013 Page 449

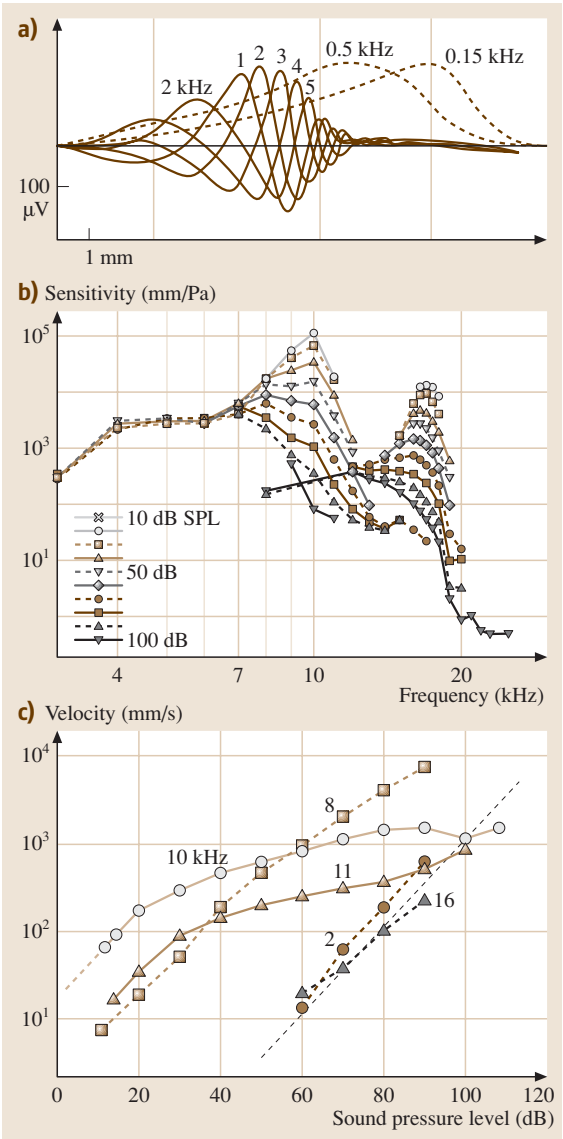


Fig. 12.7 **(a)** An estimate of the instantaneous displacement of the basilar membrane in the guinea-pig cochlea for a 2 kHz tone (*solid lines*) and estimates of the envelope of the displacement for two lower-frequency tones (*dashed lines*). Distance along the basilar membrane runs from left (*base*) to right (*apex*) and the displacement of the membrane is plotted vertically. The estimates were obtained by measuring the cochlear microphonic at three sites, indicated by the three vertical lines, and then interpolating or extrapolating to other locations using an informal method [12.35]; this method can be expected to produce only qualitatively correct results. The 2 kHz estimates are shown at 0.1 ms intervals. Waveforms at successive times are identified with the *numbers "1"* through *"5"*. The *dashed curves* were drawn through the maximum displacements of similar data for 0.5 and 0.15 kHz tones, which were scaled to have the same maximum displacement as the 2 kHz data. The cochlear microphonic is produced by hair cells, mainly OHCs, and is roughly proportional to basilar-membrane displacement. The ordinate scale has been left as microvolts of cochlear microphonic. The scale markers at lower left show distance from the origin. **(b)** Measurements of the gain of the basilar membrane (root mean square displacement divided by pressure at the eardrum) for two data sets. In each case, basilar-membrane displacement at one place was measured with an interferometer. Both frequency (abscissa) and sound level (symbols) were varied. The curves peaking near 9–10 kHz are from the chinchilla cochlea [12.36] and the curves peaking near 20 kHz are from the guinea pig cochlea [12.37] (After [12.34]). **(c)** Basilar-membrane input–output curves, showing the velocity of membrane displacement (*ordinate*) versus sound intensity (*abscissa*), at various frequencies, marked on the plots (kHz). The *dashed line at right* shows the linear growth of basilar-membrane motion (velocity proportional to sound pressure). The curve for 10 kHz is extrapolated to low sound levels (*dashed line at left*) under the assumption of linear growth at low sound levels (Fig. 12.7a after [12.35], Fig. 12.7b after [12.34], and Fig. 12.7c after [12.36])

closely to log than to linear frequency. The logarithmic frequency organization corresponds to a number of phenomena in the perception of sound, such as the fact that musical notes are spaced logarithmically in frequency.

The basilar-membrane responses in Fig. 12.7a illustrate the property of tuning, in that the displacement in response to a tone is confined to a region of the membrane centered on a place of maximal displacement. Tuning can be further understood by plotting basilar-membrane displacement as in Fig. 12.7b. This plot shows the displacement plotted against frequency at a fixed location, thus providing a direct measure of the frequency sensitivity of a place on the membrane. Data are shown at two locations from separate experiments (in different species). The ordinate actually shows basilar-membrane *gain*, defined as membrane displacement divided by sound pressure at the eardrum. At low sound levels (20 dB, plotted with empty circles), the gain peaks at a particular best frequency (BF) and decreases at adjacent frequencies. If such measurements could be repeated at multiple locations along the basilar membrane

Amazon
EX1013 Page 450

(which is difficult because of the restricted anatomical access to the basilar membrane), these gain functions would move toward higher frequencies at more basal locations in the cochlea, as predicted by the displacement functions in Fig. 12.7a.

Notice that the gain and tuning of the basilar-membrane response near the BF varies with sound level. At low sound levels (20 dB in Fig. 12.7b), gain is high and the tuning is sharp (i. e. the width of the gain functions is small), giving a clear BF; at high sound levels (100 dB), the gain is substantially reduced, especially in the vicinity of the low-sound-level BF, and the frequency at which the gain is maximum moves to a lower frequency, often about a half-octave lower. The tuning also becomes much broader.

Basilar Membrane and Compression

The fact that basilar-membrane gain changes with sound level was first described by *Rhode* [12.38]; this finding revolutionized our understanding of the auditory system, most importantly by introducing the idea of an active cochlea, meaning one in which the acoustic energy entering through the middle ear is amplified by an internal energy-utilizing process to produce cochlear responses [12.39–44]. The existence of amplification has been demonstrated by calculations of energy flow in the cochlea. However, a simpler evidence for cochlear amplification is the change in cochlear gain functions with death. The dependence of gain on sound level seen in Fig. 12.7b occurs only in the live, intact cochlea; after death, the cochlear gain functions resemble those at the highest sound levels in living animals (e.g. at 100 dB in Fig. 12.7b). Most important, the post-mortem gain functions are linear, meaning that the gain is constant, regardless of stimulus intensity, at all frequencies. In the live cochlea, as shown in Fig. 12.7b, the gain functions are linear only at low frequencies (below 6 kHz or 12 kHz for the two sets of data shown). Presumably, the post-mortem gain functions are the result of passive basilar-membrane mechanics, i. e., the result of the mechanical properties of the cochlea without any energy sources. The difference between the gain at the highest sound levels in Fig. 12.7b and those at low sound levels thus reflects an amplification process, often called the cochlear amplifier.

Additional evidence for a cochlear amplifier comes from the study of otoacoustic emissions, OAE [12.45, 46]. OAEs are sounds produced in the cochlea that can be measured at the eardrum, after propagating in the backwards direction through the middle ear. OAEs can be produced by reflection of the cochlear traveling wave from irregularities in the cochlea [12.47, 48] or from nonlinear distortion in elements of the cochlea [12.49]. In either case amplification of the sound is necessary to explain the characteristics of the emitted sound [12.46], but see also [12.50]. There are several different kinds of OAEs, varying in the mode of production. Most relevant for the present discussion are the spontaneous emissions, sounds that are present in the ear canal without any external sound source. Such sounds necessitate an acoustic energy source in the cochlea, of the type postulated for the cochlear amplifier.

A variety of evidence suggests that cochlear amplification depends on the OHCs. When these cells are destroyed, as by an ototoxic antibiotic, the sharply tuned high-sensitivity portion of the tuning of auditory-nerve fibers is lost [12.51]; presumably this change reflects a similar change at the level of the basilar membrane, i. e., loss of the high-gain sharply tuned portion of the basilar-membrane response. Electrical stimulation of the MOC (Fig. 12.5), which affects only the OHC, produces similar losses in both basilar-membrane motion [12.52] and neural responses [12.53, 54]. Finally, stimulation of the MOC reduces OAE [12.55, 56]. In each case, a neural input to a cellular element of the cochlea, the OHC, affects mechanical processes in the cochlea. These data suggest that the OHCs participate in producing basilar-membrane motion. The possible mechanisms for this effect are discussed in Sect. 12.2.4.

The sound-level dependence of gain observed in Fig. 12.7b shows that the sensitivity of the cochlea is regulated across sound levels. The gain is high at low sound levels and decreases at higher sound levels; as a result, the output varies over a narrower dynamic range than the input, which is compression. A different view of compression is shown in Fig. 12.7c, which plots basilar-membrane input/output functions at various frequencies for a 10 kHz site on a chinchilla basilar membrane [12.36]. Basilar-membrane response is measured as velocity, instead of displacement as in previous figures. However, velocity and displacement are proportional for a fixed frequency, so in terms of input/output functions it does not matter which variable is plotted. Figure 12.7c has logarithmic axes, so a linear growth of response, in which the output is proportional to the input, corresponds to a line with slope 1, as for the dashed line at right. At low sound levels (< 40 dB), the response is largest at 10 kHz, as it should be for the BF. The response to 10 kHz is roughly linear at very low sound levels (< 20 dB), but has a lower slope at higher levels. Over the range of input sound levels from 20–80 dB SPL, the output velocity increases only 10-

Amazon
EX1013 Page 451

fold or 20 dB, a compression factor of about 0.3. At lower frequencies (8 and 2 kHz), the slope is closer to 1, as expected from the constant gain at low frequencies in Fig. 12.7b. At higher frequencies (11 kHz), the growth of response is also compressive, but with a lower gain. Finally, at very high frequencies (16 kHz), the growth is again approximately linear.

The behavior shown in Figs. 12.7b,c is typical of cochleae that are judged to be in the best condition during the measurements. As the condition of the cochlea deteriorates, the input/output functions become more linear and their gain decreases [12.57], reflecting a loss of cochlear amplification. Although such data are not available for the example in Fig. 12.7c, the typical post-mortem behavior of input/output functions at BF is something like the dashed line used to illustrate linear growth. Post-mortem functions have a slope of 1 at all sound levels and typically intercept the compressive BF input/output function at levels of 70–100 dB. Assuming that the cochlear amplifier is not functioning post-mortem, the gain of the amplifier can be defined as the horizontal distance between the linear-growth portion of the BF curve in the normal ear and the post-mortem curve. With this definition of gain, the compression region of the input/output function can be understood as resulting from a gradual decrease in amplification as the sound grows louder.

The compression of basilar-membrane response shown in Fig. 12.7c has a number of perceptual correlates [12.58]. The degree of compression can be measured in human observers using a masking technique, motivated by the behavior of basilar-membrane data like Fig. 12.7 [12.59, 60]; the compression measured in these studies is comparable to that measured in basilar-membrane data. Moreover, compression can be used to explain a number of perceptual phenomena, including aspects of temporal processing and loudness growth. In hearing-impaired persons, compression is lost to varying degrees, consistent with the effects of loss of OHCs on basilar-membrane responses discussed above (i. e., the slopes of basilar-membrane input/output functions steepen). Some of the effects of hearing impairment can be explained by a change in compression ratio from 0.2–0.3 to a value near 1. An important example is loudness recruitment. If loudness is somehow proportional to the overall degree of basilar-membrane motion, then a steepening of the growth of basilar-membrane motion with sound intensity should lead to a steepening of loudness growth with intensity. This is the major effect observed on loudness growth with hearing impairment.

Mechanisms Underlying Basilar-Membrane Tuning

The current understanding of how the properties of basilar-membrane motion arise is incomplete. Work on this subject has proceeded along two lines:

1. collection of evermore accurate data on the motion of the basilar membrane and the components of the organ of Corti; and
2. models of the observed motion of the basilar membrane based on physical principles of mechanics and fluid mechanics.

Progress in this field has been limited by the difficulty of obtaining data, however, and modelers have generally responded vigorously to each new advance in observing and measuring basilar-membrane motion. At present, the major impasse in this work seems to be the difficulty of observing the independent motion of the components of the organ of Corti. Such data are necessary to resolve questions about the mode of stimulation of hair cells by the motion of the basilar membrane and questions about the effects of OHC and stereociliary movements on the basilar membrane.

Modeling of basilar-membrane motion has been approached with a variety of methods in an attempt to account for the properties of the motion. Comprehensive reviews of this literature and its relationship to important aspects of data on basilar-membrane motion are available [12.61, 62]. The next paragraphs provide a rough summary of the current understanding of this topic.

Tuning or the separation of frequencies in basilar-membrane motion is usually attributed to the variation in the stiffness and (sometimes) the mass of the basilar membrane along its length. In most basilar-membrane models, the points on the basilar membrane are assumed to move independently; that is, the longitudinal stiffness coupling adjacent points on the membrane is assumed to be small. With this assumption, the resonant frequency of a point on the basilar membrane should vary approximately as the square root of its stiffness divided by its mass, using the usual formula for the resonant frequency of a mass–spring oscillator. In so-called long-wave models, it is this resonant frequency that determines the mapping of frequency into place in the cochlea. Experimentally, the stiffness of the basilar membrane varies from high near the base to low at the apex [12.33, 63]. Thus the stiffness gradient is qualitatively consistent with the observed place–frequency map; however, whether the stiffness gradient is sufficient

Amazon
EX1013 Page 452

to fully account for tuning in the cochlea is not settled, because of the difficulty of making measurements.

Recent efforts in cochlear modeling have been devoted to micromechanical models, i. e., models of the detailed mechanics of the organ of Corti and the tectorial membrane [12.49, 62]. Movement of the basilar membrane is coupled to the hair cells through the organ of Corti, particularly through the relative movement of the top surface of the organ (the reticular lamina) in which the hair cells are inserted and the overlying tectorial membrane (refer to Fig. 12.4c). Mechanical feedback from the OHCs is similarly coupled into the basilar membrane through the same structures. The structure of the organ of Corti is complex and heterogeneous, with considerable variation in the stiffness of its different parts [12.64]. Thus a full account of the properties of the basilar membrane awaits an adequate micromechanical model, based on accurate data on the mechanical properties of its components.

In both models and measurements, the sound pressure difference across the basilar membrane drops rapidly to zero past the point of resonance. This behavior is illustrated by the high-frequency part of the data in Fig. 12.7b and the lack of deflection beyond the peak of the responses in Fig. 12.7a. This fact means that there is no pressure difference across the basilar membrane at the apical end of the cochlea, so the helicotrema does not affect basilar-membrane mechanics (in particular it does not *short out* the pressure difference driving the basilar membrane) except at very low frequencies, where the displacement is large near the apex. The role of the helicotrema is presumably to prevent the generation of a constant pressure difference across the basilar membrane; such a difference would interfere with the delicate transducer system in the organ of Corti.

The frequency analysis done by the basilar membrane is preserved by the arrangement of hair cells and nerve fibers in the cochlea (Fig. 12.6b). A particular hair cell senses only the local motion of the basilar membrane, so that the cochlear place map is recapitulated in the population of hair cells spread out along the basilar membrane. Moreover, each AN fiber innervates only one IHC, so the cochlear place map is again recapitulated in the AN fiber array. As a result, AN fibers that innervate, say, the 3 kHz place on the basilar membrane are maximally activated by 3 kHz energy in the sound and it is the level of activation of those fibers that provides information to the brain about the 3 kHz energy in the sound. The representation of sound in an array of neurons tuned to different frequencies is the basic organizing principle of the whole auditory system and is called *tonotopic* organization.

12.2.3 Representation of Sound in the Auditory Nerve

Before discussing cochlear transduction in hair cells, it is useful to introduce some properties of the responses to sound of AN fibers. Of course, the properties of AN fibers derive directly from the properties of the basilar membrane and the transduction process in IHCs. AN fibers encode sound using trains of action potentials, which are pulses fired by a fiber when stimulated by a hair-cell synapse. Information is encoded in the rate at which the fiber discharges action potentials or by the temporal pattern of action potentials. Fibers are active spontaneously at rates that vary from near 0 to over 100 spikes/s. When stimulated by an appropriate sound, the fiber's discharge rate increases and the temporal patterning of action potentials often changes as well. Figure 12.8 shows some basic features of the encoding process. Reviews of the representation of sound by auditory neurons are provided by *Sachs* [12.65], *Eggermont* [12.66] and *Moore* [12.67]. Another view is provided by recent attempts to model comprehensively the responses of AN fibers [12.68–70] or to analyze the information encoded using theoretical methods [12.71].

The tuning of the basilar membrane is reflected in the tuning of AN fibers as is shown by the tuning curves in Fig. 12.8a. This figure shows tuning curves from 11 AN fibers with different BFs. Each curve shows how intense a tone has to be (ordinate) in order to produce a criterion change in discharge rate from a fiber, as a function of sound frequency (abscissa). The tuning curves qualitatively resemble the basilar-membrane response functions shown in Fig. 12.7b for the lowest sound levels (after being turned upside down). The comparison should not be exact because the data are from different species and the tuning curves in Fig. 12.8a are constant-response contours whereas the basilar-membrane functions in Fig. 12.7b are constant-input-level response functions. However, a quantitative comparison of threshold functions for the basilar membrane and AN fibers from the same species yields a good agreement, and it is generally considered that AN tuning is accounted for by basilar-membrane tuning [12.34]. The AN consists of an array of fibers tuned across the range of frequencies that the animal can hear. The frequency content of a sound is conveyed by which fibers are activated, the tonotopic representation discussed in the previous section.

Amazon
EX1013 Page 453

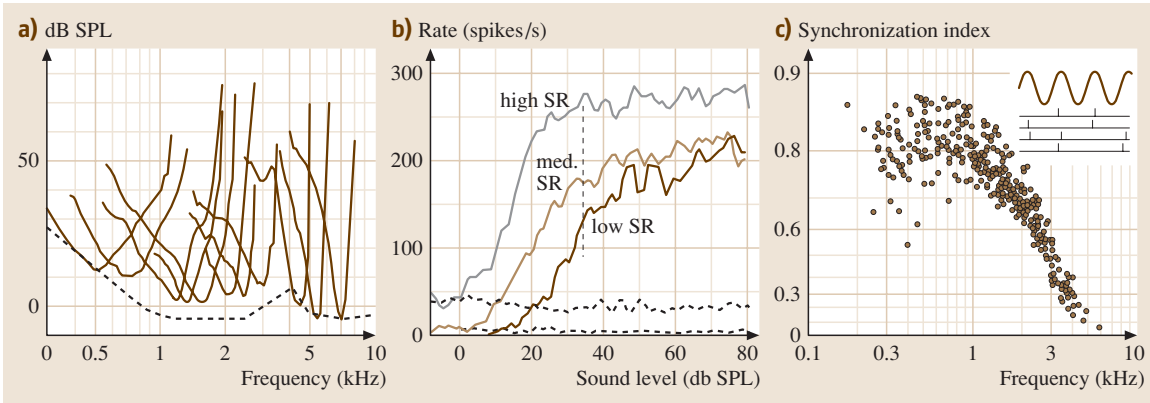


Fig. 12.8a–c Basic properties of responses to sound in AN fibers. **(a)** Tuning curves, showing the threshold sound level for a response plotted versus frequency. the *dashed line* shows the lowest thresholds across a population of animals. **(b)** Rate versus sound-level plots for responses to BF tones in three AN fibers. Rate in response to a 200 ms tone is shown by the *solid lines* and SR by the *dashed lines*, which actually plot the SR during the 400 ms immediately preceding each stimulus. The fibers had similar BFs (5.36, 6.18, and 5.74 kHz) but different spontaneous rates and were recorded successively in the same experimental preparation. The fluctuations in the curves derive from the randomness in the responses remaining after three-point smoothing of the curves. **(c)** Strength of phase-locking in a population of AN fibers to a BF tone plotted as a function of BF. Phase-locking is measured as the synchronization index, equal to the magnitude of the Fourier transform of the spike train at the stimulus frequency divided by the average rate. The *inset* illustrates phase-locked spike trains (Fig. 12.8a after [12.74], Fig. 12.8c after [12.75])

The perceptual sense of sound frequency, of the highness or lowness of a simple sound like a tone, is strongly correlated with stimulus frequency and, therefore, with the population of AN fibers activated by the sound. However, the pitch of a complex sound, as for musical sounds or speech, is a more complex attribute both in terms of the relationship of pitch to the physical qualities of the sound [12.72] and in terms of the representation of pitch in the AN. A discussion of this issue is provided by *Cariani* and *Delgutte* [12.73].

The intensity of a sound is conveyed by the rate at which fibers respond. AN fibers increase their discharge rates as sound intensity increases (Fig. 12.8b). In mammalian ears, fibers vary in their threshold sensitivity, where threshold means the sound level at which the rate increases from the spontaneous rate. The threshold variation is correlated with spontaneous discharge rate (SR), so that very sensitive fibers, those with the lowest thresholds, have relatively high SR and less-sensitive fibers have lower SR [12.76]. Fibers are broken into three classes, low, medium, and high, using the somewhat arbitrary criteria of 0.5–1 spikes/s to divide low from medium and 15–20 spikes/s to divide medium from high, depending on the experimenters. Figure 12.8b shows plots of discharge rate versus sound level for an example fiber in each SR group, labeled on the figure. Fibers of all SRs have similar rate-level functions; for tones, these increase monotonically with sound level (except for the effect of noisy fluctuations) from spontaneous rate to a saturation rate. In high-SR fibers, the dynamic range, the range over which sound intensity increases produce rate increases, is narrow and the saturation of discharge rate at high sound levels is clear. In low- and medium-SR fibers, the saturation is gradual (sloping) and the dynamic range is wider.

The sloping saturation in low- and medium-SR fibers is thought to reflect compression in basilar-membrane responses [12.21, 77]. High-SR fibers do not show sloping saturation because the fiber reaches its maximum discharge rate at sound levels where growth of the basilar-membrane response is linear (e.g. below 20 dB in Fig. 12.7). By contrast, the low- and medium-SR fibers sample both the linear and compressed portion of the basilar-membrane input/output function. The vertical dashed line in Fig. 12.8 shows the approximate threshold for basilar-membrane compression inferred for these fibers.

The relationship of AN responses to the perceptual sense of loudness is an apparently simple problem that has not been fully solved. While it is generally assumed that loudness is proportional to the overall increase in discharge rate across all BFs [12.78], this

Amazon
EX1013 Page 454

calculation does not predict important details of loudness growth [12.79–81], especially in ears with damage to the hair cells. Presumably there are additional transformations in the central auditory system that determine the final properties of loudness growth.

A property of AN discharge that is important for many sounds is the ability to represent the temporal waveform. Sounds such as speech have information encoded at multiple levels of temporal precision [12.82]; these include

1. syllables and other aspects of the envelope at frequencies below 50 Hz;
2. periodicity (pitch) in sounds such as vowels at frequencies from 50–500 Hz; and
3. the fine structure of the actual oscillations in the acoustic waveform at frequencies above 500 Hz.

AN fibers represent the waveform of sounds (the fine structure) by phase-locking to the stimulus. A schematic example is shown in the inset of Fig. 12.8c, which shows a sinusoidal acoustic waveform (a tone) and four examples of AN spike trains in response to the stimulus. The important point is that the spikes in the responses do not occur at random, but rather at a particular stimulus phase, near the positive peak of the stimulus waveform in this example. The phase-locking is not perfect, in that a spike does not occur on every cycle of the stimulus, and the alignment of spikes and the waveform varies somewhat. However, a histogram of spike-occurrence times shows a strong locking to the stimulus waveform and the Fourier transform of the spike train generally has its largest component at the frequency of the stimulus.

Phase-locking occurs at stimulus frequencies up to a few kHz, depending on the animal. The main part of Fig. 12.8c shows the strength of phase-locking in terms of the synchronization index (defined in the caption) plotted against the frequency of the tone. These data are from the cat ear [12.75] where phase-locking is strongest below 1 kHz and disappears by about 6 kHz. Phase-locking always shows this low-pass property of being strong at low frequencies and nonexistent at high frequencies. The highest frequency at which phase-locking is seen varies in different species, being lower in rodents than in cats [12.83].

AN phase-locking is essential for the perception of sound localization. The relative time of occurrence of spikes from the two ears is used to compute the delays in the stimulus waveform across the head that provide information about the azimuth of a sound source [12.84]. Phase-locking has also been suggested as a basis for processing of speech and other complex stimuli [12.71,85]. Examples of the usefulness of phase-locking in analyzing responses to complex stimuli are given in Sect. 12.3.1.

12.2.4 Hair Cells

The transduction of basilar-membrane motion into electrical signals occurs in the IHCs. The OHCs participate in the generation of sensitive and sharply tuned basilar-membrane motion. In this section, the physiology of IHCs and OHCs are described with reference to these tasks. At the time of this writing, hair-cell research is one of the most active areas in auditory neuroscience and many of the details of hair-cell function are just being worked out. Useful recent reviews of the literature are available [12.86–89].

IHCs and Transduction

Both IHCs and OHCs transduce the motion of the basilar membrane. This section describes transduction with reference to the IHCs, but transduction works in essentially the same fashion in OHCs. The transduction apparatus depends on the arrangement of stereocilia on the apical surface of hair cells. Stereocilia are structures that protrude from the hair cells into the endolymphatic space of the SM (Fig. 12.9a). They consist of bundles of actin filaments anchored in an actin/myosin matrix just under the apical surface of the cell. The membrane of the cell wraps the actin rods so that they are intracellular and the membrane potential of the cell appears across the membranes of the stereocilia. The structure of stereocilia is intricate and involves a number of structural proteins that are important for organizing the development of the cilia and for maintaining them in proper position on the cell [12.90]. The stereocilia form a precisely organized bundle; in the cochlea, the bundle consists of several roughly V- (IHCs) or W-shaped (OHCs) rows of cilia (typically three, as in Fig. 12.9) of graded length, so that the cilia in the row nearest the lateral edge of the hair cell are the longest and those in the adjacent rows are successively shorter. Each row consists of 20–30 cilia precisely aligned with the cilia in adjacent rows. The tips of the cilia are connected by tip links [12.91], a string-like bundle of protein that connects the tip of a shorter cilium to the side of the taller adjacent cilium (Fig. 12.9b).

Transduction occurs when vertical motion of the basilar membrane is converted into a shearing motion of the stereocilia (the arrow in Fig. 12.9b). It has long been known that hair cells are functionally polarized, in the sense that they respond most strongly to displacement of the cilia in the direction of the arrow, are

Amazon
EX1013 Page 455

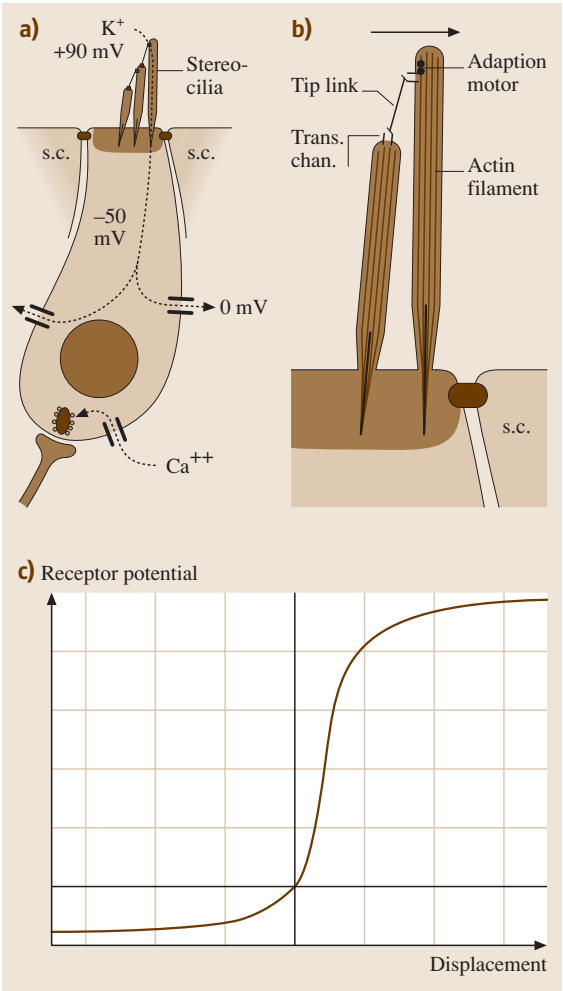


Fig. 12.9a–c Transduction in IHCs. **(a)** Schematic of an IHC showing the components important for transduction. The stereocilia protrude into the endolymphatic space (*top*), containing mainly K^+; the extracellular potential here is about +90 mV (the endolymphatic potential). The transduction apparatus consists of tip links and channels near the tips of the cilia. The *dashed lines* show the transduction current. It enters the cell through the transduction channel and exits through K^+ channels in the basolateral membrane of the cell into the perilymphatic space. The intracellular and extracellular potentials of the cell and the perilymphatic space are given. There is a mechanically stiff and electrically insulating boundary between the endolymphatic and perilymphatic spaces, formed by tight junctions between the apical surfaces of hair cells and supporting cells (s.c.), indicated by the *small black ovals*. The synapse, at the base of the cell, is a standard chemical synapse, in which voltage-gated Ca^{2+} channels open in response to the depolarization caused by the transduction current and admit Ca^{2+} to activate neurotransmitter release. **(b)** Detailed view of two stereocilia with a tip link connecting ion channels in the membrane of each. The upper ion channel is connected to an adaptation motor that moves up and down along the actin filaments to tension the tip link. **(c)** Transduction function showing the receptor potential (the depolarization or hyperpolarization) of a hair cell versus the displacement of the stereocilia tips. Typical axis scales would be 0.01–1 μm per tick on the abscissa and ≈2 mV per tick on the ordinate

inhibited by displacement in the opposite direction, and respond weakly to displacement of the cilia in lateral directions [12.92, 93]. In fact, hair cells are depolarized when the stereocilia are displaced in the direction that stretches the tip links. Further evidence to associate the tip links to the transduction process is that when the tip links are destroyed by lowering the extracellular calcium concentration, transduction disappears and reappears as the tip links are regenerated [12.94].

The mechanical connection between basilar-membrane motion and cilia motion is not fully understood. As shown in Fig. 12.4c, the stereocilia bundles are oriented so that lateral displacement of the cilia (i. e., displacement away from the central axis of the cochlea) should be excitatory. Furthermore the tips of the OHC cilia are embedded in the tectorial membrane, whereas those of the IHC are not. Based on the geometry of the organ, upward motion of the basilar membrane (i. e., away from the ST and toward the SM and SV) results in lateral displacement of the cilia through the relative motions of the organ of Corti and the tectorial membrane. Presumably the coupling to the OHC is a direct mechanical one, whereas the coupling to the IHC is via fluid movements, so the IHC stereocilia are bent through viscous drag of the endolymph in the space between the organ of Corti and the tectorial membrane. While this model is commonly offered and is probably basically correct, it is not consistent in detail with the phase relations between the stimulus (or basilar-membrane motion) and action potential phase locking in AN fibers [12.95]. Presumably the inconsistencies reflect complexities in the motions of the organ of Corti that connect basilar-membrane displacement to stereociliary displacement.

The current model for the tip link and transduction channel is shown in Fig. 12.9b. The tip link connects (probably) two transduction channels, one in each cilium. When the bundle moves in the excitatory direction (the direction of the arrow), the tension in the tip link

Amazon
EX1013 Page 456

increases, which opens the transduction channels and allows current flow into the cell. Movement in the opposite direction relaxes the tension in the tip link and allows the channels to close. The current through the transduction channels is shown by the dashed line in Fig. 12.9a. Outside of the stereociliary membrane is the endolymph of the SM. The transduction channels are nonspecific channels which allow small cations to pass; in the cochlea, these are mainly Na^+, K^+, and Ca^{2+}. However, the predominant ion in both the extracellular (endolymph) and intracellular spaces is K^+, so the transduction current is mostly K^+. The energy source producing the current is the potential difference between the SM (the endolymphatic potential, approximately +90 mV) and the intracellular space in the hair cell (approximately −50 mV). The endolymphatic potential is produced by the active transport of K^+ into the endolymph and Na^+ out of the endolymph in the stria vascularis (Fig. 12.4b). This transport process also produces the endolymphatic potential [12.22]. The negative potential inside the hair cell is produced by the usual mechanisms in which active transport of K^+ into the cell and Na^+ out of the cell produces a negative diffusion potential through the primarily K^+ conductances of the hair cell membrane. Because the transduction current is primarily K^+, it does not burden the hair cell with the necessity for additional active transport of ions brought into the hair cell by transduction.

When the transduction channels open, the hair cell is depolarized by the entry of positive charge into the cell. The transduction current flows out of the cell through the potassium channels in its membrane. Figure 12.9c shows the typical transduction curve for a cochlear hair cell, as the depolarization of the membrane (ordinate) produced by a certain displacement of the stereociliary bundle (abscissa). At rest (zero displacement), some transduction channels are open, so that both positive and negative displacements of the bundle can be signaled. However, hair cells generally have only a fraction of channels open at rest so that a much larger depolarization is possible than hyperpolarization, i. e., more channels are available for opening than for closing. The transduction curve saturates when all channels are closed or all channels are open.

The transduction process is completed by the opening of voltage-gated Ca^{2+} channels in the hair cell membrane, which allows Ca^{2+} ions to enter the cytoplasm (dotted line at the bottom of Fig. 12.9a) and activate the synapse. The hair-cell synapse has an unusual morphology, containing a synaptic ribbon surrounded by vesicles [12.96,97]. This ribbon presumably reflects molecular specializations that allow the synapse to be activated steadily over a long period of time without losing its effectiveness and also to produce an action potential in the postsynaptic fiber on each release event, so that summation of postsynaptic events is not necessary, as is required for phase-locking at kilohertz frequencies. The synapse appears to be glutamatergic [12.98] with specializations for fast recovery from one synaptic release, also necessary for high-frequency phase-locking.

The final component of the transduction apparatus is the adaptation motor, shown as two black circles in Fig. 12.9b [12.99]. Presumably the transduction channel and the adaptation motor are mechanically linked together so that when the motor moves, the channel moves also. The sensitivity of the transduction process depends on having an appropriate tension in the tip link. If the tension is too high, the channels are always open; if the tip link is slack, then the threshold for transduction will by elevated because larger motions of the cilia will be required to open a transduction channel. The tip-link tension is adjusted by a myosin motor that moves upward along the actin filaments when the transduction channel is closed. When the channel opens, Ca^{2+} flows into the stereocilium through the transduction channel and causes the motor to slip downward. Thus the adaptation motor serves as a negative feedback system to set the resting tension in the tip link. The zero-displacement point on the transduction curve in Fig. 12.9c is set by an equilibrium between the motor's tendency to climb upward and the tension in the tip link, which causes the channel to open and admit Ca^{2+}, pulling it downward. Adaptation also regulates the sensitivity of the transduction process in the presence of a steady stimulus.

There is a second, fast, mechanism for adaptation in which Ca^{2+} entering through an open transduction channel acts directly on the channel to close it, thus moving the transduction curve in the direction of the applied stimulus [12.88]. This mechanism is discussed again in a later section, because it also can serve as a kind of cochlear amplifier.

IHC Transduction and the Properties of AN Fibers

The description of transduction in the previous section suggests that the stimulus to an AN fiber should be approximately the basilar-membrane motion at the point in the cochlea where the IHC innervated by the fiber is located. It follows that the tuning and rate responses of the fiber should be similar to the tuning and response amplitude of the basilar membrane, discussed in connection with Fig. 12.7.

Amazon
EX1013 Page 457

The receptor potentials in an IHC responding to tones are shown in Fig. 12.10 [12.83]. The traces are the membrane potentials recorded in an IHC, each trace at a different stimulus frequency. These potentials are the driving force for the Ca^{2+} signal to the synapse, and thus indirectly for the AN fibers connected to the cell. The properties of these potentials can be predicted from Fig. 12.9c by the thought experiment of moving a stimulus sinusoidally back and forth on the abscissa and following the potential traced out on the ordinate. The potential should be a distorted sinusoid, because depolarizing responses are larger than hyperpolarizing ones, and it should have a steady (DC) offset, also because of the asymmetry favoring depolarization. At low frequencies, both components can be seen in Fig. 12.10; the response is a distorted sinusoid riding on a steady depolarization. At high frequencies (>1000 Hz), the sinusoidal component is reduced by low-pass filtering by the cell's membrane capacitance and only the steady component is seen. The transition from sinusoidal receptor potentials at low frequencies to rectified DC potentials at high frequencies corresponds to the loss of phase-locking at higher frequencies in AN fibers (Fig. 12.8c). At low frequencies, the IHC/AN synapse releases transmitter on the positive half cycles of the receptor potential and the AN fiber is phase-locked; at high frequencies, there is a steady release of neurotransmitter and the AN fiber responds with a rate increase but without phase-locking.

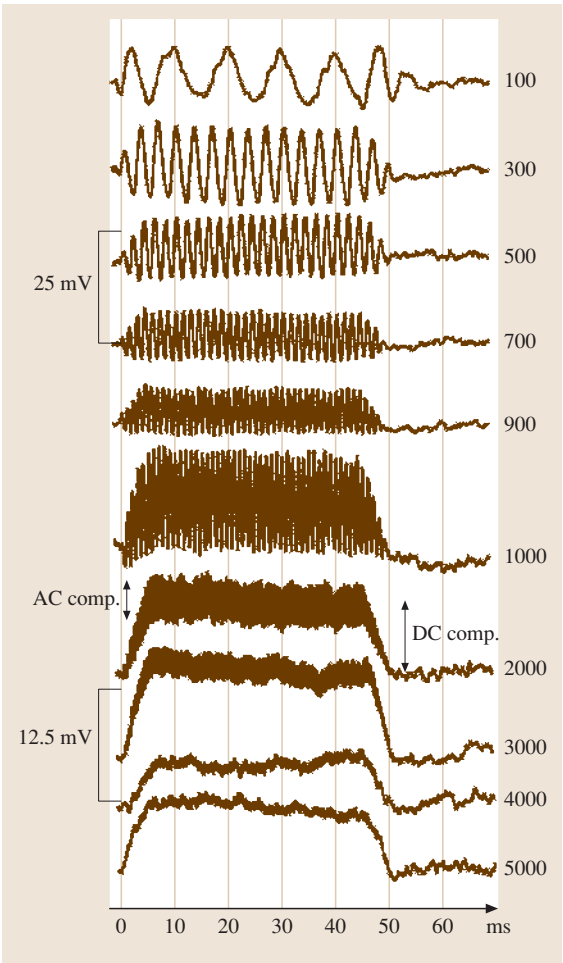


Fig. 12.10 Receptor potentials in an IHC in the basal turn of a guinea-pig cochlea. The stimulus level was set at 80 dB SPL so that the cell would respond across a wide range of frequencies. Stimulus frequency is given at right. Notice the change in ordinate scale between the 900 Hz and 1000 Hz traces. (After [12.83], with permission)

OHCs and the Cochlear Amplifier

In the discussion of the basilar membrane, the need for a source of energy in the cochlea, a cochlear amplifier, was discussed. At present, two possible sources of the amplification have been suggested. The first is OHC motility [12.86, 100] and the second is hair bundle motility [12.101]. Although hair bundle motility could function in both IHCs and OHCs, it seems likely that

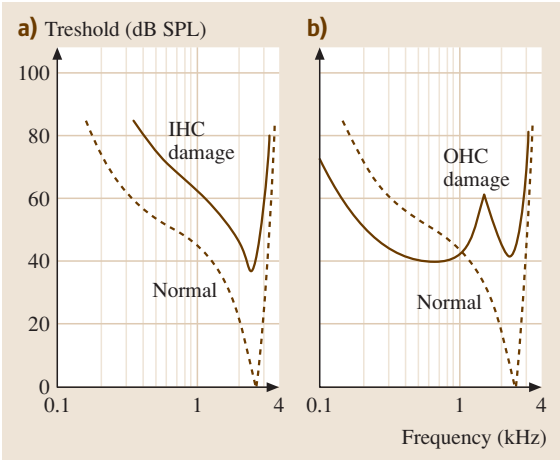


Fig. 12.11a,b Schematic summary of AN tuning curves in normal ears and in ears with damaged cochleae. **(a)** In regions of the cochlea with remaining OHCs but damaged IHCs tuning curves mainly show a threshold elevation. **(b)** In regions with missing OHCs but intact IHCs there is an elevation of threshold at frequencies near BF and a substantial broadening of tuning. (After [12.51])

Amazon
EX1013 Page 458

the OHC is the principal element of the cochlear amplifier. The evidence for this is that the properties of the cochlea change in a way consistent with modification of cochlear mechanics when the OHC are damaged or otherwise manipulated, but not the IHC. For example, electrical stimulation of the MOC, which projects only to the OHCs (Fig. 12.5), modifies otoacoustic emissions [12.31] and reduces the amplitude of motion of the basilar membrane [12.102, 103].

Figure 12.11 shows tuning curves in AN fibers for intact ears ("normal") and for ears with IHC or OHC damage [12.51, 105]. These tuning curves are schematics that summarize the results of studies in ears with damage due to acoustic trauma or ototoxic antibiotics. Complete destruction of IHCs, of course, eliminates AN activity entirely. However, it is possible to find regions of the cochlea with surviving IHCs that have damaged stereocilia and OHCs that appear intact. The sensitivity of AN fibers in such regions is reduced, reflected in the elevated threshold in Fig. 12.11a, but the tuning is still sharp and appears to be minimally altered. By contrast, in regions with intact IHCs and OHCs that are damaged or missing, the tuning curves are dramatically altered, as in Fig. 12.11b. There is a threshold shift reflecting a loss of sensitivity but also a substantial broadening of tuning. These are the effects that are expected from the loss of the cochlear amplifier. Note that the thresholds well below BF (<1 kHz in Fig. 12.11b) can actually be lower than normal with OHC damage. This phenomenon is not fully understood and seems to reflect the existence of two modes of stimulation of IHCs [12.106].

Some properties of OHCs are shown in Fig. 12.12. These cells have a transduction apparatus similar to that in IHCs, but Fig. 12.12 focuses on unique aspects of OHC anatomy and physiology. The OHCs are located in an unusual system of structural or supporting cells. Unlike virtually all tissues, the OHC are surrounded by a fluid space containing perilymph for distances comparable to the cells' widths (asterisks in Fig. 12.4c). The OHC are held by supporting cells called Deiter's cells

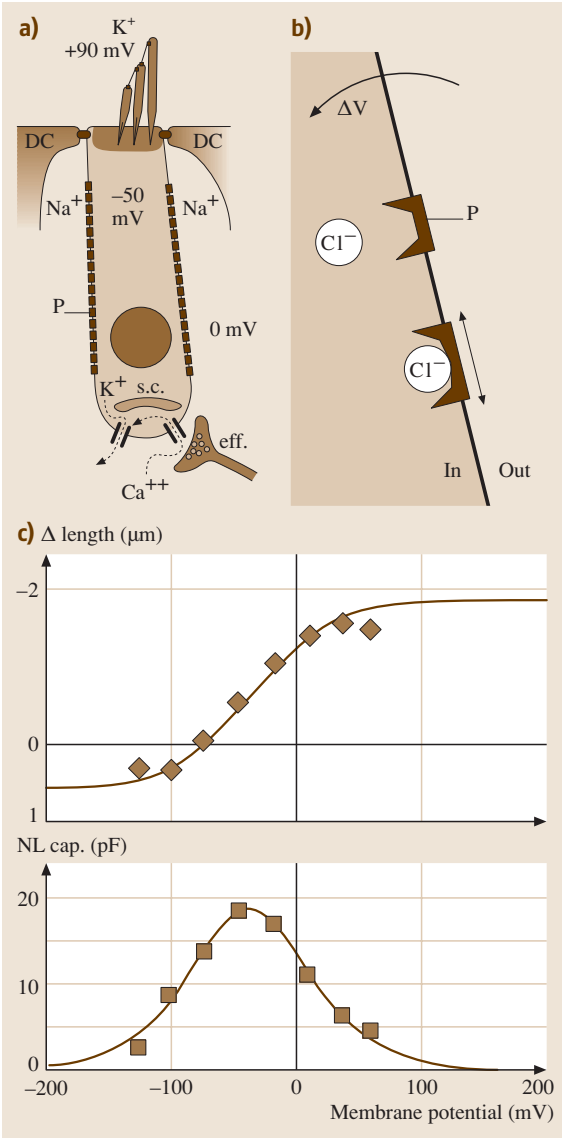


Fig. 12.12 **(a)** Schematic picture of an OHC showing elements important for OHC function. The diagram is labeled as for Fig. 12.9a except for the following: DC – Deiter's cell extensions that are linked to OHCs by tight junctions to form the top surface of the organ of Corti; Na^+ – the extracellular spaces around the lateral membranes of the OHCs are filled with perilymph, a high-Na^+, low-K^+ fluid; P – prestin molecules in the lateral membrane; eff. – efferent terminal of a MOC neuron on the OHC; s.c. subsynaptic cistern associated with the efferent synapse. K^+ and Ca^{2+} currents activated by the synapse are shown by dashed and dotted lines, respectively. **(b)** Schematic of prestin molecules in the lateral membrane of the cell. When the membrane potential ΔV is negative, Cl^- ions are driven into the prestin molecule, increasing its membrane area (as indicated by the *double-headed arrow*). **(c)** The *top plot* shows the length of an isolated OHC as a function of membrane potential showing the amplitude of the electromotility. The *bottom plot* shows the nonlinear capacitance of the same cell. (Fig. 12.12c after [12.104] with permission)

Amazon
EX1013 Page 459

that form a cup around the base of the cell and have a stiff extension up to the top surface of the organ of Corti. The extension terminates next to the apical surface of the OHC, where a system of tight junctions similar to that in the IHC region holds the OHCs and Deiter's cells to form a mechanically stiff and electrically insulating barrier between the endolymph and perilymph.

In response to depolarization of their membranes, say in response to a transduction current, OHCs shorten, a process called electromotility [12.107, 108]. Electromotility does not use a direct chemical energy source [such as adenosine triphosphate (ATP)] as do other cellular motility processes. It is also very fast, responding to frequencies of 10 kHz or above, limited mainly by the limits of the experimental observations. This unusual motile process is produced by a molecule called prestin ("P" in Fig. 12.12a) found in a dense array in the lateral wall of the OHCs [12.109–111], where it is associated with a complex mechanical matrix that forms the cell's skeleton [12.112]. Prestin causes the cell to shorten in response to electrical depolarization of its membrane, so the energy source for electromotility is the electrical membrane potential. The mechanism of electromotility is shown schematically in Fig. 12.12b. The prestin molecule is similar to anion transporters, molecules that normally move anions through the cell membrane. It is thought that prestin is a modified transporter that can bind an anion, but the ion moves only part of the way through the membrane. In OHCs, the relevant anion is Cl^- because electromotility is blocked by removing Cl^- from the cytoplasm [12.113]. When a Cl^- ion binds to a prestin molecule, the cross-sectional area of the molecule increases, thus increasing the membrane area and lengthening the cell (indicated by the double-headed arrow). Cl^- ions are driven into the membrane by negative membrane potentials, so depolarization pulls them out and decreases the membrane area, shortening the cell.

When a Cl^- ion binds to a prestin molecule, it moves part way through the cell's electrical membrane potential (ΔV in Fig. 12.12b). This movement behaves like charging a capacitance in the cell's membrane and appears in electrical recordings as an additional membrane capacitance. However the extent of charge movement depends on the membrane potential, making the capacitance nonlinear. At very negative membrane potentials, all the prestin molecules have a bound Cl^- and no charge movement can occur; similarly, at positive potentials, the Cl^- is unlikely to bind to prestin and again no charge movement can occur. Thus the nonlinear capacitance is significant only over the range of membrane potentials where the prestin molecules are partially bound to Cl^-. The bottom part of Fig. 12.12c shows the nonlinear capacitance as a function of membrane potential, showing the predicted behavior [12.104]. The top part of Fig. 12.12c shows the electrically produced change in cell length; as expected, changes in cell length are observed only over the range of membrane potentials where the nonlinear membrane capacitance is observed.

Prestin is found essentially only in OHCs and, in particular, is not seen in IHCs [12.111]. When expressed by genetic methods in cells that normally do not contain it, prestin confers electromotility on those cells. In addition, when the prestin gene was knocked out, the morphology of the cochlea and the hair cells was not affected (except that the OHCs were shorter), but electromotility was not observed and the sensitivity of auditory neurons was decreased; otoacoustic emissions were also decreased [12.114]. These are the effects expected if prestin is the energy source for the cochlear amplifier.

The means by which OHC motility amplifies basilar-membrane motion is somewhat uncertain. Models that use physiologically accurate OHC motility to produce cochlear amplification have been suggested and analyzed, e.g., [12.115]. However, such models require assumptions about the details of the mechanical interactions within the organ of Corti, assumptions that have not been tested experimentally. For example, the OHCs lie at an angle to the vertical extensions of the Deiter's cells, so that successive locations along the basilar membrane are coupled to one another [12.108]. The mechanical effects of structures like this are quite important in models of basilar-membrane movement that incorporate OHC electromotility [12.62]. An additional major uncertainty at present about a prestin-derived cochlear amplifier is the fact that prestin depends on the membrane potential as its driving signal [12.88]. Because the membrane potential is low-pass filtered by the cell's membrane capacitance, in the fashion seen in Fig. 12.10, it is not clear that a prestin-based mechanism can generate sufficient force at high frequencies.

The second possible mechanism for the cochlear amplifier is by stereocilia bundle movements that can amplify hair-cell stimulation either through nonlinearities in the compliance of the bundle or through active hair-bundle movements [12.89, 101]. Essentially these mechanisms work by moving the bundle further in the direction of its displacement, which would amplify the stimulation of the hair cell. Calculations suggest that the speed of bundle movement is sufficient and that the stiffness of the bundles is a significant fraction of the basilar-membrane stiffness, both necessary conditions for a stereociliary-bundle motor to be the cochlear ampli-

Amazon
EX1013 Page 460

fier. Of course both the prestin and stereociliary-bundle mechanisms may operate.

The final aspect of OHC physiology shown in Fig. 12.12a is the efferent synapse made by the MOC neuron on the OHC. This synapse activates an unusual cholinergic postsynaptic receptor which admits a significant Ca^{2+} current (dotted line), along with other cations, to the cytoplasm. The Ca^{2+} current activates a calcium-dependent potassium channel producing a larger potassium current (dashed line) [12.116]. The efferent synapse inhibits OHC function, decreasing the sensitivity of AN fibers and broadening their tuning, effects consistent with a decrease in cochlear amplification [12.31]. The mechanism is thought to be an increase in the membrane conductance of the OHC, from opening the calcium-dependent K^+ channel, which hyperpolarizes the cell and shorts the transducer current, producing a smaller receptor potential.

12.3 Auditory Nerve and Central Nervous System

In a previous section, the basic properties of AN fiber responses to acoustic stimuli were described. These properties can be related directly to the properties of the basilar membrane and hair cells. The auditory system does not usually deal with the kinds of simple stimuli that have been considered so far, i. e., tones of a fixed frequency. When multiple-tone complexes or other stimuli with multiple frequency components are presented to the ear, there are interactions among the frequency components that are important in shaping the responses. This section describes two such interactions and then provides an overview of the tasks performed by the remainder of the auditory system.

12.3.1 AN Responses to Complex Stimuli

Many of the features of responses of AN fibers to complex stimuli can be seen from the responses to a two-tone complex. An example is shown in Fig. 12.13 which shows responses of AN fibers to 2.17 and 2.79 kHz tones (called f_1 and f_2, respectively) presented simultaneously and separately [12.117]. In this experiment a large population of AN fibers were recorded in one animal and the same set of stimuli were presented to each fiber. Because fibers were recorded across a wide range of BFs, this approach allows the construction of an estimate of the response of the whole AN population. Responses are plotted in Fig. 12.13 as the strength of phase-locking to various frequency components of the stimulus. Phase-locking is used here to allow the complex responses of the fibers to be separated into responses to the various individual frequency components. The abscissa is plotted as BF, on a reversed frequency scale, and also in terms of the distance of the fibers' points of innervation from the stapes.

The distribution of responses to the f_1 frequency component (2.17 kHz) presented alone is shown in Fig. 12.13a by the solid line. The response peaks near the point of maximum response to f_1, indicated by the arrow labeled f_1 at the top of the plot. The dotted line shows the phase-locking to f_1 when the stimulus is a simultaneous presentation of f_1 and f_2. The response to f_1 is similar in both cases, except near the point of maximum response to f_2 (at the arrow marked f_2) where the response to f_1 is reduced by the addition of f_2 to the stimulus. This phenomenon is called two-tone suppression [12.118, 119]; it can be suppression of phase-locking as in this case or it can be a decrease of the discharge rate in response to an excitor tone when a suppressor tone is added. Suppression can be seen on the basilar membrane as a reduction in the membrane motion at one frequency caused by the addition of a second frequency [12.34], and is usually explained as resulting from compression in the basilar-membrane input/output relationship.

The importance of suppression for responses to complex stimuli is that it improves the tonotopic separation of the components of the stimulus. In Fig. 12.13a, for example, f_1 presented by itself (solid line) gives a broad response that spreads away from the f_1 place and encompasses the f_2 place. In the presence of f_2, the f_1 response is confined to points near the f_1 place, thus improving the effective *tuning* of the fibers for f_1. The bottom part of Fig. 12.13b shows that the responses at the f_2 place are dominated by phase-locking to f_2 when f_1 and f_2 are presented simultaneously (dashed curve). The suppression effect is symmetric and a dip in the response to f_2 is observed at the f_1 place (at the asterisk in the bottom plot of Fig. 12.13b) which represents suppression of f_2 by f_1.

The top part of Fig. 12.13b shows responses to combination tones, another phenomenon that can be important in responses to complex stimuli. When two tones are presented simultaneously, observers can hear

Amazon
EX1013 Page 461

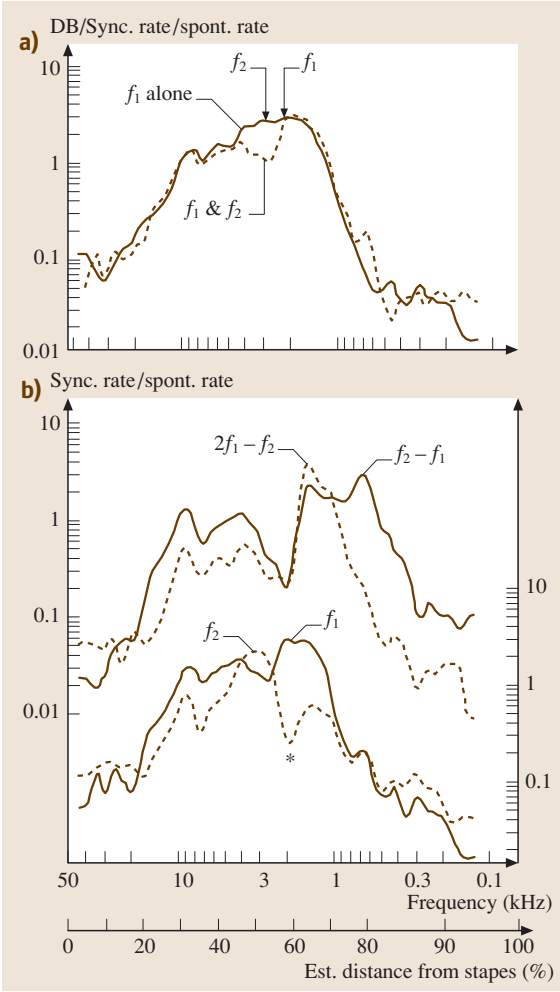


Fig. 12.13a,b Responses of a population of AN fibers to a two-tone complex consisting of f_1=2.17 kHz and f_2=2.79 kHz at 65 dB SPL. The *ordinates* show the strength of the phase-locked response as the synchronized rate divided by the spontaneous rate. Synchronized rate is the Fourier transform of the spike train at the appropriate frequency normalized to have units of spikes/s. Responses of a representative sample of fibers of different BFs were recorded. The lines are smoothed versions of the data computed as a moving-window average of the responses of individual fibers, for fibers with SR>15 /s. **(a)** Phase-locking to f_1 for responses to the f_1 stimulus tone alone (*solid line*) and to both f_1 and f_2 (*dashed line*). The arrows at top show the BF places for the two frequencies. **(b)** The *top plots* and *left ordinate* show phase-locking to the combination frequencies $2f_1 - f_2$ (*dashed*) and $f_2 - f_1$ (*solid*) when the stimulus was f_1 and f_2. The *bottom plots* and the *right ordinate* show phase-locking to f_1 (*solid*) and f_2 (*dashed*). For both plots, data were taken only from fibers to which all the stimuli shown were presented. Thus the f_1 phase-locking in **(a)** (*dashed line*) and in the bottom part of **(b)** (*solid line*) differ slightly, even though they estimate the same response to the same stimulus, because they were computed from somewhat different populations of fibers. (After [12.117] with permission)

distortion tones, most strongly at the cubic distortion frequency $2f_1 - f_2$ (1.55 kHz here). This distortion component is produced in the cochlea and is a prominent part of otoacoustic emissions for two-tone stimuli. It is used for both diagnostic and research purposes as a measure of OHC function [12.46]. In the cochlea, phase-locking at the frequency $2f_1 - f_2$ is seen, as shown by the dotted curve in the top part of Fig. 12.13b. It is important that the response to $2f_1 - f_2$ peaks at the place appropriate to the frequency $2f_1 - f_2$ (indicated by the arrow). The distribution of phase-locking to $2f_1 - f_2$ is similar to the distribution of phase-locking to a single tone at the frequency $2f_1 - f_2$, when account is taken of suppression effects and of sources of distortion in the region near the f_1 and f_2 places [12.117]. This behavior has been interpreted as showing that cochlear nonlinearities in the region of the primary tones (f_1 and f_2) lead to the creation of energy at the frequency $2f_1 - f_2$, which propagates on the basilar membrane in a fashion similar to an acoustic tone of that frequency presented at the ear. The phase of the responses (not shown) is consistent with this conclusion.

There is also a distortion tone at the frequency $f_2 - f_1$ (solid line in the top part of Fig. 12.13b), which appears to propagate on the basilar membrane and shows a peak of phase-locking at the appropriate place. This difference tone can also be heard by observers; however, there are differences in the rate of growth of the percept of $f_2 - f_1$ versus $2f_1 - f_2$, such that the former grows nonlinearly and the latter linearly as the level of f_1 and f_2 are increased. As a result, $f_2 - f_1$ is audible at high sound levels only.

Responses to a somewhat more complex stimulus are shown in Fig. 12.14, in this case an approximation to the vowel /eh/, as in "*met*" [12.65]. The magnitude spectrum of the stimulus is shown in Fig. 12.14a. It consists of the harmonics of a 125 Hz fundamental with their amplitudes adjusted to approximate the spectrum of /eh/. The actual vowel has peaks of energy at the resonant frequencies of the vocal tract, called formants; these are indicated by F1, F2, and F3 in Fig. 12.14a. Again the ex-

Amazon
EX1013 Page 462

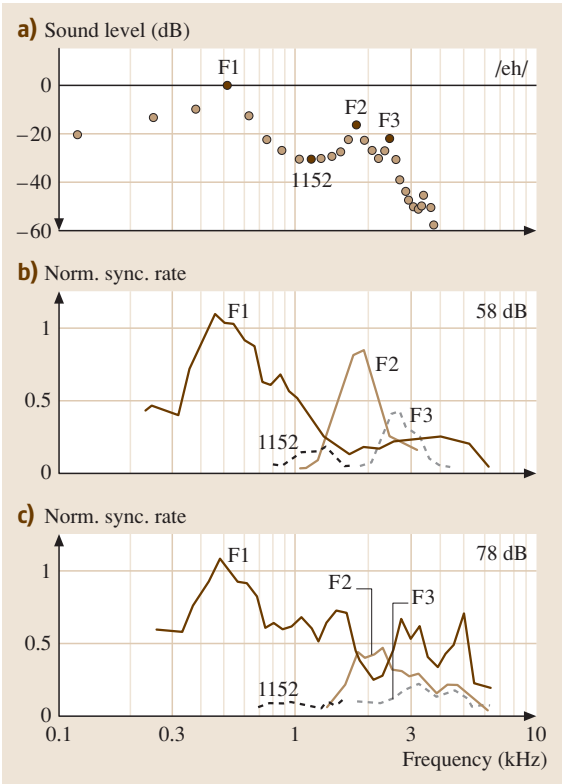


Fig. 12.14a–c Responses to a stimulus similar to the vowel /eh/ presented at two sound levels in a population of AN fibers. **(a)** The magnitude spectrum of the stimulus. It is periodic and consists of harmonics of a 125 Hz fundamental. There are peaks of energy at the formant frequencies of a typical /eh/, near 0.5 (F1), 1.75 (F2), and 2.275 kHz (F3). The 1.152 kHz component in the trough between F1 and F2 is used for comparison in parts **(b)** and **(c)**. **(b)** Distribution of phase-locking to four frequency components of the vowel when presented at 58 dB SPL, as labeled. These are moving-window averages of the phase-locking of individual fibers. Phase-locking here is the magnitude of the Fourier transform of the spike train at the frequency of interest normalized by the maximum discharge rate of the fiber. Maximum rate is the rate in response to a BF tone 50 dB above threshold. **(c)** Same for responses to the vowel at a sound level of 78 dB SPL. (After [12.120, 121])

perimental approach is to record from a large population of AN fibers in one animal and present the same stimulus to each fiber. The response of the whole population is estimated by a moving average of the data points for individual fibers. This stimulus contains many frequency components, so the responses are quite complex. However, the data show that formant frequencies dominate the responses. Note that combination tones always occur at the frequency of a real acoustic stimulus tone for a harmonic series like this, so it is not possible to analyze combination tones without making assumptions that allow the responses to combination and real tones to be separated [12.120].

The distribution of phase-locking in response to the formants and to one non-formant frequency are shown in Figs. 12.14b and 12.14c; these show responses at two sound levels. At 58 dB (Fig. 12.14b), the formant responses are clearly separated into different populations of fibers, so that the response to a formant is largest among fibers with BFs near the formant frequency. There are significant responses to all three formants, but there is little response to frequency components between the formants, such as the response to 1.152 kHz (dotted line). These results demonstrate a clear tonotopic representation of the vowel.

At 78 dB (Fig. 12.14c), the responses to F2 and F3 decrease in amplitude and the response to F1 spreads to occupy almost the entire population. This behavior is typically observed at high sound levels, where the phase-locking to a large low-frequency component of the stimulus spreads to higher BFs in the population [12.122]. Suppression plays two roles in this process. First, the spread of F1 and the decrease in response to F2 and F3 behave quantitatively like suppression of a tone at F2 or F3 by a tone at F1 and therefore seem to represent suppression of F2 and F3 by F1 [12.121]. Second, the response to F1 is suppressed among fibers with BFs near F2, as shown by dips in the phase-locking to F1 near the F2 place. Suppression acts to improve the representation in the latter case, but has the opposite effect in the former case.

12.3.2 Tasks of the Central Auditory System

The discussion so far concerns only the most basic aspects of physiological acoustics. The representation of sound in the AN is the input to a complex of neural pathways in the central auditory system. Discussion of specifics of the central processing of auditory stimuli is beyond the scope of this chapter. However, several recent books provide comprehensive coverage of the anatomy and physiology of this system, e.g., [12.123–126].

The representation of the auditory stimulus provided to the brain by the AN is a spectrotemporal one. That is, the responses of AN fibers provide an accurate representation of the moment-by-moment distribution of energy across frequency for the sound entering the ear.

Amazon
EX1013 Page 463

In terms of processing sound, the major steps taken in the cochlea are frequency analysis, compression, and suppression. Frequency analysis is essential to all of the processing that follows in the brain. Indeed central auditory centers are all organized by frequency and most properties of the perception of sound appear to include frequency analysis as a fundamental component. Compression is important for extending the dynamic range of hearing. Its importance is illustrated by the problems of loudness abnormality and oversensitivity to loud sounds in persons with damaged cochleae that are missing compression [12.127]. Suppression is important for maintaining the quality of the spectrotemporal representation, at least at moderate sound levels, and is the first example of interaction across frequencies in the auditory system. In the following paragraphs, some of the central nervous system's secondary analyses on the cochlear spectrotemporal representation are described.

The first function of the central auditory system is to stabilize the spectrotemporal representation provided in the AN. As shown in Fig. 12.14, the representation changes across sound level and the quality of the representation is generally lower at high sound levels and at low sound levels. In the cochlear nucleus the representation is more stable as sound level changes, e.g., for speech [12.128]. Presumably the stabilization occurs through optimal combination of information across SR groups in the AN and perhaps also through inhibitory interactions.

A second function of the lower central auditory system is binaural interaction. The location of sound sources is computed by the auditory system from small differences in the sounds at the two ears [12.84, 129]. Essentially, a sound is delayed in reaching the ear on the side of the head away from the source and is less intense there. Differences in interaural stimulus timing are small and require a specialized system of neurons in the first and second stages of the central auditory system to allow accurate representation of the interaural timing cues. A similar system is present for interaural differences in sound level. The representation of these cues is elaborated in higher levels, where neurons show response characteristics that might explain perceptual phenomena like binaural masking level differences [12.130] and the precedence effect [12.131].

A third aspect of central processing is that the representation of sound switches away from a spectrotemporal description of the stimulus, as in the AN, and moves to derived representations, such as one that represents auditory objects [12.132]. For example, the representation of speech in the auditory cortex does not take a recognizable tonotopic form, in that peaks of activity at BFs corresponding to the formant frequencies are not seen [12.133]. The nature of the representation used in the cortex is unknown; it is clearly based on a tonotopic axis, but neurons' responses are not determined in a straightforward way by the amount of energy in their tuning curves, as is observed in lower auditory nuclei [12.134]. In some cases, neurons specialized for particular auditory tasks have been found; neurons in one region of the marmoset cortex are tuned to pitch, for example [12.135]. Perhaps the best-studied cases are neurons that respond specifically to species-specific vocalizations in marmosets [12.136] and songbirds [12.137, 138].

12.4 Summary

Research on the auditory system has defined the means by which the system efficiently captures sound and couples it into the cochlea, how the frequency analysis of the cochlea is conducted, how transduction occurs, how nonlinear mechanisms involving the OHCs sharpen the frequency tuning, increase the sensitivity, and compress the stimulus intensity, and how suppression acting at the level of AN fibers maintains the sharpness of the frequency analysis for complex, multicomponent stimuli. Over the next decade, research on the auditory system will increasingly be focused on the organization and function of the central auditory system, with particular reference to the way in which central neurons derive information from the spectrotemporal display received from the AN.

Amazon
EX1013 Page 464

References

12.1 G.A. Manley, A.N. Popper, R.R. Fay: *Evolution of the Vertebrate Auditory System* (Springer, Berlin, New York 2004)

12.2 J.J. Rosowski: Outer and middle ears. In: *Comparative Hearing: Mammals*, ed. by R.R. Fay, A.N. Popper (Springer, Berlin, New York 1994)

12.3 J.J. Rosowski: External- and middle-ear function. In: *Auditory Computation*, ed. by H.L. Hawkins, T.A. McMullen, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1996)

12.4 E.A.G. Shaw: Transformation of sound pressure level from the free field to the eardrum in the horizontal plane, J. Acoust. Soc. Am. **56**, 1848–1861 (1974)

12.5 F.M. Wiener, R.R. Pfeiffer, A.S.N. Backus: On the sound pressure transformation by the head and auditory meatus of the cat, Acta Otolaryngol. **61**, 255–269 (1966)

12.6 J.J. Rice, B.J. May, G.A. Spiron, E.D. Young: Pinna-based spectral cues for sound localization in cat, Hearing Res. **58**, 132–152 (1992)

12.7 E.A.G. Shaw: External ear response and sound localization. In: *Localization of Sound: Theory and Applications*, ed. by R.W. Gatehouse (Amphora, Groton 1982)

12.8 M. Brodel: *Three Unpublished Drawings of the Anatomy of the Human Ear* (Saunders, Philadelphia 1946)

12.9 E.A.G. Shaw: 1979 Rayleigh medal lecture: The elusive connection. In: *Localization of Sound: Theory and Applications*, ed. by R.W. Gatehouse (Amphora, Groton 1982)

12.10 J.J. Rosowski, L.H. Carney, W.T. Peake: The radiation impedance of the external ear of cat: measurements and applications, J. Acoust. Soc. Am. **84**, 1695–1708 (1988)

12.11 J.C. Middlebrooks, D.M. Green: Sound localization by human listeners, Annu. Rev. Psychol. **42**, 135–159 (1991)

12.12 B.J. May, A.Y. Huang: Sound orientation behavior in cats. I. Localization of broadband noise, J. Acoust. Soc. Am. **100**, 1059–1069 (1996)

12.13 J. Blauert: *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge 1997)

12.14 G.F. Kuhn: Physical acoustics and measurements pertaining to directional hearing. In: *Directional Hearing*, ed. by W.A. Yost, G. Gourevitch (Springer, Berlin, New York 1987)

12.15 E.D. Young, J.J. Rice, S.C. Tong: Effects of pinna position on head-related transfer functions in the cat, J. Acoust. Soc. Am. **99**, 3064–3076 (1996)

12.16 A.D. Musicant, J.C.K. Chan, J.E. Hind: Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons, J. Acoust. Soc. Am. **87**, 757–781 (1990)

12.17 A.Y. Huang, B.J. May: Spectral cues for sound localization in cats: Effects of frequency domain on minimum audible angles in the median and horizontal planes, J. Acoust. Soc. Am. **100**, 2341–2348 (1996)

12.18 P. Dallos: *The Auditory Periphery, Biophysics and Physiology* (Academic, New York 1973)

12.19 J. Zwislocki: Analysis of some auditory characteristics. In: *Handbook of Mathematical Psychology*, ed. by R.D. Luce, R.R. Bush, E. Galanter (Wiley, New York 1965)

12.20 W. Bialek: Physical limits to sensation and perception, Ann. Rev. Biophys. Biophys. Chem. **16**, 455–478 (1987)

12.21 M.B. Sachs, P.J. Abbas: Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli, J. Acoust. Soc. Am. **56**, 1835–1847 (1974)

12.22 P. Wangemann, J. Schacht: Homeostatic mechanisms in the cochlea. In: *The Cochlea*, ed. by P. Dallos, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1996)

12.23 W. Bloom, D.W. Fawcett: *A Textbook of Histology* (Saunders, Philadelphia 1975)

12.24 D.K. Ryugo: The auditory nerve: Peripheral innervation, cell body morphology, and central projections. In: *The Mammalian Auditory Pathway: Neuroanatomy*, ed. by D.B. Webster, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1992)

12.25 W.B. Warr: Organization of olivocochlear efferent systems in mammals. In: *The Mammalian Auditory Pathway: Neuroanatomy*, ed. by D.B. Webster, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1992)

12.26 M.C. Brown: Antidromic responses of single units from the spiral ganglion, J. Neurophysiol. **71**, 1835–1847 (1994)

12.27 D. Robertson: Horseradish peroxidase injection of physiologically characterized afferent and efferent neurons in the guinea pig spiral ganglion, Hearing Res. **15**, 113–121 (1984)

12.28 D. Robertson, P.M. Sellick, R. Patuzzi: The continuing search for outer hair cell afferents in the guinea pig spiral ganglion, Hearing Res. **136**, 151–158 (1999)

12.29 D.J. Jagger, G.D. Housley: Membrane properties of type II spiral ganglion neurones identified in a neonatal rat cochlear slice, J. Physiol. **552**, 525–533 (2003)

12.30 M.A. Reid, J. Flores-Otero, R.L. Davis: Firing patterns of type II spiral ganglion neurons in vitro, J. Neurosci. **24**, 733–742 (2004)

12.31 J.J. Guinan: Physiology of olivocochlear efferents. In: *The Cochlea*, ed. by P. Dallos, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1996)

Amazon
EX1013 Page 465

12.32 M.B. Sachs, I.C. Bruce, E.D. Young: The biological basis of hearing-aid design, Ann. BME **30**, 157–168 (2002)

12.33 G. von Bekesy: *Experiments in Hearing* (McGraw-Hill, New York 1960)

12.34 L. Robles, M.A. Ruggero: Mechanics of the mammalian cochlea, Physiol. Rev. **81**, 1305–1352 (2001)

12.35 D.H. Eldredge: Electrical equivalents of the Bekesy traveling wave in the mammalian cochlea. In: *Evoked Electrical Activity in the Auditory Nervous System*, ed. by R.F. Naunton, C. Fernandez (Academic, New York 1978)

12.36 M.A. Ruggero, N.C. Rich, A. Recio, S.S. Narayan: Basilar-membrane responses to tones at the base of the chinchilla cochlea, J. Acoust. Soc. Am. **101**, 2151–2163 (1997)

12.37 N.P. Cooper, W.S. Rhode: Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea, J. Neurophysiol. **78**, 261–270 (1997)

12.38 W.S. Rhode: Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique, J. Acoust. Soc. Am. **49**, 1218–1231 (1971)

12.39 D. Brass, D.T. Kemp: Analyses of Mossbauer mechanical measurements indicate that the cochlea is mechanically active, J. Acoust. Soc. Am. **93**, 1502–1515 (1993)

12.40 P. Dallos: The active cochlea, J. Neurosci. **12**, 4575–4585 (1992)

12.41 H. Davis: An active process in cochlear mechanics, Hear Res. **9**, 79–90 (1983)

12.42 R.J. Diependaal, E. de Boer, M.A. Viergever: Cochlear power flux as an indicator of mechanical activity, J. Acoust. Soc. Am. **82**, 917–926 (1987)

12.43 G.A. Manley: Evidence for an active process and a cochlear amplifier in nonmammals, J. Neurophysiol. **86**, 541–549 (2001)

12.44 E. Zwicker: A model describing nonlinearities in hearing by active processes with saturation at 40 dB, Biol. Cybern. **35**, 243–250 (1979)

12.45 D.T. Kemp: Stimulated acoustic emissions from within the human auditory system, J. Acoust. Soc. Am. **64**, 1386–1391 (1978)

12.46 C.A. Shera: Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions, Ear Hearing **25**, 86–97 (2004)

12.47 C.L. Talmadge, A. Tubis, G.R. Long, C. Tong: Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure, J. Acoust. Soc. Am. **108**, 2911–2932 (2000)

12.48 G. Zweig, C.A. Shera: The origin of periodicity in the spectrum of evoked otoacoustic emissions, J. Acoust. Soc. Am. **98**, 2018–2047 (1995)

12.49 R. Patuzzi: Cochlear micromechanics and macromechanics. In: *The Cochlea*, ed. by P. Dallos, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1996)

12.50 E. de Boer, A.L. Nuttall, N. Hu, Y. Zou, J. Zheng: The Allen-Fahey experiment extended, J. Acoust. Soc. Am. **117**, 1260–1266 (2005)

12.51 M.C. Liberman, L.W. Dodds: Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves, Hearing Res. **16**, 55–74 (1984)

12.52 E. Murugasu, I.J. Russell: The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea, J. Neurosci. **16**, 325–332 (1996)

12.53 M.C. Brown, A.L. Nuttall: Efferent control of cochlear inner hair cell responses in the guinea-pig, J. Physiol. **354**, 625–646 (1984)

12.54 M.L. Wiederhold, N.Y.S. Kiang: Effects of electric stimulation of the crossed olivocochlear bundle on single auditory-nerve fibers in the cat, J. Acoust. Soc. Am. **48**, 950–965 (1970)

12.55 M.C. Liberman, S. Puria, J.J. Guinan: The ipsilaterally evoked olivocochlear reflex causes rapid adaptation of the $2f_1-f_2$ distortion product otoacoustic emission, J. Acoust. Soc. Am. **99**, 3572–3584 (1996)

12.56 D.C. Mountain: Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics, Science **210**, 71–72 (1980)

12.57 M.A. Ruggero, N.C. Rich, A. Recio: The effect of intense acoustic stimulation on basilar-membrane vibrations, Aud. Neurosci. **2**, 329–345 (1996)

12.58 A.J. Oxenham, S.P. Bacon: Cochlear compression: Perceptual measures and implications for normal and impaired hearing, Ear Hearing **24**, 352–366 (2003)

12.59 D.A. Nelson, A.C. Schroder, M. Wojtczak: A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners, J. Acoust. Soc. Am. **110**, 2045–2064 (2001)

12.60 A.J. Oxenham, C.J. Plack: A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing, J. Acoust. Soc. Am. **101**, 3666–3675 (1997)

12.61 E. de Boer: Mechanics of the cochlea: Modeling efforts. In: *The Cochlea*, ed. by P. Dallos, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1996)

12.62 A.E. Hubbard, D.C. Mountain: Analysis and synthesis of cochlear mechanical function using models. In: *Auditory Computation*, ed. by H.L. Hawkins, T.A. McMullen, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1996)

12.63 R.C. Naidu, D.C. Mountain: Measurements of the stiffness map challenge a basic tenet of cochlear theories, Hearing Res. **124**, 124–131 (1998)

12.64 M.P. Sherer, A.W. Gummer: Impedance analysis of the organ of corti with magnetically actuated probes, Biophys. J. **87**, 1378–1391 (2004)

Amazon
EX1013 Page 466

12.65 M.B. Sachs: Speech encoding in the auditory nerve. In: *Hearing Science Recent Advances*, ed. by C.I. Berlin (College Hill, San Diego 1984)

12.66 J.J. Eggermont: Between sound and perception: reviewing the search for a neural code, Hear Res. **157**, 1–42 (2001)

12.67 B.C.J. Moore: Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants, Otol. Neurotol. **24**, 243–254 (2003)

12.68 I.C. Bruce, M.B. Sachs, E.D. Young: An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses, J. Acoust. Soc. Am. **113**, 369–388 (2003)

12.69 S.D. Holmes, C. Sumner, L.P. O'Mard, R. Meddis: The temporal representation of speech in a nonlinear model of the guinea pig cochlea, J. Acoust. Soc. Am. **116**, 3534–3545 (2004)

12.70 X. Zhang, M.G. Heinz, I.C. Bruce, L.H. Carney: A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression, J. Acoust. Soc. Am. **109**, 648–670 (2001)

12.71 H.S. Colburn, L.H. Carney, M.G. Heinz: Quantifying the information in auditory-nerve responses for level discrimination, J. Assoc. Res. Otolaryngol. **4**, 294–311 (2003)

12.72 B.C.J. Moore: *An Introduction to the Psychology of Hearing* (Elsevier, London 2004)

12.73 P.A. Cariani, B. Delgutte: Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch, J. Neurophysiol. **76**, 1717–1734 (1996)

12.74 R.L. Miller, J.R. Schilling, K.R. Franck, E.D. Young: Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers, J. Acoust. Soc. Am. **101**, 3602–3616 (1997)

12.75 D.H. Johnson: The relationship of post-stimulus time and interval histograms to the timing characteristics of spike trains, Biophys. J. **22**(3), 413–430 (1978)

12.76 M.C. Liberman: Auditory-nerve response from cats raised in a low-noise chamber, J. Acoust. Soc. Am. **63**, 442–455 (1978)

12.77 G.K. Yates, I.M. Winter, D. Robertson: Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range, Hearing Res. **45**, 203–220 (1990)

12.78 B.C.J. Moore: *Perceptual Consequences of Cochlear Damage* (Oxford Univ. Press, Oxford 1995)

12.79 M.G. Heinz, J.B. Issa, E.D. Young: Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment, J. Assoc. Res. Otolaryngol. **6**, 91–105 (2005)

12.80 J.O. Pickles: Auditory-nerve correlates of loudness summation with stimulus bandwidth, in normal and pathological cochleae, Hearing Res. **12**, 239–250 (1983)

12.81 E.M. Relkin, J.R. Doucet: Is loudness simply proportional to the auditory nerve spike count?, J. Acoust. Soc. Am. **101**, 2735–2740 (1997)

12.82 S. Rosen: Temporal information in speech: acoustic, auditory and linguistic aspects, Philos. Trans. R. Soc. London B. **336**, 367–373 (1992)

12.83 A.R. Palmer, I.J. Russell: Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells, Hear Res. **24**, 1–15 (1986)

12.84 T.C.T. Yin: Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In: *Integrative Functions in the Mammalian Auditory Pathway*, ed. by D. Oertel, A.N. Popper, R.R. Fay (Springer, Berlin, New York 2002)

12.85 S.A. Shamma: Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve, J. Acoust. Soc. Am. **78**, 1622–1632 (1985)

12.86 P. Dallos, B. Fakler: Prestin, a new type of motor protein, Nature Rev. Molec. Bio. **3**, 104–111 (2002)

12.87 R.A. Eatock, K.M. Hurley: Functional development of hair cells, Curr. Top. Dev. Bio. **57**, 389–448 (2003)

12.88 R. Fettiplace, C.M. Hackney: The sensory and motor roles of auditory hair cells, Nature Rev. Neurosci. **7**, 19–29 (2006)

12.89 M. LeMasurier, P.G. Gillespie: Hair-cell mechanotransduction and cochlear amplification, Neuron. **48**, 403–415 (2005)

12.90 G.I. Frolenkov, I.A. Belyantseva, T.B. Friedmann, A.J. Griffith: Genetic insights into the morphogenesis of inner ear hair cells, Nature Rev. Genet. **5**, 489–498 (2004)

12.91 J.O. Pickles, S.D. Comis, M.P. Osborne: Cross-links between stereocilia in the guinea pig organ of Corti, and their possible relation to sensory transduction, Hear Res. **15**, 103–112 (1984)

12.92 A. Flock: Transducing mechanisms in the lateral line canal organ receptors, Cold Spring Harb. Symp. Quant. Biol. **30**, 133–145 (1965)

12.93 S.L. Shotwell, R. Jacobs, A.J. Hudspeth: Directional sensitivity of individual vertebrate hair cells to controlled deflection of their hair bundles, Ann. NY Acad. Sci. **374**, 1–10 (1981)

12.94 J.A. Assad, G.M.G. Shepherd, D.P. Corey: Tip-link integrity and mechanical transduction in vertebrate hair cells, Neuron **7**, 985–994 (1991)

12.95 M.A. Ruggero, N.C. Rich: Timing of spike initiation in cochlear afferents: dependence on site of innervation, J. Neurophysiol. **58**, 379–403 (1987)

12.96 E. Glowatzki, P.A. Fuchs: Transmitter release at the hair cell ribbon synapse, Nature Neurosci. **5**, 147–154 (2002)

12.97 L.O. Trussell: Transmission at the hair cell synapse, Nature Neurosci. **5**, 85–86 (2002)

Amazon
EX1013 Page 467

12.98 Y. Raphael, R.A. Altschuler: Structure and innervation of the cochlea, Brain Res. Bull. **60**, 397–422 (2003)
12.99 P.G. Gillespie, J.L. Cyr: Myosin-1c, the hair cell's adaptation motor, Annu. Rev. Physiol. **66**, 521–545 (2004)
12.100 M.C. Holley: Outer hair cell motility. In: *The Cochlea*, ed. by P. Dallos, A.N. Popper, R.R. Fay (Springer, Berlin, New York 1996)
12.101 R. Fettiplace: Active hair bundle movements in auditory hair cells, J. Physiol. **576**, 29–36 (2006)
12.102 M.C. Brown, A.L. Nuttall, R.I. Masta: Intracellular recordings from cochlear inner hair cells: effects of stimulation of the crossed olivocochlear efferents, Science **222**, 69–72 (1983)
12.103 D.F. Dolan, M.H. Guo, A.L. Nuttall: Frequency-dependent enhancement of basilar membrane velocity during olivocochlear bundle stimulation, J. Acoust. Soc. Am. **102**, 3587–3596 (1997)
12.104 J. Santos-Sacchi: Reversible inhibition of voltage-dependent outer hair cell motility and capacitance, J. Neurosci. **11**, 3096–3110 (1991)
12.105 P. Dallos, D. Harris: Properties of auditory nerve responses in absence of outer hair cells, J. Neurophysiol. **41**, 365–383 (1978)
12.106 M.C. Liberman, N.Y. Kiang: Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions, Hearing Res. **16**, 75–90 (1984)
12.107 J.F. Ashmore: A fast motile response in guinea pig outer hair cells: the basis of the cochlear amplifier, J. Physiol. **388**, 323–347 (1987)
12.108 W.E. Brownell, C.R. Bader, D. Bertrand, Y. de Ribaupierre: Evoked mechanical responses of isolated cochlear hair cells, Science **227**, 194–196 (1985)
12.109 I.A. Belyantseva, H.J. Adler, R. Curi, G.I. Frolenkov, B. Kachar: Expression and localization of prestin and the sugar transporter GLUT-5 during development of electromotility in cochlear outer hair cell, J. Neurosci. **20**(RC116), 1–5 (2000)
12.110 F. Kalinec, M.C. Holley, K.H. Iwasa, D.J. Lim, B. Kachar: A membrane-based force generation mechanism in auditory sensory cells, Proc. Natl. Acad. Sci. USA **89**, 8671–8675 (1992)
12.111 J. Zheng, W. Shen, D.Z. He, K.B. Long, D. Madison, P. Dallos: Prestin is the motor protein of cochlear outer hair cells, Nature **405**, 149–155 (2000)
12.112 M.C. Holley, F. Kalinec, B. Kachar: Structure of the cortical cytoskeleton in mammalian outer hair cells, J. Cell Sci. **102**, 569–580 (1992)
12.113 D. Oliver, D.Z. He, N. Klocker, J. Ludwig, U. Schutte, S. Waldegger, J.P. Ruppersberg, P. Dallos, B. Fakler: Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein, Science **292**, 2340–2343 (2001)
12.114 M.C. Liberman, J. Galilo, D.Z.Z. He, X. Wu, S. Jia, J. Zuo: Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier, Nature **419**, 300–304 (2002)
12.115 C.D. Geisler, C. Sang: A cochlear model using feed-forward outer-hair-cell forces, Hear Res. **86**, 132–146 (1995)
12.116 P.A. Fuchs: Synaptic transmission at vertebrate hair cells, Curr. Opin. Neurobiol. **6**, 514–519 (1996)
12.117 D.O. Kim, C.E. Molnar, J.W. Matthews: Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure, J. Acoust. Soc. Am. **67**, 1704–1721 (1980)
12.118 E. Javel, C.D. Geisler, A. Ravindran: Two-tone suppression in auditory nerve of the cat: rate-intensity and temporal analyses, J. Acoust. Soc. Am. **63**(4), 1093–1104 (1978)
12.119 M.B. Sachs, N.Y. Kiang: Two-tone inhibition in auditory-nerve fibers, J. Acoust. Soc. Am. **43**(5), 1120–1128 (1968)
12.120 E.D. Young, M.B. Sachs: Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers, J. Acoust. Soc. Am. **66**, 1381–1403 (1979)
12.121 M.B. Sachs, E.D. Young: Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate, J. Acoust. Soc. Am. **66**, 470–479 (1979)
12.122 J.C. Wong, R.L. Miller, B.M. Callhoun, M.B. Sachs, E.D. Young: Effects of high sound levels on responses to the vowel /ɛ/ in cat auditory nerve, Hear Res. **123**, 61–77 (1998)
12.123 G. Ehret, R. Romand: *The Central Auditory System* (Oxford Univ. Press, New York 1997)
12.124 M.S. Malmierca, D.R.F. Irvine: *International Review of Neurobiology – Auditory Spectral Processing*, Vol. 70 (Elsevier, Amsterdam 2005)
12.125 D. Oertel, A.N. Popper, R.R. Fay: *Integrative Functions in the Mammalian Auditory Pathway* (Springer, Berlin, New York 2001)
12.126 J.A. Winer, C.E. Schreiner: *The Inferior Colliculus* (Springer, Berlin, New York 2005)
12.127 S.P. Bacon, A.J. Oxenham: Psychophysical manifestations of compression: Hearing-impaired listeners. In: *Compression: From cochlea to cochlear implants*, ed. by S.P. Bacon, R.R. Fay, A.N. Popper (Springer, Berlin, New York 2004)
12.128 C.C. Blackburn, M.B. Sachs: The representations of the steady-state vowel sound /ɛ/ in the discharge patterns of cat anteroventral cochlear nucleus neurons, J. Neurophysiol. **63**, 1191–1212 (1990)
12.129 A.R. Palmer: Reassessing mechanisms of low-frequency sound localisation, Curr. Opin. Neurobiol. **14**, 457–460 (2004)
12.130 A.R. Palmer, D. Jiang, D. McAlpine: Neural responses in the inferior colliculus to binaural masking level differences created by inverting the

Amazon
EX1013 Page 468

noise in one ear, J. Neurophysiol. **84**, 844–852 (2000)
12.131 T.C. Yin: Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat, J. Neurosci. **14**, 5170–5186 (1994)
12.132 I. Nelken: Processing of complex stimuli and natural scenes in the auditory cortex, Curr. Opin. Neurobiol. **14**, 474–480 (2004)
12.133 S.W. Wong, C.E. Schreiner: Representation of CV-sounds in cat primary auditory cortex: Intensity dependence, Speech Commun. **41**, 93–106 (2003)
12.134 C.K. Machens, M.S. Wehr, A.M. Zador: Linearity of cortical receptive fields measured with natural sounds, J. Neurosci. **24**, 1089–1100 (2004)
12.135 D. Bendor, X. Wang: The neuronal representation of pitch in primate auditory cortex, Nature **436**, 1161–1165 (2005)
12.136 X. Wang, S.C. Kadia: Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat, J. Neurophysiol. **86**, 2616–2620 (2001)
12.137 H. Cousillas, H.J. Leppelsack, E. Leppelsack, J.P. Richard, M. Mathelier, M. Hausberger: Functional organization of the forebrain auditory centres of the European starling: A study based on natural sounds, Hear Res. **207**, 10–21 (2005)
12.138 J.A. Grace, N. Amin, N.C. Singh, F.E. Theunissen: Selectivity for conspecific song in the zebra finch auditory forebrain, J. Neurophysiol. **89**, 472–487 (2003)

Amazon
EX1013 Page 469

13. Psychoacoustics

Psychoacoustics is concerned with the relationships between the physical characteristics of sounds and their perceptual attributes. This chapter describes: the absolute sensitivity of the auditory system for detecting weak sounds and how that sensitivity varies with frequency; the frequency selectivity of the auditory system (the ability to resolve or *hear out* the sinusoidal components in a complex sound) and its characterization in terms of an array of auditory filters; the processes that influence the masking of one sound by another; the range of sound levels that can be processed by the auditory system; the perception and modeling of loudness; level discrimination; the temporal resolution of the auditory system (the ability to detect changes over time); the perception and modeling of pitch for pure and complex tones; the perception of timbre for steady and time-varying sounds; the perception of space and sound localization; and the mechanisms underlying *auditory scene analysis* that allow the construction of percepts corresponding to individual sounds sources when listening to complex mixtures of sounds.

13.1 **Absolute Thresholds** 460

13.2 **Frequency Selectivity and Masking** 461
13.2.1 The Concept of the Auditory Filter 462
13.2.2 Psychophysical Tuning Curves 462
13.2.3 The Notched-Noise Method 463
13.2.4 Masking Patterns and Excitation Patterns 464
13.2.5 Forward Masking 465
13.2.6 Hearing Out Partials in Complex Tones 467

13.3 **Loudness** 468
13.3.1 Loudness Level and Equal-Loudness Contours 468
13.3.2 The Scaling of Loudness 469
13.3.3 Neural Coding and Modeling of Loudness 469
13.3.4 The Effect of Bandwidth on Loudness 470
13.3.5 Intensity Discrimination 472

13.4 **Temporal Processing in the Auditory System** 473
13.4.1 Temporal Resolution Based on Within-Channel Processes 473
13.4.2 Modeling Temporal Resolution 474
13.4.3 A Modulation Filter Bank? 475
13.4.4 Duration Discrimination 476
13.4.5 Temporal Analysis Based on Across-Channel Processes 476

13.5 **Pitch Perception** 477
13.5.1 Theories of Pitch Perception 477
13.5.2 The Perception of the Pitch of Pure Tones 478
13.5.3 The Perception of the Pitch of Complex Tones 480

13.6 **Timbre Perception** 483
13.6.1 Time-Invariant Patterns and Timbre 483
13.6.2 Time-Varying Patterns and Auditory Object Identification 483

13.7 **The Localization of Sounds** 484
13.7.1 Binaural Cues 484
13.7.2 The Role of the Pinna and Torso 485
13.7.3 The Precedence Effect 485

13.8 **Auditory Scene Analysis** 485
13.8.1 Information Used to Separate Auditory Objects 486
13.8.2 The Perception of Sequences of Sounds 489
13.8.3 General Principles of Perceptual Organization 492

13.9 **Further Reading and Supplementary Materials** 494

References 495

Amazon
EX1013 Page 470

13.1 Absolute Thresholds

The absolute threshold of a sound is the lowest detectable level of that sound in the absence of any other sounds. In practice, there is no distinct sound level at which a sound suddenly becomes detectable. Rather, the probability of detecting a sound increases progressively as the sound level is increased from a very low value. Hence, the absolute threshold is defined as the sound level at which an individual detects the sound with a certain probability, such as 75% (in a two-alternative forced-choice task, where guessing leads to 50% correct, on average). Typically, results are averaged across many listeners with normal hearing (i. e., with no known history of hearing disorders and no obvious signs of hearing problems) to obtain representative results.

The absolute threshold for detecting sinusoids is partly determined by the sound transmission through the outer and middle ear (see Chap. 12); to a first approximation, the inner ear (the cochlea) is equally sensitive to all frequencies, except perhaps at very low frequencies and very high frequencies [13.3, 4]. Figure 13.1 shows estimates of the absolute threshold, measured in two ways. For the curve labeled MAP, standing for minimum audible pressure, the sound level was measured at a point close to the eardrum [13.2]. For the curve labeled MAF, standing for minimum audible field, the measurement of sound level was made after the listener had been removed from the sound field, at the point which had been occupied by the center of the listener's head [13.1]. For the MAF curve, the sound was presented in free field (in an anechoic chamber) from a direction directly in front of the listener. Note that the MAP estimates are for monaural listening and the MAF estimates are for binaural listening. On average, thresholds are about 2 dB lower when two ears are used as opposed to one, although the exact value of the difference varies across studies from 0 to 3 dB, and it can depend on the interaural phase of the tone [13.5–7]. Both curves represent the average data from many young listeners with normal hearing. It should be noted, however, that individual people may have thresholds as much as 20 dB above or below the mean at a specific frequency and still be considered as normal.

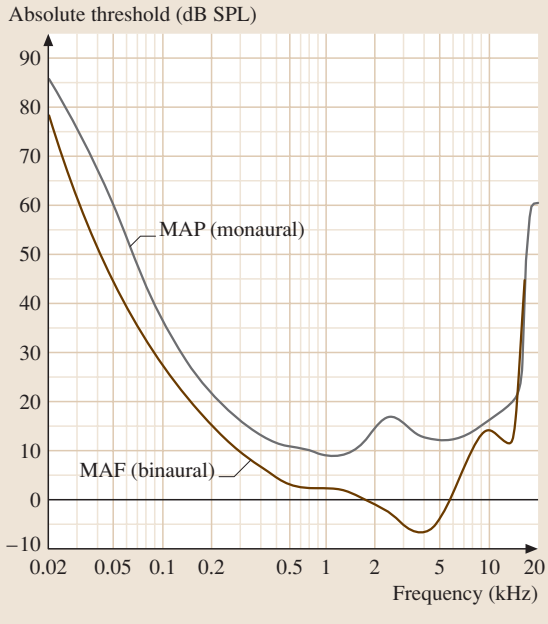


Fig. 13.1 The minimum audible sound level as a function of frequency. The *solid curve* shows the minimum audible field (MAF) for binaural listening published in an International Standards Organization (ISO) standard [13.1]. The *dashed curve* shows the minimum audible pressure (MAP) for monaural listening [13.2]

The MAP and MAF curves are somewhat differently shaped, since the head, the pinna and the meatus have an influence on the sound field. The MAP curve shows only minor peaks and dips (±5 dB) for frequencies between about 0.2 kHz and 13 kHz, whereas the MAF curve shows a distinct dip around 3–4 kHz and a peak around 8–9 kHz. The difference derives mainly from a broad resonance produced by the meatus and pinna. The sound level at the eardrum is enhanced markedly for frequencies in the range 1.5–6 kHz, with a maximum enhancement at 3 kHz of about 15 dB.

The highest audible frequency varies considerably with age. Young children can often hear tones as high as 20 kHz, but for most adults the threshold rises rapidly above about 15 kHz. The loss of sensitivity with increasing age (presbyacusis) is much greater at high frequencies than at low, and the variability between different people is also greater at high frequencies. There seems to be no well-defined low-frequency limit to hearing, although very intense low-frequency tones can sometimes be felt as vibration before they are heard. The low-frequency limit for the true hearing of pure tones probably lies at about 20 Hz. This is close to the lowest frequency which evokes a pitch sensation [13.8].

A third method of specifying absolute thresholds is commonly used when measuring hearing in clinical situations, for example, when a hearing impairment is

Amazon
EX1013 Page 471

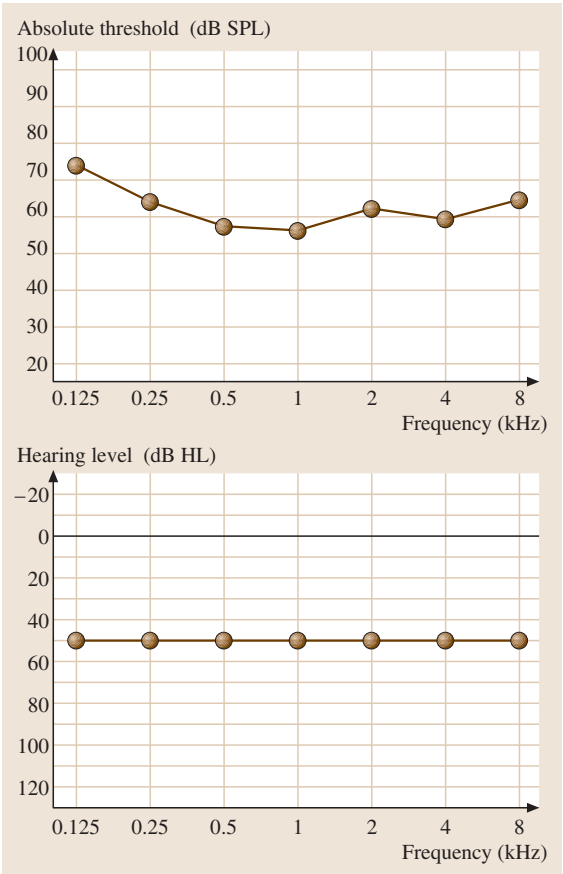


Fig. 13.2 Comparison of a clinical audiogram for a 50 dB hearing loss at all frequencies *(bottom)* and the absolute threshold curve for the same hearing loss plotted in terms of the MAP *(top)*

suspected; thresholds are specified relative to the average threshold at each frequency for young, healthy listeners with normal hearing. In this case, the sound level is usually specified relative to standardized values produced by a specific earphone in a specific coupler. A coupler is a device which contains a cavity or series of cavities and a microphone for measuring the sound produced by the earphone. The preferred earphone varies from one country to another. For example, the Telephonics TDH49 or TDH50 is often used in the UK and USA, while the Beyer DT48 is used in Germany. Thresholds specified in this way have units of dB HL (hearing level) in Europe or dB HTL (hearing threshold level) in the USA. For example, a threshold of 40 dB HL at 1 kHz would mean that the person had a threshold which was 40 dB higher than normal at that frequency. In psychoacoustic work, thresholds are normally plotted with threshold increasing upward, as in Fig. 13.1. However, in audiology, threshold elevations are shown as hearing losses, plotted downward. The average normal threshold is represented as a horizontal line at the top of the plot, and the degree of hearing loss is indicated by how much the threshold falls below this line. This type of plot is called an audiogram. Figure 13.2 compares an audiogram for a hypothetical hearing-impaired person with a flat hearing loss, with a plot of the same thresholds expressed as MAP values. Notice that, although the audiogram is flat, the corresponding MAP curve is not flat. Note also that thresholds in dB HL can be negative. For example, a threshold of −10 dB simply means that the individual is 10 dB more sensitive than the average.

The absolute thresholds described above were measured using tone bursts of relatively long duration. For durations exceeding about 500 ms, the sound level at threshold is roughly independent of duration. However, for durations less than about 200 ms, the sound level necessary for detection increases as duration decreases, by about 3 dB per halving of the duration [13.9]. Thus, the sound energy required for threshold is roughly constant.

13.2 Frequency Selectivity and Masking

Frequency selectivity refers to the ability to resolve or separate the sinusoidal components in a complex sound. It is a key aspect of the analysis of sounds by the auditory system, and it influences many aspects of auditory perception, including the perception of loudness, pitch and timbre. It is often demonstrated and measured by studying masking, which has been defined as 'The process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound' [13.10]. It has been known for many years that a signal is most easily masked by a sound having frequency components close to, or the same as, those of the signal [13.11]. This led to the idea that our ability to separate the components of a complex sound depends, at least in part, on the frequency analysis that takes place on the basilar membrane (see Chap. 12).

Amazon
EX1013 Page 472

13.2.1 The Concept of the Auditory Filter

Fletcher [13.13], following *Helmholtz* [13.14], suggested that the peripheral auditory system behaves as if it contains a bank of bandpass filters, with overlapping passbands. These filters are now called the auditory filters. Fletcher thought that the basilar membrane provided the basis for the auditory filters. Each location on the basilar membrane responds to a limited range of frequencies, so each different point corresponds to a filter with a different center frequency. When trying to detect a signal in a broadband noise background, the listener is assumed to make use of a filter with a center frequency close to that of the signal. This filter passes the signal but removes a great deal of the noise. Only the components in the noise which pass through the filter have any effect in masking the signal. It is usually assumed that the threshold for detecting the signal is determined by the amount of noise passing through the auditory filter; specifically, threshold is assumed to correspond to a certain signal-to-noise ratio at the output of the filter. This set of assumptions has come to be known as the *power spectrum model* of masking [13.15], since the stimuli are represented by their long-term power spectra, i. e., the short-term fluctuations in the masker are ignored.

The question considered next is: What is the shape of the auditory filter? In other words, how does its relative response change as a function of the input frequency? Most methods for estimating the shape of the auditory filter at a given center frequency are based on the assumptions of the power spectrum model of masking. The threshold of a signal whose frequency is fixed is measured in the presence of a masker whose spectral content is varied. It is assumed, as a first approximation, that the signal is detected using the single auditory filter which is centered on the frequency of the signal, and that threshold corresponds to a constant signal-to-masker ratio at the output of that filter. The methods described below both measure the shape of the filter using this technique.

13.2.2 Psychophysical Tuning Curves

One method of measuring the shape of the filter involves a procedure which is analogous in many ways to the determination of a neural tuning curve (see Chap. 12), and the resulting function is often called a psychophysical tuning curve (PTC). To determine a PTC, the signal is fixed in level, usually at a very low level, say, 10 dB above absolute threshold (called 10 dB sensation level, SL). The masker can be either a sinusoid or a band of noise covering a small frequency range.

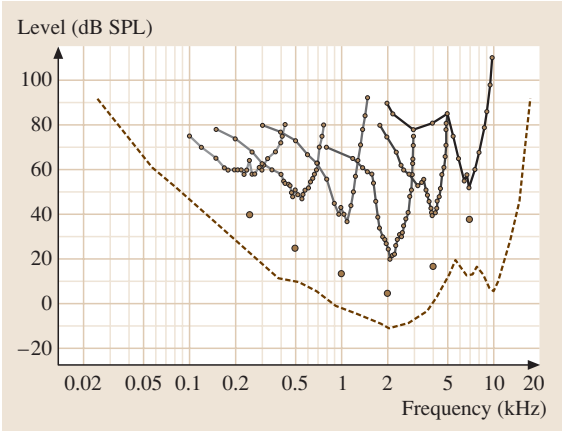


Fig. 13.3 Psychophysical tuning curves (PTCs) determined in simultaneous masking, using sinusoidal signals at 10 dB SL. For each curve, the *circle* below it indicates the frequency and level of the signal. The masker was a sinusoid which had a fixed starting phase relationship to the 50 ms signal. The masker level required for threshold is plotted as a function of masker frequency on a logarithmic scale. The *dashed line* shows the absolute threshold for the signal. (After *Vogten* [13.12])

For each of several masker frequencies, the level of the masker needed just to mask the signal is determined. Because the signal is at a low level it is assumed that it produces activity primarily at the output of a single auditory filter. It is assumed further that at threshold the masker produces a constant output from that filter, in order to mask the fixed signal. Thus the PTC indicates the masker level required to produce a fixed output from the auditory filter as a function of frequency. Normally a filter characteristic is determined by plotting the output from the filter for an input varying in frequency and fixed in level. However, if the filter is linear the two methods give the same result. Thus, assuming linearity, the shape of the auditory filter can be obtained simply by inverting the PTC. Examples of some PTCs are given in Fig. 13.3.

One problem in interpreting PTCs is that, in practice, the listener may use the information from more than one auditory filter. When the masker frequency is above the signal frequency the listener might do better to use the information from a filter centered just below the signal frequency. If the filter has a relatively flat top, and sloping edges, this will considerably attenuate the masker at the filter output, while only slightly attenuating

Amazon
EX1013 Page 473

the signal. By using this off-center filter the listener can improve performance. This is known as *off-frequency listening*, and there is now good evidence that humans do indeed listen off-frequency when it is advantageous to do so [13.16,17]. The result of off-frequency listening is that the PTC has a sharper tip than would be obtained if only one auditory filter were involved [13.18].

13.2.3 The Notched-Noise Method

Patterson [13.19] described a method of determining auditory filter shape which limits off-frequency listening. The method is illustrated in Fig. 13.4. The signal (indicated by the bold vertical line) is fixed in frequency, and the masker is a noise with a bandstop or notch centered at the signal frequency. The deviation of each edge of the noise from the center frequency is denoted by Δf. The width of the notch is varied, and the threshold of the signal is determined as a function of notch width. Since the notch is symmetrically placed around the signal frequency, the method cannot reveal asymmetries in the auditory filter, and the analysis assumes that the filter is symmetric on a linear frequency scale. This assumption appears not unreasonable, at least for the top part of the filter and at moderate sound levels, since PTCs are quite symmetric around the tips. For a signal symmetrically placed in a bandstop noise, the optimum signal-to-masker ratio at the output of the auditory filter is achieved with a filter centered at the signal frequency, as illustrated in Fig. 13.4.

As the width of the spectral notch is increased, less noise passes through the auditory filter. Thus the threshold of the signal drops. The amount of noise passing through the auditory filter is proportional to the area under the filter in the frequency range covered by the noise. This is shown as the dark shaded areas in Fig. 13.4. Assuming that threshold corresponds to a constant signal-to-masker ratio at the output of the filter, the change in signal threshold with notch width indicates how the area under the filter varies with Δf. The area under a function between certain limits is obtained by integrating the value of the function over those limits. Hence by differentiating the function relating threshold to Δf, the relative response of the filter at that value of Δf is obtained. In other words, the relative response of the filter for a given deviation, Δf, from the center frequency is equal to the slope of the function relating signal threshold to notch width, at that value of Δf.

A typical auditory filter derived using this method is shown in Fig. 13.5. It has a rounded top and quite steep skirts. The sharpness of the filter is often specified as the bandwidth of the filter at which the response has fallen by a factor of two in power, i. e., by 3 dB. The 3 dB bandwidths of the auditory filters derived using the notched-noise method are typically between 10% and

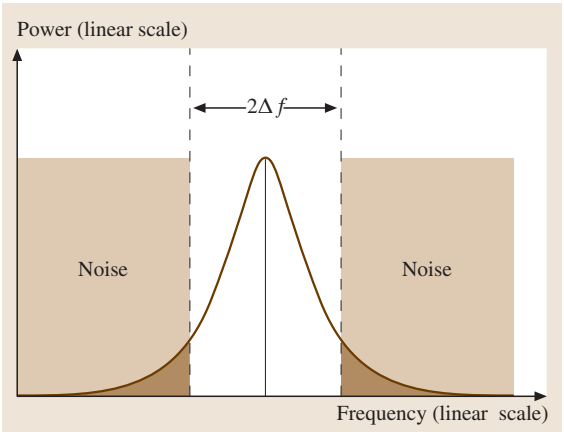


Fig. 13.4 Schematic illustration of the technique used by *Patterson*. [13.19] to determine the shape of the auditory filter. The threshold of the sinusoidal signal (indicated by the *bold vertical line*) is measured as a function of the width of a spectral notch in the noise masker. The amount of noise passing through the auditory filter centered at the signal frequency is proportional to the shaded areas

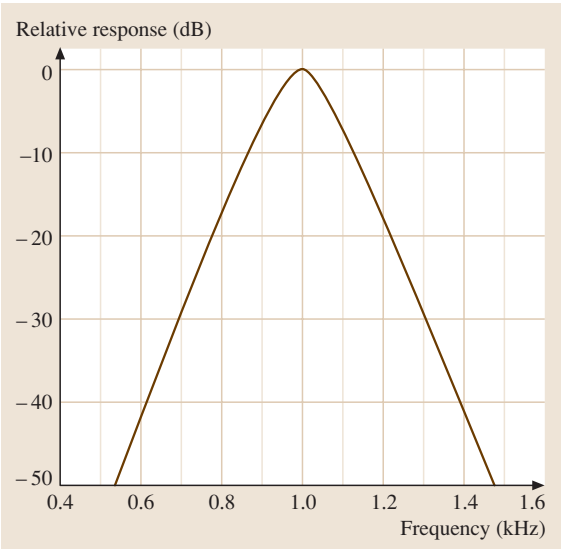


Fig. 13.5 A typical auditory filter shape determined using Patterson's method. The filter is centered at 1 kHz. The relative response of the filter (in dB) is plotted as a function of frequency

Amazon
EX1013 Page 474

15% of the center frequency. An alternative measure is the equivalent rectangular bandwidth (ERB), which is the bandwidth of a rectangular filter which has the same peak transmission as the filter of interest and which passes the same total power for a white noise input. The ERB of the auditory filter is a little larger than the 3 dB bandwidth. In what follows, the mean ERB of the auditory filter determined using young listeners with normal hearing and using a moderate noise level is denoted $\mathrm{ERB_N}$ (where the subscript N denotes normal hearing). An equation describing the value of $\mathrm{ERB_N}$ as a function of center frequency, F (in Hz), is [13.20]:

$$\mathrm{ERB_N} = 24.7(0.00437F + 1)\,. \qquad (13.1)$$

Sometimes it is useful to plot psychoacoustical data on a frequency scale related to $\mathrm{ERB_N}$. Essentially, the value of $\mathrm{ERB_N}$ is used as the unit of frequency. For example, the value of $\mathrm{ERB_N}$ for a center frequency of 1 kHz is about 132 Hz, so an increase in frequency from 934 to 1066 Hz represents a step of one $\mathrm{ERB_N}$. A formula relating $\mathrm{ERB_N}$ number to frequency is [13.20]:

$$\mathrm{ERB_N\ number} = 21.4\log_{10}(0.00437F + 1), \qquad (13.2)$$

where F is frequency in Hz. This scale is conceptually similar to the Bark scale proposed by *Zwicker* and coworkers [13.22], although it differs somewhat in numerical values.

The notched-noise method has been extended to include conditions where the spectral notch in the noise is placed asymmetrically about the signal frequency. This allows the measurement of any asymmetry in the auditory filter, but the analysis of the results is more difficult, and has to take off-frequency listening into account [13.23]. It is beyond the scope of this chapter to give details of the method of analysis; the interested reader is referred to [13.15, 20, 24, 25]. The results show that the auditory filter is reasonably symmetric at moderate sound levels, but becomes increasingly asymmetric at high levels, the low-frequency side becoming shallower than the high-frequency side. The filter shapes derived using the notched-noise method are quite similar to inverted PTCs [13.26], except that PTCs are slightly sharper around their tips, probably as a result of off-frequency listening.

13.2.4 Masking Patterns and Excitation Patterns

In the masking experiments described so far, the frequency of the signal was held constant, while the masker was varied. These experiments are most appropriate for estimating the shape of the auditory filter at a given center frequency. However, many of the early experiments on masking did the opposite; the masker was held constant in both level and frequency and the signal threshold was measured as a function of the signal frequency. The resulting functions are called masking patterns or masked audiograms.

Masking patterns show steep slopes on the low-frequency side (when the signal frequency is below that of the masker), of between 55 and 240 dB/octave. The slopes on the high-frequency side are less steep and depend on the level of the masker. A typical set of results is shown in Fig. 13.6. Notice that on the high-frequency side the curve is shallower at the highest level. Around the tip of the masking pattern, the growth of masking is approximately linear; a 10 dB increase in masker level leads to roughly a 10 dB increase in the signal threshold. However, for signal frequencies in the range from about 1300 to 2000 Hz, when the level of the masker is increased by 10 dB (e.g., from 70 to 80 dB SPL), the masked threshold increases by more than 10 dB; the amount of masking grows nonlinearly on the high-frequency side. This has been called the *upward spread of masking*.

The masking patterns do not reflect the use of a single auditory filter. Rather, for each signal frequency the listener uses a filter centered close to the signal frequency. Thus the auditory filter is shifted as the signal frequency is altered. One way of interpreting the masking pattern

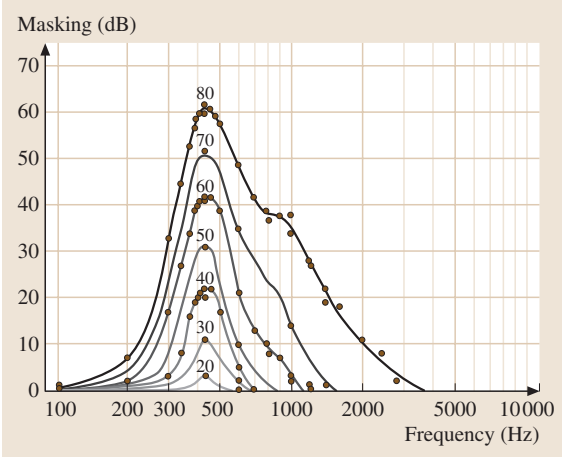


Fig. 13.6 Masking patterns for a narrow-band noise masker centered at 410 Hz. Each curve shows the elevation in threshold of a pure-tone signal as a function of signal frequency. The overall noise level for each curve is indicated in the figure (After *Egan* and *Hake* [13.21])

Amazon
EX1013 Page 475

Amazon EX1013 Page 476

is as a crude indicator of the excitation pattern of the masker [13.27]. The excitation pattern is a representation of the effective amount of excitation produced by a stimulus as a function of characteristic frequency (CF) on the basilar membrane (BM; see Chap. 12), and is plotted as effective level (in dB) against CF. In the case of a masking sound, the excitation pattern can be thought of as representing the relative amount of vibration produced by the masker at different places along the basilar membrane. The signal is detected when the excitation it produces is some constant proportion of the excitation produced by the masker at places with CFs close to the signal frequency. Thus the threshold of the signal as a function of frequency is proportional to the masker excitation level. The masking pattern should be parallel to the excitation pattern of the masker, but shifted vertically by a small amount. In practice, the situation is not so straightforward, since the shape of the masking pattern is influenced by factors such as off-frequency listening, and the detection of beats and combination tones [13.28].

Moore and *Glasberg* [13.29] have described a way of deriving the shapes of excitation patterns using the concept of the auditory filter. They suggested that the excitation pattern of a given sound can be thought of as the output of the auditory filters plotted as a function of their center frequency. To calculate the excitation pattern of a given sound, it is necessary to calculate the output of each auditory filter in response to that sound, and to plot the output as a function of the filter center frequency. The characteristics of the auditory filters are determined using the notched-noise method described earlier. Figure 13.7 shows excitation patterns calculated in this way for 1000 Hz sinusoids with various levels. The patterns are similar in form to the masking patterns shown in Fig. 13.6. Software for calculating excitation patterns can be downloaded from http://hearing.psychol.cam.ac.uk/Demos/demos.html.

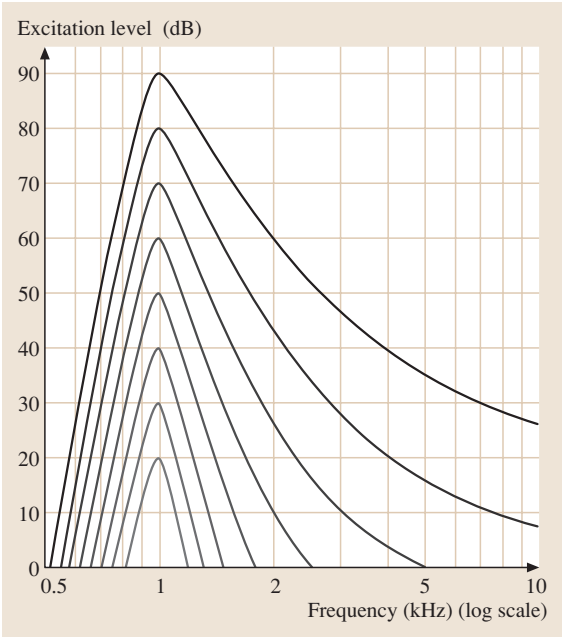


Fig. 13.7 Calculated psychoacoustical excitation patterns for a 1 kHz sinusoid at levels ranging from 20 to 90 dB SPL in 10 dB steps

13.2.5 Forward Masking

Masking can occur when the signal is presented just before or after the masker. This is called non-simultaneous masking and it is studied using brief signals, often called *probes*. Two basic types of non-simultaneous masking can be distinguished: (1) backward masking, in which the probe precedes the masker; and (2) forward masking, in which the probe follows the masker. Although many studies of backward masking have been published, the phenomenon is poorly understood. The amount of backward masking obtained depends strongly on how much practice the subjects have received, and practiced subjects often show little or no backward masking [13.30, 31]. The larger masking effects found for unpracticed subjects may reflect some sort of confusion of the signal with the masker. In the following, I will focus on forward masking, which can be substantial even in highly practiced subjects. The main properties of forward masking are as follows:

1. Forward masking is greater the nearer in time to the masker that the signal occurs. This is illustrated in the left panel of Fig. 13.8. When the delay D of the signal after the end of the masker is plotted on a logarithmic scale, the data fall roughly on a straight line. In other words, the amount of forward masking, in dB, is a linear function of log D.
2. The rate of recovery from forward masking is greater for higher masker levels. Regardless of the initial amount of forward masking, the masking decays to zero after 100–200 ms.
3. A given increment in masker level does not produce an equal increment in amount of forward masking. For example, if the masker level is increased by 10 dB, the masked threshold may only increase by 3 dB. This contrasts with simultaneous mask-

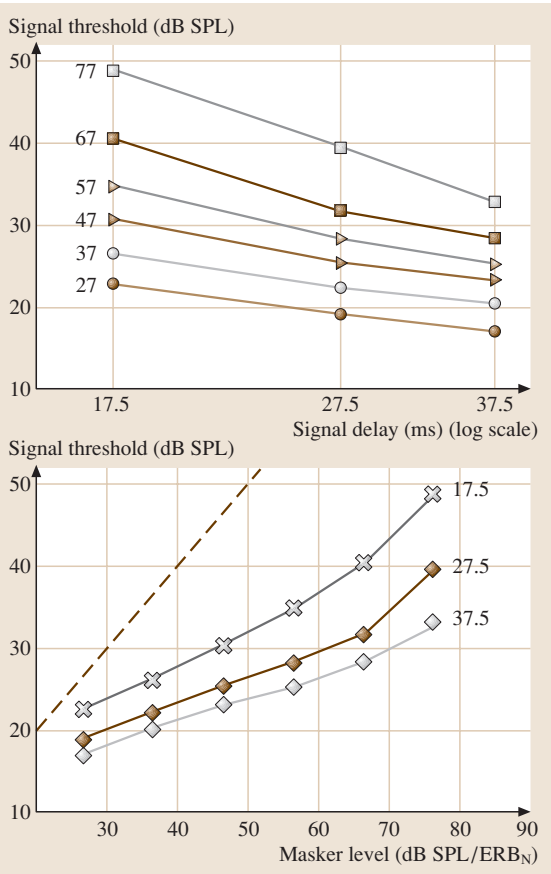


Fig. 13.8 The top panel shows the amount of forward masking of a brief 4 kHz signal, plotted as a function of the time delay of the signal after the end of the noise masker. Each curve shows results for a different noise level, specified as the level in a one-ERB$_N$-wide band centered at 4 kHz. The results for each level fall roughly on a straight line when the signal delay is plotted on a logarithmic scale, as here. The bottom panel shows the same thresholds plotted as a function of masker level. Each curve shows results for a different signal delay time (17.5, 27.5, or 37.5 ms). Note that the slopes of these growth of masking functions decrease with increasing signal delay. The *dashed line* indicates a slope of 1. (After *Moore* and *Glasberg* [13.32])

ing, where, at least for wide-band maskers, threshold corresponds approximately to a constant signal-to-masker ratio. This effect can be quantified by plotting the signal threshold as a function of the masker level. The resulting function is called a growth of masking function. Several such functions are shown in the right panel of Fig. 13.8. In simultaneous masking such functions would have slopes close to one, as indicated by the dashed line. In forward masking the slopes are usually less than one, and the slopes decrease as the value of D increases.

4. The amount of forward masking increases with increasing masker duration for durations up to at least 50 ms. The results for greater masker durations vary somewhat across studies. Some studies show an effect of masker duration for durations up to 200 ms [13.33, 34], while others show little effect for durations beyond 50 ms [13.35].
5. Forward masking is influenced by the relation between the frequencies of the signal and the masker (just as in the case of simultaneous masking).

The basis of forward masking is still not clear. Four factors may contribute:

1. The response of the BM to the masker continues for some time after the end of the masker, an effect known as *ringing*. If the ringing overlaps with the response to the signal, then this may contribute to the masking of the signal. The duration of the ringing is less at places tuned to high frequencies, whose bandwidth is larger than at low frequencies. Hence, ringing plays a significant role only at low frequencies [13.36–38]. For frequencies above 200–300 Hz, the amount of forward masking for a given masker level and signal delay time is roughly independent of frequency [13.39].
2. The masker produces short-term adaptation or fatigue in the auditory nerve or at higher centers in the auditory system, which reduces the response to a signal presented just after the end of the masker [13.40]. However, the effect in the auditory nerve appears to be too small to account for the forward masking observed behaviorally [13.41].
3. The neural activity evoked by the masker persists at some level in the auditory system higher than the auditory nerve, and this persisting activity masks the signal [13.42].
4. The masker may evoke a form of inhibition in the central auditory system, and this inhibition persists for some time after the end of the masker [13.43].

Oxenham and *Moore* [13.44] have suggested that the shallow slopes of the growth of masking functions, as shown in the bottom panel of Fig. 13.8, can be explained, at least qualitatively, in terms of the compressive input–output function of the BM (see Chap. 12). Such an input–output function is shown schematically

Amazon
EX1013 Page 477

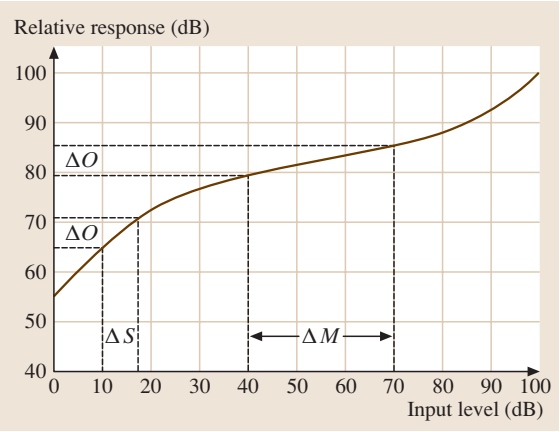


Fig. 13.9 Illustration of why growth of masking functions in forward masking usually have shallow slopes. The *solid curve* shows a schematic input–output function on the basilar membrane. The relative response is plotted on a dB scale with an arbitrary origin. When the masker is increased in level by ΔM, this produces an increase in response of ΔO. To restore signal threshold, the response to the signal also has to be increased by ΔO. This requires an increase in signal level, ΔS, which is markedly smaller than ΔM

in Fig. 13.9. It has a shallow slope for medium input levels, but a steeper slope at very low input levels. Assume that, for a given time delay of the signal relative to the masker, the response evoked by the signal at threshold is directly proportional to the response evoked by the masker. Assume, as an example, that a masker with a level of 40 dB produces a signal threshold of 10 dB. Consider now what happens when the masker level is increased by 30 dB. The increase in masker level, denoted by ΔM in Fig. 13.9, produces a relatively small increase in response, ΔO. To restore the signal to threshold, the signal has to be increased in level so that the response to it also increases by ΔO. However, this requires a relatively small increase in signal level, ΔS, as the signal level falls in the range where the input–output function is relatively steep. Thus, the growth of masking function has a shallow slope.

According to this explanation, the shallow slope of the growth of masking function arises from the fact that the signal level is lower than the masker level, so the masker is subject to more compression than the signal. The input–output function on the BM has a slope which decreases progressively with increasing level over the range 0 to about 50 dB. Hence the slope of the growth of masking function should decrease with increasing difference in level between the masker and signal. This can account for the progressive decrease in the slopes of the growth of masking functions with increasing delay between the signal and masker (see the right-hand panel of Fig. 13.8); longer delays are associated with greater differences in level between the signal and masker. Another prediction is that the growth of masking function for a given signal delay should increase in slope if the signal level is high enough to fall in the compressive region of the input–output function. Such an effect can be seen in the growth of masking function for the shortest delay time in Fig. 13.8; the function steepens for the highest signal level.

In summary, the processes underlying forward masking are not fully understood. Contributions from a number of different sources may be important. Temporal overlap of patterns of vibration on the BM may be important, especially for small delay times between the signal and masker. Short-term adaptation or fatigue in the auditory nerve may also play a role. At higher neural levels, a persistence of the excitation or inhibition evoked by the masker may occur. The form of the growth of masking functions can be explained, at least qualitatively, in terms of the nonlinear input–output functions observed on the BM.

13.2.6 Hearing Out Partials in Complex Tones

A complex tone, such as a tone produced by a musical instrument, usually evokes a single pitch; pitches corresponding to the frequencies of individual partials are not usually perceived. However, such pitches can be heard if attention is directed appropriately. In other words, individual partials can be *heard out*. *Plomp* [13.45] and *Plomp* and *Mimpen* [13.46] used complex tones with 12 equal-amplitude sinusoidal components to investigate the limits of this ability. The listener was presented with two comparison tones, one of which was of the same frequency as a partial in the complex; the other lay halfway between that frequency and the frequency of the adjacent higher or lower partial. The listener had to judge which of these two tones was a component of the complex. Plomp used two types of complex: a harmonic complex containing harmonics 1 to 12, where the frequencies of the components were integer multiples of that of the fundamental; and a nonharmonic complex, where the frequencies of the components were mistuned from simple frequency ratios. He found that for both kinds of complex, partials could only be heard out if they were sufficiently far in frequency from neighboring partials. The data, and other more recent data [13.47], are consis-

Amazon
EX1013 Page 478

tent with the hypothesis that a partial can be heard out (with 75% accuracy) when it is separated from neighboring (equal-amplitude) partials by 1.25 times the ERB_N of the auditory filter. For harmonic complex tones, only the first (lowest) five to eight harmonics can be heard out, as higher harmonics are separated by less than 1.25 ERB_N.

It seems likely that the analysis of partials from a complex sound depends in part on factors other than the frequency analysis that takes place on the basilar membrane. *Soderquist* [13.48] compared musicians and non-musicians in a task very similar to that of Plomp, and found that the musicians were markedly superior. This result could mean that musicians have smaller auditory filter bandwidths than non-musicians. However, *Fine* and *Moore* [13.49] showed that auditory filter bandwidths, as estimated in a notched-noise masking experiment, did not differ for musicians and non-musicians. It seems that some mechanism other than peripheral filtering is involved in hearing out partials from complex tones and that musicians, because of their greater experience, are able to make more efficient use of this mechanism.

13.3 Loudness

The human ear is remarkable both in terms of its absolute sensitivity and the range of sound intensities to which it can respond. The most intense sound that can be heard without damaging the ear has a level about 120 dB above the absolute threshold; this range is referred to as the dynamic range of the auditory system and it corresponds to a ratio of intensities of 1 000 000 000 000:1.

Loudness corresponds to the subjective impression of the magnitude of a sound. The formal definition of loudness is: that intensive attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from soft to loud [13.10]. Because loudness is subjective, it is very difficult to measure in a quantitative way. Estimates of loudness can be strongly affected by bias and context effects of various kinds [13.50,51]. For example the impression of loudness of a sound with a moderate level (say, 60 dB SPL) can be affected by presenting a high level sound (say, 100 dB SPL) before the moderate level sound.

13.3.1 Loudness Level and Equal-Loudness Contours

It is often useful to be able to compare the loudness of sounds with that of a standard, reference sound. The most common reference sound is a 1000 Hz sinusoid, presented binaurally in a free field, with the sound coming from directly in front of the listener. The loudness level of a sound is defined as the intensity level of a 1000 Hz sinusoid that is equal in loudness to the sound. The unit of loudness level is the phon. Thus, the loudness level of any sound in phons is the level (in dB SPL) of the 1000 Hz sinusoid to which it sounds equal in loudness. For example, if a sound appears to be as loud as a 1000 Hz sinusoid with a level of 45 dB SPL, then the sound has a loudness level of 45 phons. To determine the loudness level of a given sound, the subject is asked to adjust the level of a 1000 Hz sinusoid until it appears to have the same loudness as that sound. The 1000 Hz sinusoid and the test sound are presented alternately rather than simultaneously.

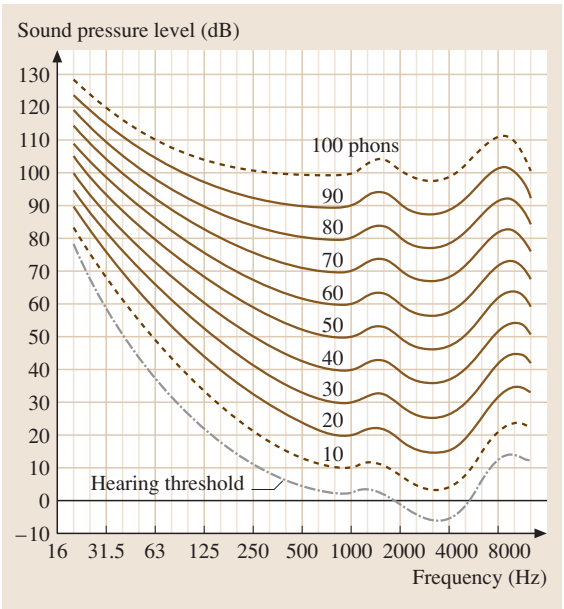


Fig. 13.10 Equal-loudness contours for loudness levels from 10 to 100 phons for sounds presented binaurally from the frontal direction. The absolute threshold curve (the MAF) is also shown. The curves for loudness levels of 10 and 100 phons are *dashed*, as they are based on interpolation and extrapolation, respectively

Amazon
EX1013 Page 479

In a variation of this procedure, the 1000 Hz sinusoid is fixed in level, and the test sound is adjusted to give a loudness match, again with alternating presentation. If this is repeated for various different frequencies of a sinusoidal test sound, an equal-loudness contour is generated [13.52, 53]. For example, if the 1000 Hz sinusoid is fixed in level at 40 dB SPL, then the 40 phon equal-loudness contour is generated. Figure 13.10 shows equal-loudness contours as published in a recent standard [13.53]. The figure shows equal-loudness contours for binaural listening for loudness levels from 10–100 phons, and it also includes the absolute threshold (MAF) curve. The listening conditions were similar to those for determining the MAF curve; the sound came from a frontal direction in a free field. The equal-loudness contours are of similar shape to the MAF curve, but tend to become flatter at high loudness levels. Note that the MAF curve in Fig. 13.10 differs somewhat from that in Fig. 13.1, as the two curves are based on different standards.

Note that the subjective loudness of a sound is not directly proportional to its loudness level in phons. For example, a sound with a loudness level of 80 phons sounds much more than twice as loud as a sound with a loudness level of 40 phons. This is discussed in more detail in the next section.

13.3.2 The Scaling of Loudness

Several methods have been developed that attempt to measure directly the relationship between the physical magnitude of sound and perceived loudness [13.54]. In one, called magnitude estimation, sounds with various different levels are presented, and the subject is asked to assign a number to each one according to its perceived loudness. In a second method, called magnitude production, the subject is asked to adjust the level of a sound until it has a loudness corresponding to a specified number.

On the basis of results from these two methods, Stevens suggested that loudness, L, was a power function of physical intensity, I:

$$L = kI^{0.3}x\,, \tag{13.3}$$

where k is a constant depending on the subject and the units used. In other words, the loudness of a given sound is proportional to its intensity raised to the power 0.3. Note that this implies that loudness is *not* linearly related to intensity; rather, it is a *compressive* function of intensity. An approximation to this equation is that the loudness doubles when the intensity is increased by a factor of 10, or, equivalently, when the level is increased by 10 dB. In practice, this relationship only holds for sound levels above about 40 dB SPL. For lower levels than this, the loudness changes with intensity more rapidly than predicted by the power-law equation.

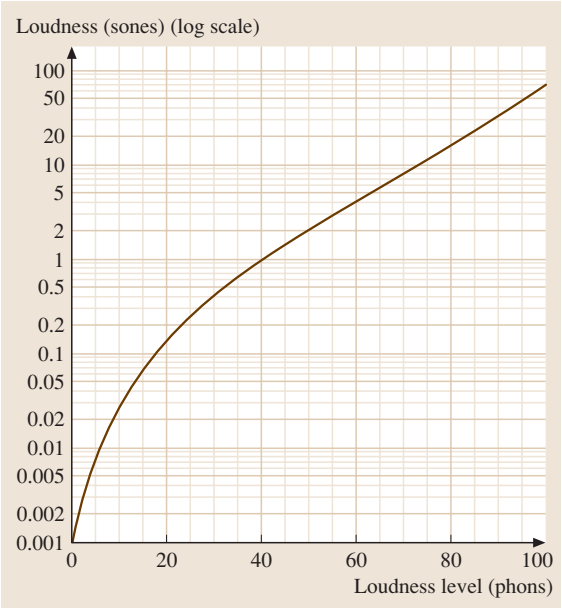


Fig. 13.11 The relationship between loudness in sones and loudness level in phons for a 1000 Hz sinusoid. The curve is based on the loudness model of *Moore* et al.[13.3]

The unit of loudness is the *sone*. One sone is defined arbitrarily as the loudness of a 1000 Hz sinusoid at 40 dB SPL, presented binaurally from a frontal direction in a free field. Fig. 13.11 shows the relationship between loudness in sones and the physical level of a 1000 Hz sinusoid, presented binaurally from a frontal direction in a free-field; the level of the 1000 Hz tone is equal to its loudness level in phons. This figure is based on predictions of a loudness model [13.3], but it is consistent with empirical data obtained using scaling methods [13.55]. Since the loudness in sones is plotted on a logarithmic scale, and the decibel scale is itself logarithmic, the curve shown in Fig. 13.11 approximates a straight line for levels above 40 dB SPL. The slope corresponds to a doubling of loudness for each 10 dB increase in sound level.

13.3.3 Neural Coding and Modeling of Loudness

The mechanisms underlying the perception of loudness are not fully understood. A common assumption is that

Amazon
EX1013 Page 480

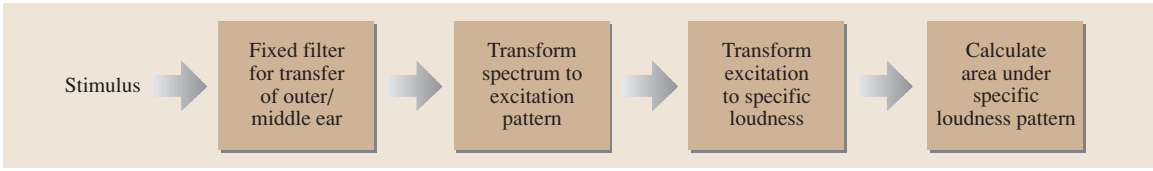


Fig. 13.12 Block diagram of a typical loudness model

loudness is somehow related to the total neural activity evoked by a sound, although this concept has been questioned [13.56]. In any case, it is commonly assumed that loudness depends upon a summation of loudness contributions from different frequency channels (i. e., different auditory filters). Models incorporating this basic concept have been proposed by *Fletcher* and *Munson* [13.57], by *Zwicker* [13.58, 59] and by *Moore* and coworkers [13.3, 60]. The models attempt to calculate the average loudness that would be perceived by a large group of listeners with normal hearing under conditions where biases are minimized as far as possible. The models all have the basic form illustrated in Fig. 13.12. The first stage is a fixed filter to account for the transmission of sound through the outer and middle ear. The next stage is the calculation of an excitation pattern for the sound under consideration. In most of the models, the excitation pattern is calculated from psychoacoustical masking data, as described earlier. From the excitation-pattern stage onwards, the models should be considered as multichannel; the excitation pattern is sampled at regular intervals along the $\mathrm{ERB_N}$ number scale, each sample value corresponding to the amount of excitation at a specific center frequency.

The next stage is the transformation from excitation level (dB) to specific loudness, which is a kind of *loudness density*. It represents the loudness that would be evoked by the excitation within a small fixed distance along the basilar membrane if it were possible to present that excitation alone (without any excitation at adjacent regions on the basilar membrane). In the model of *Moore* et al. [13.3], the distance is 0.89 mm, which corresponds to one $\mathrm{ERB_N}$, so the specific loudness represents the loudness per $\mathrm{ERB_N}$. The specific loudness plotted as a function of $\mathrm{ERB_N}$ number is called the *specific loudness pattern*. The specific loudness cannot be measured either physically or subjectively. It is a theoretical construct used as an intermediate stage in the loudness models. The transformation from excitation level to specific loudness involves a compressive nonlinearity. Although the models are based on psychoacoustical data, this transformation can be thought of as representing the way that physical excitation is transformed into neural activity; the specific loudness is assumed to be related to the amount of neural activity at the corresponding CF. The overall loudness of a given sound, in sones, is assumed to be proportional to the total area under the specific loudness pattern. In other words, the overall loudness is calculated by summing the specific loudness values across all $\mathrm{ERB_N}$ numbers (corresponding to all center frequencies).

Loudness models of this type have been rather successful in accounting for experimental data on the loudness of both simple sounds and complex sounds [13.3]. They have also been incorporated into loudness meters, which can give an appropriate indication of the loudness of sounds even when they fluctuate over time [13.27, 60, 61]. Software for calculating loudness using the model of *Moore* et al. [13.3] can be downloaded from http://hearing.psychol.cam.ac.uk/Demos/demos.html. The new American National Standards Institute (ANSI) standard for calculating loudness [13.62] is based on this model.

13.3.4 The Effect of Bandwidth on Loudness

Consider a complex sound of fixed energy (or intensity) having a bandwidth W. If W is less than a certain bandwidth, called the critical bandwidth for loudness, $\mathrm{CB_L}$, then the loudness of the sound is almost independent of W; the sound is judged to be about as loud as a pure tone or narrow band of noise of equal intensity lying at the center frequency of the band. However, if W is increased beyond $\mathrm{CB_L}$, the loudness of the complex begins to increase. This has been found to be the case for bands of noise [13.27, 63] and for complex sounds consisting of several pure tones whose frequency separation is varied [13.64, 65]. An example for bands of noise is given in Fig. 13.13. The $\mathrm{CB_L}$ for the data in Fig. 13.13 is about 250–300 Hz for a center frequency of 1420 Hz, although the exact value of $\mathrm{CB_L}$ is hard to determine. The value of $\mathrm{CB_L}$ is similar to, but a little greater than, the $\mathrm{ERB_N}$ of the auditory filter. Thus, for a given amount of energy, a complex sound is louder if its bandwidth exceeds one $\mathrm{ERB_N}$ than if its bandwidth is less than one $\mathrm{ERB_N}$.

Amazon
EX1013 Page 481

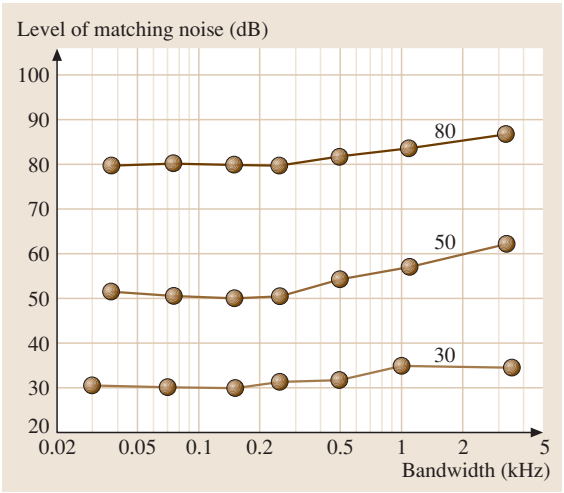


Fig. 13.13 The level of a 210 Hz-wide noise required to match the loudness of a noise of variable bandwidth is plotted as a function of the variable bandwidth; the bands of noise were geometrically centered at 1420 Hz. The number by each curve is the overall noise level in dB SPL. (After *Zwicker* et al.[13.63])

The pattern of results shown in Fig. 13.13 can be understood in terms of the loudness models described above. First, a qualitative account is given to illustrate the basic concept. When a sound has a bandwidth less than one ERB_N at a given center frequency, the excitation pattern and the specific loudness pattern are roughly independent of the bandwidth of the sound [13.66]. Further, the overall loudness is dominated by the specific loudness at the peak of the pattern. Consider now the effect of increasing the bandwidth of a sound from one ERB_N to N ERB_N, keeping the overall power, $P_{overall}$, constant. The power in each one-ERB_N-wide band is now $P_{overall}/N$. However, the peak specific loudness evoked by each band decreases by much less than a factor of N, because of the compressive relationship between excitation and specific loudness. For example, if $N = 10$, the peak specific loudness evoked by each band decreases by a factor of about two. Thus, in this example, there are ten bands each about half as loud as the original one-ERB_N-wide band, so the overall loudness is about a factor of five greater for the ten-ERB_N-wide than for the one-ERB_N-wide band. Generally, once the bandwidth is increased beyond one ERB_N, the loudness goes up because the decrease in loudness per ERB_N is more than offset by the increase in the number of channels contributing significantly to the overall loudness.

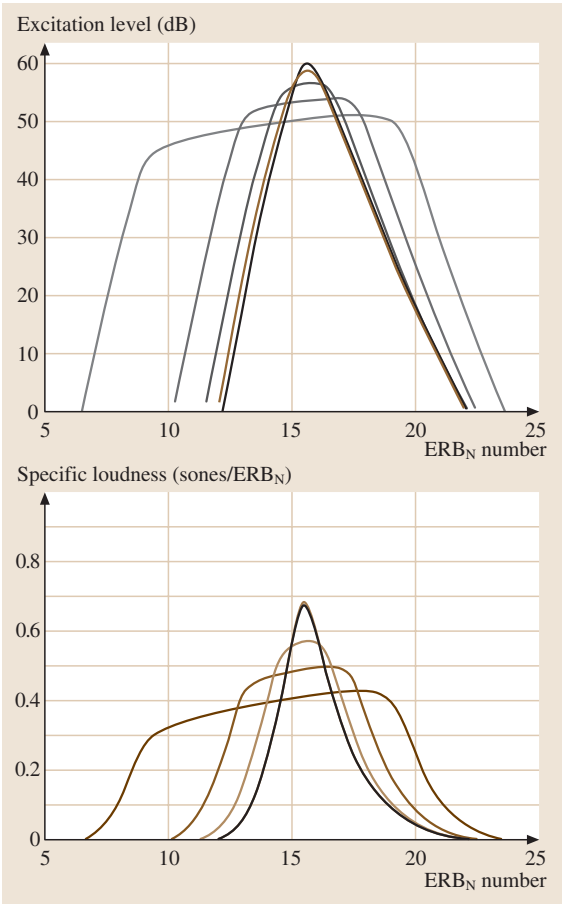


Fig. 13.14 The upper panel shows excitation patterns for a 1 kHz sinusoid with a level of 60 dB SPL (the narrowest pattern with the highest peak) and for noise bands of various widths, all centered at 1 kHz and with an overall level of 60 dB SPL. The frequency scale has been transformed to an ERB_N number scale. The noise bands have widths of 20, 60, 140, 300, 620 and 1260 Hz. As the bandwidth is increased, the patterns decrease in height but spread over a greater number of ERB_Ns. The lower panel shows specific loudness patterns corresponding to the excitation patterns in the upper panel. For bandwidths up to 140 Hz, the area under the specific loudness patterns is constant. For greater bandwidths, the total area increases

This argument is illustrated in a more quantitative way in Fig. 13.14, which shows excitation patterns (top) and specific loudness patterns (bottom) for a sinusoid and for bands of noise of various widths (20, 40, 60, 140, 300 and 620 Hz), all with a level of 60 dB SPL, calculated using the model of *Moore* et al. [13.3]. With

Amazon
EX1013 Page 482

increasing bandwidth, the specific loudness patterns become lower at their tips, but broader. For the first three bandwidths, the small decrease in area around the tip is almost exactly canceled by the increase on the skirts, so the total area remains almost constant. When the bandwidth is increased beyond 140 Hz, the increase on the skirts is greater than the decrease around the tip, and so the total area, and hence the predicted loudness, increases. The value of CB_L in this case is a little greater than 140 Hz. Since the increase in loudness depends on the summation of specific loudness at different center frequencies, the increase in loudness is often described as loudness summation.

At low sensation levels (around 10–20 dB SL), the loudness of a complex sound is roughly independent of bandwidth. This can also be explained in terms of the loudness models described above. At these low levels, specific loudness changes rapidly with excitation level, and so does loudness. As a result, the total area under the specific loudness pattern remains almost constant as the bandwidth is altered, even for bandwidths greater than CB_L. Thus, loudness is independent of bandwidth. When a narrow-band sound has a very low sensation level (below 10 dB), then if the bandwidth is increased keeping the total energy constant, the output of each auditory filter becomes insufficient to make the sound audible. Accordingly, near threshold, loudness must decrease as the bandwidth of a complex sound is increased from a small value. As a consequence, if the intensity of a complex sound is increased slowly from a subthreshold value, the rate of growth of loudness is greater for a wide-band sound than for a narrow-band sound.

13.3.5 Intensity Discrimination

The smallest detectable change in intensity of a sound has been measured for many different types of stimuli by a variety of methods. The three main methods are:

1. Modulation detection. The stimulus is amplitude modulated (i. e., made to vary in amplitude) at a slow regular rate and the listener is required to detect the modulation. Usually, the modulation is sinusoidal.
2. Increment detection. A continuous background stimulus is presented, and the subject is required to detect a brief increment in the level of the background. Often the increment is presented at one of two possible times, indicated by lights, and the listener is required to indicate whether the increment occurred synchronously with the first light or the second light.
3. Intensity discrimination of gated or pulsed stimuli. Two (or more) separate pulses of sound are presented successively, one being more intense than the other(s), and the subject is required to indicate which pulse was the most intense.

In all of these tasks, the subjective impression of the listener is of a change in loudness. For example, in method 1 the modulation is heard as a fluctuation in loudness. In method 2 the increment is heard as a brief increase in loudness of the background, or sometimes as an extra sound superimposed on the background. In method 3, the most intense pulse appears louder than the other(s). Although there are some minor discrepancies in the experimental results for the different methods, the general trend is similar. For wide-band noise, or for bandpass-filtered noise, the smallest detectable intensity change, ΔI, is approximately a constant fraction of the intensity of the stimulus, I. In other words, $\Delta I/I$ is roughly constant. This is an example of Weber's Law, which states that the smallest detectable change in a stimulus is proportional to the magnitude of that stimulus. The value of $\Delta I/I$ is called the Weber fraction. Thresholds for detecting intensity changes are often specified as the change in level at threshold, ΔL, in decibels. The value of ΔL is given by

$$\Delta L = 10 \log_{10}[(I + \Delta I)/I] \,. \tag{13.4}$$

As $\Delta I/I$ is constant, ΔL is also constant, regardless of the absolute level, and for wide-band noise has a value of about 0.5–1 dB. This holds from about 20 dB above threshold to 100 dB above threshold [13.67]. The value of ΔL increases for sounds which are close to the absolute threshold.

For sinusoidal stimuli, the situation is somewhat different. If ΔI (in dB) is plotted against I (also in dB), a straight line is obtained with a slope of about 0.9; Weber's law would give a slope of 1.0. Thus discrimination, as measured by the Weber fraction, improves at high levels. This has been termed the "near miss" to Weber's Law. The data of Riesz [13.68] for modulation detection show a value of ΔL of 1.5 dB at 20 dB SL, 0.7 dB at 40 dB SL, and 0.3 dB at 80 dB SL (all at 1000 Hz). The Weber fraction may increase somewhat at very high sound levels (above 100 dB SPL) [13.69]. In everyday life, a change in level of 1 dB would hardly be noticed, but a change in level of 3 dB (corresponding to a doubling or halving of intensity) would be fairly easily heard.

Amazon
EX1013 Page 483

13.4 Temporal Processing in the Auditory System

This section is concerned mainly with temporal resolution (or acuity), which refers to the ability to detect changes in stimuli over time, for example, to detect a brief gap between two stimuli or to detect that a sound is modulated in some way. As pointed out by *Viemeister* and *Plack* [13.71], it is also important to distinguish the rapid pressure variations in a sound (the *fine structure*) from the slower overall changes in the amplitude of those fluctuations (the *envelope*). Temporal resolution normally refers to the resolution of changes in the envelope, not in the fine structure. In characterizing temporal resolution in the auditory system, it is important to take account of the filtering that takes place in the peripheral auditory system. Temporal resolution depends on two main processes: analysis of the time pattern occurring within each frequency channel and comparison of the time patterns across channels.

A major difficulty in measuring the temporal resolution of the auditory system is that changes in the time pattern of a sound are generally associated with changes in its magnitude spectrum, for example its power spectrum (see Chap. 14). Thus, the detection of a change in time pattern can sometimes depend not on temporal resolution per se, but on the detection of the change in magnitude spectrum. There have been two general approaches to getting around this problem. One is to use signals whose magnitude spectrum is not changed when the time pattern is altered. For example, the magnitude spectrum of white noise remains flat if the noise is interrupted, i. e., if a gap is introduced into the noise. The second approach uses stimuli whose magnitude spectra are altered by the change in time pattern, but extra background sounds are added to mask the spectral changes. Both approaches will be described.

13.4.1 Temporal Resolution Based on Within-Channel Processes

The threshold for detecting a gap in a broadband noise provides a simple and convenient measure of temporal resolution. Usually a two-alternative forced-choice (2AFC) procedure is used: the subject is presented with two successive bursts of noise and either the first or the second burst (at random) is interrupted to produce the gap. The task of the subject is to indicate which burst contained the gap. The gap threshold is typically 2–3 ms [13.72, 73]. The threshold increases at very low sound levels, when the level of the noise approaches the absolute threshold, but is relatively invariant with level for moderate to high levels.

The long-term magnitude spectrum of a sound is not changed when that sound is time reversed (played backward in time). Thus, if a time-reversed sound can be discriminated from the original, this must reflect a sensitivity to the difference in time pattern of the two sounds. This was exploited by *Ronken* [13.74], who used as stimuli pairs of clicks differing in amplitude. One click, labeled A, had an amplitude greater than that of the other click, labeled B. Typically the amplitude of A was twice that of B. Subjects were required to distinguish click pairs differing in the order of A and B: either AB or BA. The ability to do this was measured as a function of the time interval or gap between A and B. *Ronken* found that subjects could distinguish the click pairs for gaps down to 2–3 ms. Thus, the limit to temporal resolution found in this task is similar to that found for the detection of a gap in broadband noise. It should be noted that, in this task, subjects do not hear the individual clicks within a click pair. Rather, each click pair is heard as a single sound with its own characteristic quality.

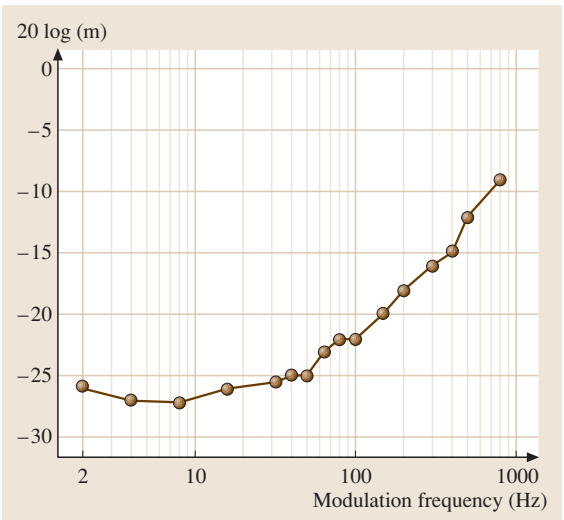


Fig. 13.15 A temporal modulation transfer function (TMTF). A broadband white noise was sinusoidally amplitude modulated, and the threshold amount of modulation required for detection is plotted as a function of modulation rate. The amount of modulation is specified as $20\log(m)$, where m is the modulation index (see Chap. 14). The higher the sensitivity to modulation, the more negative is this quantity. (After *Bacon* and *Viemeister* [13.70])

Amazon
EX1013 Page 484

The experiments described above each give a single number to describe temporal resolution. A more comprehensive approach is to measure the threshold for detecting changes in the amplitude of a sound as a function of the rapidity of the changes. In the simplest case, white noise is sinusoidally amplitude modulated, and the threshold for detecting the modulation is determined as a function of modulation rate. The function relating threshold to modulation rate is known as a temporal modulation transfer function (TMTF). Modulation of white noise does not change its long-term magnitude spectrum. An example of the results is shown in Fig. 13.15; data are taken from *Bacon* and *Viemeister* [13.70]. For low modulation rates, performance is limited by the amplitude resolution of the ear, rather than by temporal resolution. Thus, the threshold is independent of modulation rate for rates up to about 50 Hz. As the rate increases beyond 50 Hz, temporal resolution starts to have an effect; the threshold increases, and for rates above about 1000 Hz the modulation cannot be detected at all. Thus, sensitivity to amplitude modulation decreases progressively as the rate of modulation increases. The shapes of TMTFs do not vary much with overall sound level, but the ability to detect the modulation does worsen at low sound levels.

To explore whether temporal resolution varies with center frequency, *Green* [13.75] used stimuli which consisted of a brief pulse of a sinusoid in which the level of the first half of the pulse was 10 dB different from that of the second half. Subjects were required to distinguish two signals, differing in whether the half with the high level was first or second. Green measured performance as a function of the total duration of the stimuli. The threshold, corresponding to 75% correct discrimination, was similar for center frequencies of 2 and 4 kHz and was between 1 and 2 ms. However, the threshold was slightly higher for a center frequency of 1 kHz, being between 2 and 4 ms.

Moore et al. [13.76] measured the threshold for detecting a gap in a sinusoid, for signal frequencies of 100, 200, 400, 800, 1000, and 2000 Hz. A background noise was used to mask the spectral *splatter* produced by turning the sound off and on to produce the gap. The gap thresholds were almost constant, at 6–8 ms over the frequency range 400–2000 Hz, but increased somewhat at 200 Hz and increased markedly, to about 18 ms, at 100 Hz. Individual variability also increased markedly at 100 Hz.

Overall, it seems that temporal resolution is roughly independent of frequency for medium to high frequencies, but worsens somewhat at very low center frequencies.

The measurement of TMTFs using sinusoidal carriers is complicated by the fact that the modulation introduces spectral sidebands, which may be detected as separate components if they are sufficiently far in frequency from the carrier frequency. When the carrier frequency is high, the effect of resolution of sidebands is likely to be small for modulation frequencies up to a few hundred Hertz, as the auditory filter bandwidth is large for high center frequencies. Consistent with this, TMTFs for high carrier frequencies generally show an initial flat portion (sensitivity independent of modulation frequency), then a portion where threshold increases with increasing modulation frequency, presumably reflecting the limits of temporal resolution, and then a portion where threshold decreases again, presumably reflecting the detection of spectral sidebands [13.77, 78].

The initial flat portion of the TMTF extends to about 100–120 Hz for sinusoidal carriers, but only to 50 or 60 Hz for a broadband noise carrier. It has been suggested that the discrepancy occurs because the inherent amplitude fluctuations in a noise carrier limit the detectability of the imposed modulation [13.79–82]; see below for further discussion of this point. The effect of the inherent fluctuations depends upon their similarity to the imposed modulation. When a narrow-band noise carrier is used, which has relatively slow inherent amplitude fluctuations, TMTFs show the poorest sensitivity for low modulation frequencies [13.79, 80]. In principle, then, TMTFs obtained using sinusoidal carriers provide a better measure of the inherent temporal resolution of the auditory system than TMTFs obtained using noise carriers, provided that the modulation frequency is within the range where spectral resolution does not play a major role.

13.4.2 Modeling Temporal Resolution

Most models of temporal resolution are based on the idea that there is a process at levels of the auditory system higher than the auditory nerve which is sluggish in some way, thereby limiting temporal resolution. The models assume that the internal representation of stimuli is smoothed over time, so that rapid temporal changes are reduced in magnitude but slower ones are preserved. Although this smoothing process almost certainly operates on neural activity, the most widely used models are based on smoothing a simple transformation of the stimulus, rather than its neural representation.

Amazon
EX1013 Page 485

Most models include an initial stage of bandpass filtering, reflecting the action of the auditory filters. Each filter is followed by a nonlinear device. This nonlinear device is meant to reflect the operation of several processes that occur in the peripheral auditory system such as compression on the basilar membrane and neural transduction, whose effects resemble half-wave rectification (see Chap. 12). The output of the nonlinear device is fed to a smoothing device, which can be implemented either as a low-pass filter [13.83] or (equivalently) as a sliding temporal integrator [13.38, 84]. The device determines a kind of weighted average of the output of the compressive nonlinearity over a certain time interval or window. This weighting function is sometimes called the shape of the temporal window. The window itself is assumed to slide in time, so that the output of the temporal integrator is like a weighted running average of the input. This has the effect of smoothing rapid fluctuations while preserving slower ones. When a sound is turned on abruptly, the output of the temporal integrator takes some time to build up. Similarly, when a sound is turned off, the output of the integrator takes some time to decay. The shape of the window is assumed to be asymmetric in time, such that the build up of its output in response to the onset of a sound is more rapid than the decay of its output in response to the cessation of a sound. The output of the sliding temporal integrator is fed to a decision device. The decision device may use different rules depending on the task required. For example, if the task is to detect a brief temporal gap in a signal, the decision device might look for a dip in the output of the temporal integrator. If the task is to detect amplitude modulation of a sound, the device might assess the amount of modulation at the output of the sliding temporal integrator [13.83].

It is often assumed that backward and forward masking depend on the process of build up and decay. For example, if a brief signal is rapidly followed by a masker (backward masking), the response to the signal may still be building up when the masker occurs. If the masker is sufficiently intense, then its internal effects may swamp those of the signal. Similarly, if a brief signal follows soon after a masker (forward masking), the decaying response to the masker may swamp the response to the signal. The asymmetry in the shape of the window accounts for the fact that forward masking occurs over longer masker-signal intervals than backward masking.

13.4.3 A Modulation Filter Bank?

Some researchers have suggested that the analysis of sounds that are amplitude modulated depends on a specialized part of the brain that contains an array of neurons, each tuned to a different modulation rate [13.85]. Each neuron can be considered as a filter in the modulation domain, and the array of neurons is known collectively as a modulation filter bank. The modulation filter bank has been suggested as a possible explanation for certain perceptual phenomena, which are described below. It should be emphasized, however, that this is still a controversial concept.

The threshold for detecting amplitude modulation of a given carrier generally increases if additional amplitude modulation is superimposed on that carrier. This effect has been called modulation masking. *Houtgast* [13.86] studied the detection of sinusoidal amplitude modulation of a pink noise carrier. Thresholds for detecting the modulation were measured when no other modulation was present and when a masker modulator was present in addition. In one experiment, the masker modulation was a half-octave-wide band of noise, with a center frequency of 4, 8, or 16 Hz. For each masker, the masking pattern showed a peak at the masker frequency. This could be interpreted as indicating selectivity in the modulation-frequency domain, analogous to the frequency selectivity in the audio-frequency domain that was described earlier.

Bacon and *Grantham* [13.87] measured thresholds for detecting sinusoidal amplitude modulation of a broadband white noise in the presence of a sinusoidal masker modulator. When the masker modulation frequency was 16 or 64 Hz, most modulation masking occurred when the signal modulation frequency was near the masker frequency. In other words, the masking patterns were roughly bandpass, although they showed an increase for very low signal frequencies. For a 4 Hz masker, the masking patterns had a low-pass characteristic, i. e., there was a downward spread of modulation masking.

It should be noted that the sharpness of tuning of the hypothetical modulation filter bank is much less than the sharpness of tuning of the auditory filters in the audio-frequency domain. The bandwidths have been estimated as between 0.5 and 1 times the center frequency [13.80, 88–90]. The modulation filters, if they exist, are not highly selective.

Amazon
EX1013 Page 486

13.4.4 Duration Discrimination

Duration discrimination has typically been studied by presenting two successive sounds which have the same power spectrum but differ in duration. The subject is required to indicate which sound had the longer duration. Both *Creelman* [13.91] and *Abel* [13.92] found that the smallest detectable increase in duration, ΔT, increased with the baseline duration T. The data of Abel showed that, for $T = 10$, 100, and 1000 ms, ΔT was about 4, 15, and 60 ms, respectively. Thus, the Weber fraction, $\Delta T/T$, decreased with increasing T. The results were relatively independent of the overall level of the stimuli and were similar for noise bursts of various bandwidths and for bursts of a 1000 Hz sine wave.

Abel [13.93] reported somewhat different results for the discrimination of the duration of the silent interval between two *markers*. For silent durations, T, less than 160 ms, the results showed that discrimination improved as the level of the markers increased. The function relating $\Delta T/T$ to T was non-monotonic, reaching a local minimum for $T = 2.5$ ms and a local maximum for $T = 10$ ms. The value of ΔT ranged from 6 to 19 ms for a base duration of 10 ms and from 61 to 96 ms for a base duration of 320 ms.

Divenyi and *Danner* [13.94] required subjects to discriminate the duration of the silent interval defined by two 20 ms sounds. When the markers were identical high-level (86 dB SPL) bursts of sinusoids or noise, performance was similar across markers varying in center frequency (500–4000 Hz) and bandwidth. In contrast to the results of *Abel* [13.93], $\Delta T/T$ was almost independent of T over the range of T from 25 to 320 ms. Thresholds were markedly lower than those reported by *Abel* [13.93], being about 1.7 ms at $T = 25$ ms and 15 ms at $T = 320$ ms. This may have been a result of the extensive training of the subjects of Divenyi and Danner. For bursts of a 1 kHz sinusoid, performance worsened markedly when the level of the markers was decreased below 25 dB SL. Performance also worsened markedly when the two markers on either side of a silent interval were made different in level or frequency.

In summary, all studies show that, for values of T exceeding 10 ms, ΔT increases with T and ΔT is roughly independent of the spectral characteristics of the sounds. This is true both for the duration discrimination of sounds and for the discrimination of silent intervals bounded by acoustic markers, provided that the markers are identical on either side of the interval. However, ΔT increases at low sound levels, and also increases when the markers differ in level or frequency on either side of the interval.

13.4.5 Temporal Analysis Based on Across-Channel Processes

Studies of the ability to compare timing across different frequency channels can give very different results depending on whether the different frequency components in the sound are perceived as part of a single sound or as part of more than one sound. Also, it should be realized that subjects may be able to *distinguish* different time patterns, for example, a change in the relative onset time of two different frequencies, without the subjective impression of a change in time pattern; some sort of change in the quality of the sound may be all that is heard. The studies described next indicate the limits of the ability to compare timing across channels, using highly trained subjects.

Patterson and *Green* [13.95] and *Green* [13.75] have studied the discrimination of a class of signals which have the same long-term magnitude spectrum, but which differ in their short-term spectra. These sounds, called Huffman sequences, are brief, broadband click-like sounds, except that the energy in a certain frequency region is delayed relative to that in other regions. The amount of the delay, the center frequency of the delayed frequency region, and the width of the delayed frequency region can all be varied. If subjects can distinguish a pair of Huffman sequences differing, for example, in the amount of delay in a given frequency region, this implies that they are sensitive to the difference in time pattern, i. e., they must be detecting that one frequency region is delayed relative to other regions. *Green* [13.75] measured the ability of subjects to detect differences in the amount of delay in three frequency regions: 650, 1900 and 4200 Hz. He found similar results for all three center frequencies: subjects could detect differences in delay time of about 2 ms regardless of the center frequency of the delayed region.

It should be noted that subjects did not report hearing one part of the sound after the rest of the sound. Rather, the differences in time pattern were perceived as subtle changes in sound quality. Further, some subjects required extensive training to achieve the fine acuity of 2 ms, and even after this training the task required considerable concentration.

Zera and *Green* [13.96] measured thresholds for detecting asynchrony in the onset or offset of complex signals composed of many sinusoidal components.

Amazon
EX1013 Page 487

The components were either uniformly spaced on a logarithmic frequency scale or formed a harmonic series. In one stimulus, the standard, all components started and stopped synchronously. In the signal stimulus, one component was presented with an onset or offset asynchrony. The task of the subjects was to discriminate the standard stimulus from the signal stimulus. They found that onset asynchrony was easier to detect than offset asynchrony. For harmonic signals, onset asynchronies less than 1 ms could generally be detected, whether the asynchronous component was leading or lagging the other components. Thresholds for detecting offset asynchronies were larger, being about 3–10 ms when the asynchronous component ended after the other components and 10–30 ms when the asynchronous component ended before the other components. Thresholds for detecting asynchronies in logarithmically spaced complexes were generally 2–50 times larger than for harmonic complexes.

The difference between harmonically and logarithmically spaced complexes may be explicable in terms of perceptual grouping (see later for more details). The harmonic signal was perceived as a single sound source, i. e., all of the components appeared to belong together. The logarithmically spaced complex was perceived as a series of separate tones, like many notes being played at once on an organ. It seems that it is difficult to compare the timing of sound elements that are perceived as coming from different sources, a point that will be expanded later. The high sensitivity to onset asynchronies for harmonic complexes is consistent with the finding that the perceived timbres of musical instruments are partly dependent on the exact onset times and rates of rise of individual harmonics within each musical note [13.97]. This is described in more detail later on.

13.5 Pitch Perception

Pitch is an attribute of sound defined in terms of what is *heard*. It is defined formally as "that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from low to high" [13.10]. It is related to the physical repetition rate of the waveform of a sound; for a pure tone (a sinusoid) this corresponds to the frequency, and for a periodic complex tone to the fundamental frequency. Increasing the repetition rate gives a sensation of increasing pitch. Appropriate variations in repetition rate can give rise to a sense of melody. Variations in pitch are also associated with the intonation of voices, and they provide cues as to whether an utterance is a question or a statement and as to the emotion of the talker. Since pitch is a subjective attribute, it cannot be measured directly. Often, the pitch of a complex sound is assessed by adjusting the frequency of a sinusoid until the pitch of the sinusoid matches the pitch of the sound in question. The frequency of the sinusoid then gives a measure of the pitch of the sound. Sometimes a periodic complex sound, such as a pulse train, is used as a matching stimulus. In this case, the repetition rate of the pulse train gives a measure of pitch. Results are generally similar for the two methods, although it is easier to make a pitch match when the sounds to be matched do not differ very much in timbre (see Sect. 13.6.)

13.5.1 Theories of Pitch Perception

There are two traditional theories of pitch perception. One, the *place* theory, is based on the fact that different frequencies (or frequency components in a complex sound) excite different places along the basilar membrane, and hence neurons with different CFs. The place theory assumes that the pitch of a sound is related to the excitation pattern produced by that sound; for a pure tone the pitch is generally assumed to be determined by the position of maximum excitation [13.98].

An alternative theory, called the *temporal* theory, is based on the assumption that the pitch of a sound is related to the time pattern of the neural impulses evoked by that sound [13.99]. These impulses tend to occur at a particular phase of the waveform on the basilar membrane, a phenomenon called phase locking (see Chap. 18). The intervals between successive neural impulses approximate integer multiples of the period of the waveform and these intervals are assumed to determine the perceived pitch. The temporal theory cannot be applicable at very high frequencies, since phase locking does not occur for frequencies above about 5 kHz. However, the tones produced by most musical instruments, the human voice, and most everyday sound sources have fundamental frequencies well below this range.

Amazon
EX1013 Page 488

Many researchers believe that the perception of pitch involves both place mechanisms and temporal mechanisms. However, one mechanism may be dominant for a specific task or aspect of pitch perception, and the relative role of the two mechanisms almost certainly varies with center frequency.

13.5.2 The Perception of the Pitch of Pure Tones

The Frequency Discrimination of Pure Tones

It is important to distinguish between frequency selectivity and frequency discrimination. The former refers to the ability to resolve the frequency components of a complex sound. The latter refers to the ability to detect changes in frequency over time. Usually, the changes in frequency are heard as changes in pitch. The smallest detectable change in frequency is called the frequency difference limen (DL).

Place models of frequency discrimination [13.100, 101] predict that frequency discrimination should be related to frequency selectivity; both should depend on the sharpness of tuning on the basilar membrane. *Zwicker* [13.101] described a model of frequency discrimination based on changes in the excitation pattern evoked by the sound when the frequency is altered, inferring the shapes of the excitation patterns from masking patterns such as those shown in Fig. 13.6. The model is illustrated in Fig. 13.16. The figure shows two excitation patterns, corresponding to two tones with slightly different frequencies. A change in frequency results in a sideways shift of the excitation pattern. The change is assumed to be detectable whenever the excitation level at some point on the excitation pattern changes by more than about 1 dB.

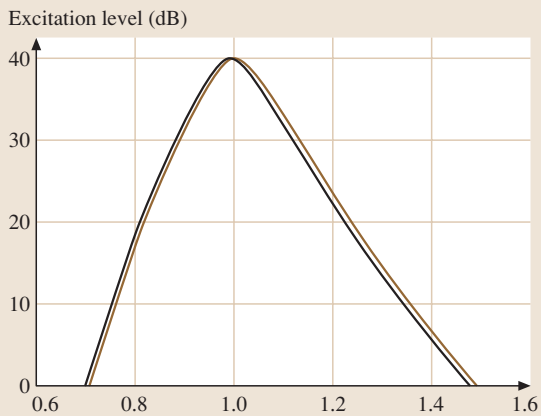


Fig. 13.16 Schematic illustration of an excitation-pattern model for frequency discrimination. Excitation patterns are shown for two sinusoidal tones differing slightly in frequency; the two tones have frequencies of 995 Hz and 1005 Hz. It is assumed that the difference in frequency, ΔF, can be detected if the excitation level changes anywhere by more than a criterion amount. The biggest change in excitation level is on the low-frequency side

The change in excitation level is greatest on the steeply sloping low-frequency (low-CF) side of the excitation pattern. Thus, in this model, the detection of a change in frequency is functionally equivalent to the detection of a change in level on the low-frequency side of the excitation pattern. The steepness of the low-frequency side is roughly constant when the frequency scale is expressed in units of $\mathrm{ERB_N}$, rather than in terms of linear frequency. The slope is about 18 dB per $\mathrm{ERB_N}$. To achieve a change in excitation level of 1 dB, the frequency has to be changed by one eighteenth of an $\mathrm{ERB_N}$. Thus, Zwicker's model predicts that the frequency DL at any given frequency should be about one eighteenth ($= 0.056$) of the value of $\mathrm{ERB_N}$ at that frequency.

To test Zwicker's model, frequency DLs have been measured as a function of center frequency. There have been two common ways of measuring frequency discrimination. One involves the discrimination of two successive steady tones with slightly different frequencies. On each trial, the tones are presented in a random order and the listener is required to indicate whether the first or second tone is higher in frequency. The frequency difference between the two tones is adjusted until the listener achieves a criterion percentage correct, for example 75%. This measure will be called the difference limen for frequency (DLF). A second measure, called the frequency modulation detection limen (FMDL), uses tones which are frequency modulated. In such tones, the frequency moves up and down in a regular periodic manner about the mean (carrier) frequency. The number of times per second that the frequency goes up and down is called the modulation rate. Typically, the modulation rate is rather low (2–20 Hz), and the change in frequency is heard as a fluctuation in pitch – a kind of *warble*. To determine a threshold for detecting frequency modulation, two tones are presented successively; one is modulated in frequency and the other has a steady frequency. The order of the tones on each trial is random. The listener is required to indicate whether the first or the second tone is modulated. The amount of modulation (also called the modulation depth) required to achieve a criterion response (e.g. 75% correct) is determined.

Amazon
EX1013 Page 489

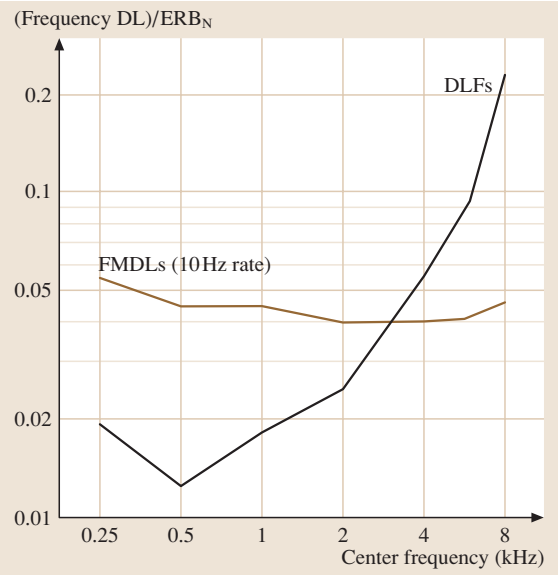


Fig. 13.17 Thresholds for detecting differences in frequency between steady pulsed tones (DLFs) and for detecting frequency modulation (FMDLs), plotted relative to the ERB$_N$ of the auditory filter at each center frequency. (After *Sek* and *Moore* [13.103])

It turns out that the results obtained with these two methods are quite different [13.102], although the difference depends on the modulation rate used to measure the FMDLs. An example of results obtained with the two methods is given in Fig. 13.17 (data from *Sek* and *Moore* [13.103]). For the FMDLs the modulation rate was 10 Hz. To test Zwicker's model, the DLFs and FMDLs were plotted as a proportion of the value of ERB$_N$ at the same center frequency. According to this model, the proportion should be independent of frequency. The proportion for FMDLs using a 10 Hz modulation rate, shown as the brighter line, is roughly constant, and its value is about 0.05, close to the value predicted by the model. However, the proportion for DLFs varies markedly with frequency [13.103, 104]. This is illustrated by the dark line in Fig. 13.17. The DLFs for frequencies of 2 kHz and below are smaller than predicted by Zwicker's model, while those for frequencies of 6 and 8 kHz are larger than predicted.

The results for the FMDLs are consistent with the place model, but the results for the DLFs are not. The reason for the deviation for DLFs is probably that DLFs at low frequencies depend on the use of temporal information from phase locking. Phase locking becomes less precise at frequencies above 1 kHz, and it is completely lost above 5 kHz. This can account for the marked increase in the DLFs at high frequencies [13.105].

The ratio FMDL/ERB$_N$ is not constant across center frequency when the modulation rate is very low (around 2 Hz), but increases with increasing frequency [13.103, 106]. For low center frequencies, FMDLs are smaller for a 2 Hz modulation rate than for a 10 Hz rate, while for high carrier frequencies (above 4 kHz) the reverse is true. Thus, for a 2 Hz modulation rate, the agreement between DLFs and FMDLs is better than for a 10 Hz rate, but discrepancies remain. For very low modulation rates, frequency modulation may be detected by virtue of the changes in phase locking to the carrier that occur over time. In other words, the frequency is determined over short intervals of time, using phase-locking information, and changes in the estimated frequency over time indicate the presence of frequency modulation. *Moore* and *Sek* [13.106] suggested that the mechanism for decoding the phase-locking information was sluggish; it had to sample the sound for a certain time in order to estimate its frequency. Hence, it could not follow rapid changes in frequency and it played little role for high modulation rates.

In summary, measures of frequency discrimination are consistent with the idea that DLFs, and FMDLs for very low modulation rates, are determined by temporal information (phase locking) for frequencies up to about 4–5 kHz. The precision of phase locking decreases with increasing frequency above 1–2 kHz, and it is almost absent above about 5 kHz. This can explain why DLFs increase markedly at high frequencies. FMDLs for medium to high modulation rates may be determined by a place mechanism, i. e., by the detection of changes in the excitation pattern. This mechanism may also account for DLFs and for FMDLs for low modulation rates, when the center frequency is above about 5 kHz.

The Perception of Musical Intervals

If temporal information plays a role in determining the pitch of pure tones, then we would expect changes in perception to occur for frequencies above 5 kHz, at which phase locking does not occur. Two aspects of perception do indeed change in the expected way, namely the perception of musical intervals, and the perception of melodies.

Two tones which are separated in frequency by an interval of one *octave* (i. e., one has twice the frequency of the other) sound similar. They are judged to have the same name on the musical scale (for example, C3 and C4). This has led several theorists to suggest that

Amazon
EX1013 Page 490

there are at least two dimensions to musical pitch. One aspect is related monotonically to frequency (for a pure tone) and is known as *tone height*. The other is related to pitch class (i. e., the name of the note) and is called *tone chroma* [13.107, 108]. For example, two sinusoids with frequencies of 220 and 440 Hz would have the same tone chroma (they would both be called A on the musical scale) but, as they are separated by an octave, they would have different tone heights.

If subjects are presented with a pure tone of a given frequency, f_1, and are asked to adjust the frequency, f_2 of a second tone (presented so as to alternate in time with the fixed tone) so that it appears to be an octave higher in pitch, they generally adjust f_2 to be roughly twice f_1. However, when f_1 lies above 2.5 kHz, so that f_2 would lie above 5 kHz, octave matches become very erratic [13.109]. It appears that the musical interval of an octave is only clearly perceived when both tones are below 5 kHz.

Other aspects of the perception of pitch also change above 5 kHz. A sequence of pure tones above 5 kHz does not produce a clear sense of melody [13.110]. It is possible to hear that the pitch changes when the frequency is changed, but the musical intervals are not heard clearly. Also, subjects with absolute pitch (the ability to assign names to notes without reference to other notes) are very poor at naming notes above 4–5 kHz [13.111].

These results are consistent with the idea that the pitch of pure tones is determined by different mechanisms above and below 5 kHz, specifically, by a temporal mechanism at low frequencies and a place mechanism at high frequencies. It appears that the perceptual dimension of tone height persists over the whole audible frequency range, but tone chroma only occurs in the frequency range below 5 kHz. Musical intervals are only clearly perceived when the frequencies of the tones lie in the range where temporal information is available.

The Effect of Level on Pitch

The pitch of a pure tone is primarily determined by its frequency. However, sound level also plays a small role. On average, the pitch of tones with frequencies below about 2 kHz decreases with increasing level, while the pitch of tones with frequencies above about 4 kHz increases with increasing sound level. The early data of *Stevens* [13.112] showed rather large effects of sound level on pitch, but other data generally show much smaller effects [13.113]. For tones with frequencies between 1 and 2 kHz, changes in pitch with level are generally less than 1%. For tones of lower and higher frequencies, the changes can be larger (up to 5%). There are also considerable individual differences both in the size of the pitch shifts with level, and in the direction of the shifts [13.113].

13.5.3 The Perception of the Pitch of Complex Tones

The Phenomenon of the Missing Fundamental

For complex tones the pitch does not, in general, correspond to the position of maximum excitation on the basilar membrane. Consider, as an example, a sound consisting of short impulses (clicks) occurring 200 times per second. This sound contains harmonics with frequencies at integer multiples of 200 Hz (200, 400, 600, 800 ... Hz). The harmonic at 200 Hz is called the fundamental frequency. The sound has a low pitch, which is very close to the pitch of its fundamental component (200 Hz), and a sharp timbre (a buzzy tone quality). However, if the sound is filtered so as to remove the fundamental component, the pitch does not alter; the only result is a slight change in timbre. This is called the phenomenon of the missing fundamental [13.99, 114]. Indeed, all except a small group of mid-frequency harmonics can be eliminated, and the low pitch remains the same, although the timbre becomes markedly different.

Schouten [13.99, 115] called the low pitch associated with a group of high harmonics the *residue*. Several other names have been used to describe this pitch, including *periodicity pitch*, *virtual pitch*, and *low pitch*. The term *low pitch* will be used here. Schouten pointed out that it is possible to hear the change produced by removing the fundamental component and then reintroducing it. Indeed, when the fundamental component is present, it is possible to hear it out as a separate sound. The pitch of that component is almost the same as the pitch of the whole sound. Therefore, the presence or absence of the fundamental component does not markedly affect the pitch of the whole sound.

The perception of a low pitch does not require activity at the point on the basilar membrane which would respond maximally to the fundamental component. *Licklider* [13.116] showed that the low pitch could be heard when low-frequency noise was present that would mask any component at the fundamental frequency. Even when the fundamental component of a complex tone is present, the pitch of the tone is usually determined by harmonics other than the fundamental.

The phenomenon of the missing fundamental is not consistent with a simple place model of pitch based on the idea that pitch is determined by the position of the peak excitation on the basilar membrane. However, more

Amazon
EX1013 Page 491

elaborate place models have been proposed, and these are discussed below.

Theories of Pitch Perception for Complex Tones

To understand theories of pitch perception for complex tones, it is helpful to consider how complex tones are represented in the peripheral auditory system. A simulation of the response of the basilar membrane to a complex tone is illustrated in Fig. 13.18. In this example, the complex tone is a regular series of brief pulses, whose spectrum contains many equal-amplitude harmonics. The number of pulses per second (also called the repetition rate) is 200, so the harmonics have frequencies that are integer multiples of 200 Hz. The lower harmonics are partly resolved on the basilar membrane, and give rise to distinct peaks in the pattern of activity along the basilar membrane. At a place tuned to the frequency of a low harmonic, the waveform on the basilar membrane is approximately a sinusoid at the harmonic frequency. For example, at the place with a characteristic frequency of 400 Hz the waveform is a 400 Hz sinusoid. At a place tuned between two low harmonics, e.g., the place tuned to 317 Hz, there is very little response. In contrast, the higher harmonics are not resolved, and do not give rise to distinct peaks on the basilar membrane. The waveforms at places on the basilar membrane responding to higher harmonics are complex, but they all have a repetition rate equal to the fundamental frequency of the sound.

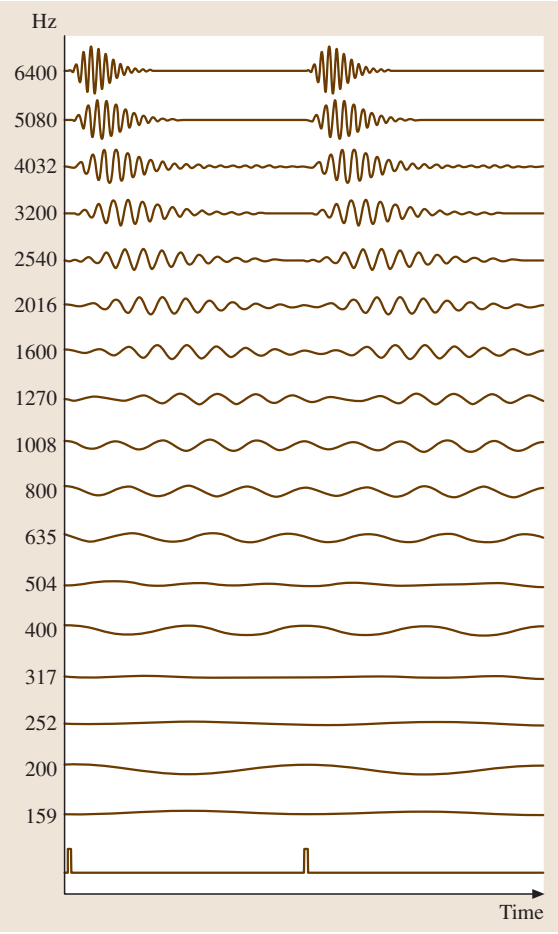


Fig. 13.18 A simulation of the responses on the basilar membrane to periodic impulses of rate 200 pulses per second. The input waveform is shown at the bottom; impulses occur every 5 ms. Each number on the left represents the frequency which would maximally excite a given point on the basilar membrane. The waveform which would be observed at that point, as a function of time, is plotted opposite that number

There are two main (nonexclusive) ways in which the low pitch of a complex sound might be extracted. Firstly, it might be derived from the frequencies of the lower harmonics that are resolved on the basilar membrane. The frequencies of the harmonics might be determined either by place mechanisms (e.g., from the positions of local maxima on the basilar membrane) or by temporal mechanisms (from the inter-spike intervals in neurons with CFs close to the frequencies of individual harmonics). For example, for the complex tone whose analysis is illustrated in Fig. 13.18, the second harmonic, with a frequency of 400 Hz, would give rise to a local maximum at the place on the basilar membrane tuned to 400 Hz. The inter-spike intervals in neurons innervating that place would reflect the frequency of that harmonic; the intervals would cluster around integer multiples of 2.5 ms. Both of these forms of information may allow the auditory system to determine that there is a harmonic at 400 Hz.

The auditory system may contain a pattern recognizer which determines the low pitch of the complex sound from the frequencies of the resolved components [13.117–119]. In essence the pattern recognizer tries to find the harmonic series giving the best match to the resolved frequency components; the fundamental frequency of this harmonic series determines the perceived pitch. Say, for example, that the initial analysis establishes frequencies of 800, 1000 and 1200 Hz to be present. The fundamental frequency whose harmonics would match these frequencies is 200 Hz. The perceived pitch corresponds to this inferred fundamental frequency of 200 Hz. Note that the inferred fundamen-

Amazon
EX1013 Page 492

tal frequency is always the highest possible value that fits the frequencies determined in the initial analysis. For example, a fundamental frequency of 100 Hz would also have harmonics at 800, 1000 and 1200 Hz, but a pitch corresponding to 100 Hz is *not* perceived. It should also be noted that when a complex tone contains only two or three low harmonics, some people, called *analytic listeners*, do not hear the low pitch, but rather hear pitches corresponding to individual harmonics [13.120, 121]. Others, called *synthetic listeners*, usually hear only the low pitch. When many harmonics are present, the low pitch is usually the dominant percept.

Evidence supporting the idea that the low pitch of a complex tone is derived by combining information from several harmonics comes from studies of the ability to detect changes in repetition rate (equivalent to the number of periods per second). When the repetition rate of a complex tone changes, all of the components change in frequency by the same *ratio*, and a change in low pitch is heard. The ability to detect such changes is better than the ability to detect changes in a sinusoid at the fundamental frequency [13.122] and it can be better than the ability to detect changes in the frequency of any of the individual sinusoidal components in the complex tone [13.123]. This indicates that information from the different harmonics is combined or integrated in the determination of low pitch. This can lead to very fine discrimination; changes in repetition rate of about 0.2% can be detected for fundamental frequencies in the range 100–400 Hz provided that low harmonics are present (e.g., the third, fourth and fifth).

The pitch of a complex tone may also be extracted from the higher unresolved harmonics. As shown in Fig. 13.18, the waveforms at places on the basilar membrane responding to higher harmonics are complex, but they all have a repetition rate equal to the fundamental frequency of the sound, namely 200 Hz. For the neurons with CFs corresponding to the higher harmonics, nerve impulses tend to be evoked by the biggest peaks in the waveform, i. e., by the waveform peaks close to envelope maxima. Hence, the nerve impulses are separated by times corresponding to the period of the sound. For example, in Fig. 13.18 the input has a repetition rate of 200 periods per second, so the period is 5 ms. The time intervals between nerve spike would cluster around integer multiples of 5 ms, i. e., 5, 10 15, 20 . . . ms. The pitch may be determined from these time intervals. In this example, the time intervals are integer multiples of 5 ms, so the pitch corresponds to 200 Hz.

Experimental evidence suggests that pitch can be extracted *both* from the lower harmonics and from the higher harmonics. Usually, the lower, resolved harmonics give a clearer low pitch, and are more important in determining low pitch, than the upper unresolved harmonics [13.124–126]. This idea is called the principle of dominance; when a complex tone contains many harmonics, including both low- and high-numbered harmonics, the pitch is mainly determined by a small group of lower harmonics. Also, the discrimination of changes in repetition rate of complex tones is better for tones containing only low harmonics than for tones containing only high harmonics [13.123, 127–129]. However, a low pitch can be heard when only high unresolvable harmonics are present. Although, this pitch is not as clear as when lower harmonics are present, it is clear enough to allow the recognition of musical intervals and of simple melodies [13.129, 130].

Several researchers have proposed theories in which both place (spectral) and temporal mechanisms play a role; these are referred to as spectro-temporal theories. The theories assume that information from both low harmonics and high harmonics contributes to the determination of pitch. The initial place/spectral analysis in the cochlea is followed by an analysis of the time pattern of the neural spikes evoked at each CF [13.131–135]. The temporal analysis is assumed to occur at a level of the auditory system higher than the auditory nerve, perhaps in the cochlear nucleus. In the model proposed by *Moore* [13.133], the sound is passed through an array of bandpass filters, each corresponding to a specific place on the basilar membrane. The time pattern of the neural impulses at each CF is determined by the waveform at the corresponding point on the basilar membrane. The inter-spike intervals at each CF are determined. Then, a device compares the time intervals present at different CFs, and searches for common time intervals. The device may also integrate information over time. In general the time interval which is found most often corresponds to the period of the fundamental component. The perceived pitch corresponds to the reciprocal of this interval. For example, if the most prominent time interval is 5 ms, the perceived pitch corresponds to a frequency of 200 Hz.

Amazon
EX1013 Page 493

13.6 Timbre Perception

13.6.1 Time-Invariant Patterns and Timbre

Almost all of the sounds that we encounter in everyday life contain a multitude of frequency components with particular levels and relative phases. The distribution of energy over frequency is one of the major determinants of the quality of a sound or its timbre. Timbre is usually defined as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" [13.10]. Differences in timbre enable us to distinguish between the same note played on, say, the piano, the violin or the flute.

Timbre depends upon more than just the frequency spectrum of the sound; fluctuations over time can play an important role (see the next section). For the purpose of this section we can adopt a more restricted definition suggested by *Plomp* [13.136] as applicable to steady tones: "Timbre is that attribute of sensation in terms of which a listener can judge that two steady complex tones having the same loudness, pitch and duration are dissimilar." Timbre defined in this way depends mainly on the relative magnitudes of the partials of the tones.

Timbre is multidimensional; there is no single scale along which the timbres of different sounds can be compared or ordered. Thus, a way is needed of describing the spectrum of a sound which takes into account this multidimensional aspect, and which can be related to the subjective timbre. A crude first approach is to look at the overall distribution of spectral energy. The *brightness* or *sharpness* [13.137] of sounds seems to be related to the spectral centroid. However, a much more quantitative approach has been described by *Plomp* and his colleagues [13.136, 138]. They showed that the perceptual differences between different sounds, such as vowels, or steady tones produced by musical instruments, were closely related to the differences in the spectra of the sounds, when the spectra were specified as the levels in 18 1/3-octave frequency bands. A bandwidth of 1/3 octave is slightly greater than the $\mathrm{ERB_N}$ of the auditory filter over most of the audible frequency range. Thus, timbre is related to the relative level produced at the output of each auditory filter. Put another way, the timbre of a sound is related to the excitation pattern of that sound.

It is likely that the number of dimensions required to characterize timbre is limited by the number of $\mathrm{ERB_N}$s required to cover the audible frequency range. This would give a maximum of about 37 dimensions. For a restricted class of sounds, however, a much smaller number of dimensions may be involved. It appears to be generally true, both for speech and non-speech sounds, that the timbres of steady tones are determined primarily by their magnitude spectra, although the relative phases of the components may also play a small role [13.139, 140].

13.6.2 Time-Varying Patterns and Auditory Object Identification

Differences in static timbre are not always sufficient to allow the absolute identification of an *auditory object*, such as a musical instrument. One reason for this is that the magnitude and phase spectrum of the sound may be markedly altered by the transmission path and room reflections [13.141]. In practice, the recognition of a particular timbre, and hence of an auditory object, may depend upon several other factors. *Schouten* [13.142] has suggested that these include: (1) whether the sound is periodic, having a tonal quality for repetition rates between about 20 and 20 000 per second, or irregular, and having a noise-like character; (2) whether the waveform envelope is constant, or fluctuates as a function of time, and in the latter case what the fluctuations are like; (3) whether any other aspect of the sound (e.g., spectrum or periodicity) is changing as a function of time; (4) what the preceding and following sounds are like.

The recognition of musical instruments, for example, depends quite strongly on onset transients and on the temporal structure of the sound envelope. The characteristic tone of a piano depends upon the fact that the notes have a rapid onset and a gradual decay. If a recording of a piano is reversed in time, the timbre is completely different. It now resembles that of a harmonium or accordion, in spite of the fact that the long-term magnitude spectrum is unchanged by time reversal. The perception of sounds with temporally asymmetric envelopes has been studied by *Patterson* [13.143, 144]. He used sinusoidal carriers that were amplitude modulated by a repeating exponential function. The envelope either increased abruptly and decayed gradually (damped sounds) or increased gradually and decayed abruptly (ramped sounds). The ramped sounds were time-reversed versions of the damped sounds and had the same long-term magnitude spectrum. The sounds were characterized by the repetition period of the envelope, which was 25 ms, and by the half

Amazon
EX1013 Page 494

life. For a damped sinusoid, the half life is the time taken for the amplitude to decrease by a factor of two.

Patterson reported that the ramped and damped sounds had different qualities. For a half life of 4 ms, the damped sound was perceived as a single source rather like a drum roll played on a hollow, resonant surface (like a drummer's wood block). The ramped sound was perceived as two sounds: a drum roll on a non-resonant surface (such as a leather table top) and a continuous tone corresponding to the carrier frequency. *Akeroyd* and *Patterson* [13.145] used sounds with similar envelopes, but the carrier was broadband noise rather than a sinusoid. They reported that the damped sound was heard as a drum struck by wire brushes. It did not have any hiss-like quality. In contrast, the ramped sound was heard as a noise, with a hiss-like quality, that was sharply cut off in time. These experiments clearly demonstrate the important role of temporal envelope in timbre perception

Pollard and *Jansson* [13.146] described a perceptually relevant way of characterizing the characteristics of time-varying sounds. The sound is filtered in bands of width 1/3 octave. The loudness of each band is calculated at 5 ms intervals. The loudness values are then converted into three coordinates, based on the loudness of:

1. the fundamental component,
2. a group containing partials 2–4, and
3. a group containing partials 5–n,

where n is the highest significant partial. This *tristimulus* representation appears to be related quite closely to perceptual judgements of musical sounds.

Many instruments have noise-like qualities which strongly influence their subjective quality. A flute, for example, has a relatively simple harmonic structure, but synthetic tones with the same harmonic structure do not sound flute-like unless each note is preceded by a small puff of noise. In general, tones of standard musical instruments are poorly simulated by the summation of steady component frequencies, since such a synthesis cannot produce the dynamic variation with time characteristic of these instruments. Thus traditional electronic organs (pre-1965), which produced only tones with a fixed envelope shape, could produce a good simulation of the bagpipes, but could not be made to sound like a piano. Modern synthesizers shape the envelopes of the sounds they produce, and hence are capable of more accurate and convincing imitations of musical instruments. For a simulation to be convincing, it is often necessary to give different time envelopes to different harmonics within a complex sound [13.97, 147].

13.7 The Localization of Sounds

13.7.1 Binaural Cues

It has long been recognized that slight differences in the sounds reaching the two ears can be used as cues in sound localization. The two major cues are differences in the time of arrival at the two ears and differences in intensity at the two ears. For example, a sound coming from the left arrives first at the left ear and is more intense in the left ear. For steady sinusoids, a difference in time of arrival is equivalent to a phase difference between the sounds at the two ears. However, phase differences are not usable over the whole audible frequency range. Experiments using sounds delivered by headphones have shown that a phase difference at the two ears can be detected and used to judge location only for frequencies below about 1500 Hz. This is reasonable because, at high frequencies, the wavelength of sound is small compared to the dimensions of the head, so the listener cannot determine which cycle in the left ear corresponds to a given cycle in the right; there may be many cycles of phase difference. Thus phase differences become ambiguous and unusable at high frequencies. On the other hand, at low frequencies our accuracy at detecting changes in relative time at the two ears is remarkably good; changes of 10–20 μs can be detected, which is equivalent to a movement of the sound source of 1–2° laterally when the sound comes from straight ahead [13.148].

Intensity differences are usually largest at high frequencies. This is because low frequencies bend or diffract around the head, so that there is little difference in intensity at the two ears whatever the location of the sound source, except when the source is very close to the head. At high frequencies, the head casts more of a shadow, and above 2–3 kHz the intensity differences provide useful cues. For complex sounds, containing a range of frequencies, the difference in spectral patterning at the two ears may also be important.

The idea that sound localization is based on interaural time differences at low frequencies and interaural intensity differences at high frequencies has been called

Amazon
EX1013 Page 495

the duplex theory of sound localization, proposed by Lord *Rayleigh* [13.149]. However, complex sounds, containing only high frequencies (above 1500 Hz), can be localized on the basis of interaural time delays, provided that they have an appropriate temporal structure. For example, a single click can be localized in this way no matter what its frequency content. Periodic sounds containing only high-frequency harmonics can also be localized on the basis of interaural time differences, provided that the envelope repetition rate (usually equal to the fundamental frequency) is below about 600 Hz [13.150, 151]. Since most of the sounds we encounter in everyday life are complex, and have repetition rates below 600 Hz, interaural time differences can be used for localization in most listening situations.

13.7.2 The Role of the Pinna and Torso

Binaural cues are not sufficient to account for all aspects of sound localization. For example, an interaural time or intensity difference will not indicate whether a sound is coming from in front or behind, or above or below, but such judgments can clearly be made. Also, under some conditions localization with one ear can be as accurate as with two. It has been shown that reflections of sounds from the pinnae and torso play an important role in sound localization [13.152, 153]. The spectra of sounds entering the ear are modified by these reflections in a way which depends upon the direction of the sound source. This direction-dependent filtering provides cues for sound source location. The spectral cues are important not just in providing information about the direction of sound sources, but also in enabling us to judge whether a sound comes from within the head or from the outside world. The pinnae alter the sound spectrum primarily at high frequencies. Only when the wavelength of the sound is comparable with or smaller than the dimensions of the pinnae is the spectrum significantly affected. This occurs mostly above about 6 kHz. Thus, cues provided by the pinnae are most effective for broadband high-frequency sounds. However, reflections from other structures, such as the shoulders, result in spectral changes at lower frequencies, and these may be important for front–back discrimination [13.154].

13.7.3 The Precedence Effect

In everyday conditions the sound from a given source reaches the ears by many different paths. Some of it arrives via a direct path, but a great deal may only reach the ears after reflections from one or more surfaces. However, listeners are not normally aware of these reflections, and the reflections do not markedly impair the ability to localize sound sources. The reason for this seems to lie in a phenomenon known as the precedence effect [13.155, 156]; for a review, see [13.157]. When several sounds reach the ears in close succession (i. e., the direct sound and its reflections) the sounds are perceptually fused into a single sound (an effect called echo suppression), and the location of the total sound is primarily determined by the location of the first (direct) sound (the precedence effect). Thus the reflections have little influence on the perception of direction. Furthermore, there is little awareness of the reflections, although they may influence the timbre and loudness of the sound.

The precedence effect only occurs for sounds of a discontinuous or transient character, such as speech or music, and it can break down if the reflections have a level 10 dB or more above that of the direct sound. However, in normal conditions the precedence effect plays an important role in the localization and identification of sounds in reverberant conditions.

13.8 Auditory Scene Analysis

It is hardly ever the case that the sound reaching our ears comes from a single source. Usually the sound arises from several different sources. However, usually we are able to decompose the mixture of sounds and to perceive each source separately. An auditory object can be defined as the percept of a group of successive and/or simultaneous sound elements as a coherent whole, appearing to emanate from a single source.

As discussed earlier, the peripheral auditory system acts as a frequency analyzer, separating the different frequency components in a complex sound. Somewhere in the brain, the internal representations of these frequency components have to be assigned to their appropriate sources. If the input comes from two sources, A and B, then the frequency components must be split into two groups; the components emanating from source A should be assigned to one source and the components emanating from source B should be assigned to another. The process of doing this is often called perceptual grouping. It is also given the name auditory scene anal-

Amazon
EX1013 Page 496

ysis [13.158]. The process of separating the elements arising from two or more different sources is sometimes called segregation.

Many different physical cues may be used to derive separate perceptual objects corresponding to the individual sources which give rise to a complex acoustic input. There are two aspects to this process: the grouping together of all the simultaneous frequency components that emanate from a single source at a given moment, and the connecting over time of the changing frequencies that a single source produces from one moment to the next [13.159]. These two aspects are sometimes described as *simultaneous grouping* and *sequential grouping*, respectively.

Most experiments on perceptual grouping have studied the effect of grouping on one specific attribute of sounds, for example, their pitch, their subjective location, or their timbre. These experiments have shown that a cue which is effective for one attribute may be less effective or completely ineffective for another attribute [13.160]. Also, the effectiveness of the cues may differ for simultaneous and sequential grouping.

13.8.1 Information Used to Separate Auditory Objects

Fundamental Frequency and Spectral Regularity

When we listen to two steady complex tones together (e.g., two musical instruments or two vowel sounds), we do not generally confuse which harmonics belong to which tone. If the complex tones overlap spectrally, two sounds are heard only if the two tones have different values of fundamental frequency (F_0) [13.161]. *Scheffers* [13.162] has shown that, if two vowels are presented simultaneously, they can be identified better when they have F_0s that differ by more than 6% than when they have the same F_0. Other researchers have reported similar findings [13.163, 164].

F_0 may be important in several ways. The components in a periodic sound have frequencies which form a simple harmonic series; the frequencies are integer multiples of F_0. This property is referred to as harmonicity. The lower harmonics are resolved in the peripheral auditory system. The regular spacing of the lower harmonics may promote their perceptual fusion, causing them to be heard as a single sound. If a sinusoidal component does not form part of this harmonic series, it tends to be heard as a separate sound. This is illustrated by some experiments of *Moore* et al.[13.165]. They investigated the effect of mistuning a single low harmonic in a harmonic complex tone. When the harmonic was mistuned sufficiently, it was heard as a separate pure tone standing out from the complex as a whole. The degree of mistuning required varied somewhat with the duration of the sounds; for 400 ms tones, a mistuning of 3% was sufficient to make the harmonic stand out as a separate tone.

Roberts and *Brunstrom* [13.166,167] have suggested that the important feature determining whether a group of frequency components sounds fused is not harmonicity, but spectral regularity; if a group of components form a regular spectral pattern, they tend to be heard as fused, while if a single component does not fit the pattern, it is heard to pop out. For example, a sequence of components with frequencies 623, 823, 1023, 1223, and 1423 Hz is heard as relatively fused. If the frequency of the middle component is shifted to, say, 923 or 1123 Hz, that component no longer forms part of the regular pattern, and it tends to be heard as a separate tone, standing out from the complex.

For the higher harmonics in a complex sound, F_0 may play a different role. The higher harmonics of a periodic complex sound are not resolved on the basilar membrane, but give rise to a complex waveform with a periodicity equal to F_0 (see Fig. 13.18). When two complex sounds with different F_0s are presented simultaneously, then each will give rise to a waveform on the basilar membrane with periodicity equal to its respective F_0. If the two sounds have different spectra, then each will dominate the response at certain points on the basilar membrane. The auditory system may group together regions with common F_0 and segregate them from regions with different F_0 [13.163]. It may also be the case that both resolved and unresolved components can be grouped on the basis of the detailed time pattern of the neural spikes [13.168].

This process can be explained in a qualitative way by extending the model of pitch perception presented earlier. Assume that the pitch of a complex tone results from a correlation or comparison of time intervals between successive nerve firings in neurons with different CFs. Only those channels which show a high correlation would be classified as belonging to the same sound. Such a mechanism would automatically group together components with a common F_0. However, *de Cheveigné* et al. [13.169] presented evidence against such a mechanism. They measured the identification of a target vowel in the presence of a background vowel; the nominal fundamental frequencies of the two vowels differed by 6.45%. Identification was better when the background was harmonic than when it was made in-

Amazon
EX1013 Page 497

harmonic (by shifting the frequency of each harmonic by a random amount between 0 and 6.45% or by less than half of the spacing between harmonics, whichever was smaller). In contrast, identification of the target did not depend upon whether or not the target was harmonic. *De Cheveigné* and coworkers [13.169–171] proposed a mechanism based on the idea that a harmonic background sound can be canceled in the auditory system, thus enhancing the representation of a target vowel.

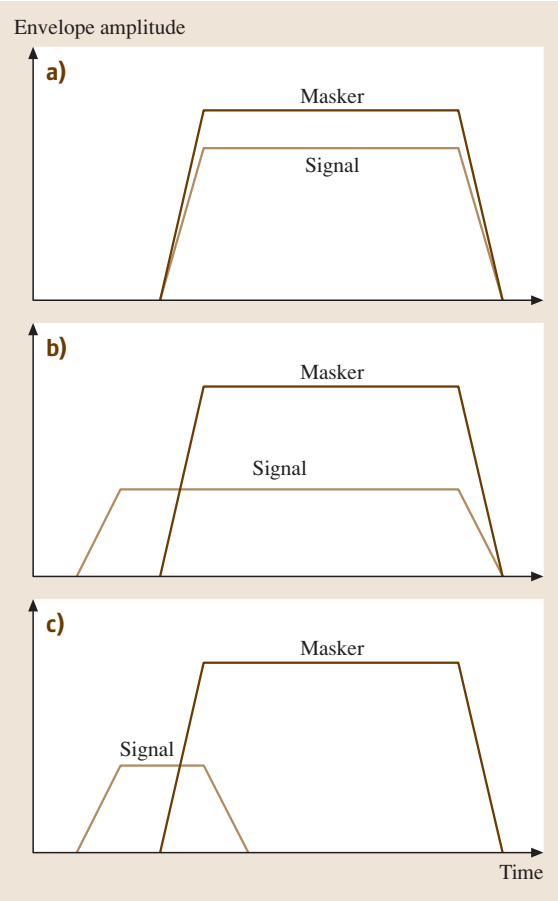


Fig. 13.19a–c Schematic illustration of the stimuli used by *Rasch* [13.172]. Both the signal and the masker were periodic complex tones, with the signal having the higher fundamental frequency. When the signal and masker were gated on and off synchronously **(a)**, the threshold for the signal was relatively high. When the signal started slightly before the masker **(b)**, the threshold was markedly reduced. When the signal was turned off as soon as the masker was turned on **(c)**, the signal was perceived as continuing through the masker, and the threshold was the same as when the signal did continue through the masker

Onset and Offset Disparities

Another cue for the perceptual separation of (near-)simultaneous sounds is onset and offset disparity. *Rasch* [13.172] investigated the ability to hear one complex tone in the presence of another. One of the tones was treated as a masker and the level of the signal tone (the higher in F_0) was adjusted to find the point where it was just detectable. When the two tones started at the same time and had exactly the same temporal envelope, the threshold of the signal was between 0 and −20 dB relative to the level of the masker (Fig. 13.19a). Thus, when a difference in F_0 was the only cue, the signal could not be heard when its level was more than 20 dB below that of the masker.

Rasch also investigated the effect of starting the signal just before the masker (Fig. 13.19b). He found that threshold depended strongly on onset asynchrony, reaching a value of −60 dB for an asynchrony of 30 ms. Thus, when the signal started 30 ms before the masker, it could be heard much more easily and with much greater differences in level between the two tones. It should be emphasized that the lower threshold was a result of the signal occurring for a brief time on its own; essentially performance was limited by backward masking of the 30 ms asynchronous segment, rather than by simultaneous masking. However, the experiment does illustrate the large benefit that can be obtained from a relatively small asynchrony.

Although the percept of his subjects was that the signal continued throughout the masker, Rasch showed that this percept was not based upon sensory information received during the presentation time of the masker. He found that identical thresholds were obtained if the signal terminated immediately after the onset of the masker (Fig. 13.19c). It appears that the perceptual system "assumes" that the signal continues, since there is no evidence to the contrary; the part of the signal that occurs simultaneously with the masker would be completely masked.

Rasch [13.172] showed that, if the two tones have simultaneous onsets but different rise times, this also can give very low thresholds for the signal, provided it has the shorter rise time. Under these conditions and those of onset asynchronies up to 30 ms, the notes sound as though they start synchronously. Thus, we do not need to be consciously aware of the onset differences for the auditory system to be able to exploit them in the perceptual separation of complex tones. Rasch also pointed out

Amazon
EX1013 Page 498

that, in ensemble music, different musicians do not play exactly in synchrony even if the score indicates that they should. The onset differences used in his experiments correspond roughly with the onset asynchronies of nominally simultaneous notes found in performed music. This supports the view that the asynchronies are an important factor in the perception of the separate parts or voices in polyphonic music.

Onset asynchronies can also play a role in determining the timbre of complex sounds. *Darwin* and *Sutherland* [13.173] showed that a tone that starts or stops at a different time from a vowel is less likely to be heard as part of that vowel than if it is simultaneous with it. For example, increasing the level of a single harmonic can produce a significant change in the quality (timbre) of a vowel. However, if the incremented harmonic starts before the vowel, the change in vowel quality is markedly reduced. Similarly, *Roberts* and *Moore* [13.174] showed that extraneous sinusoidal components added to a vowel could influence vowel quality, but the influence was markedly reduced when the extraneous components were turned on before the vowel or turned off after the vowel.

Contrast with Previous Sounds

The auditory system seems well suited to the analysis of changes in the sensory input, and particularly to changes in spectrum over time. The changed aspect stands out perceptually from the rest. It is possible that there are specialized central mechanisms for detecting changes in spectrum. Additionally, stimulation with a steady sound may result in some kind of adaptation. When some aspect of a stimulus is changed, that aspect is freed from the effects of adaptation and thus will be enhanced perceptually. While the underlying mechanism is a matter of debate, the perceptual effect certainly is not.

A powerful demonstration of this effect may be obtained by listening to a stimulus with a particular spectral structure and then switching rapidly to a stimulus with a flat spectrum, such as white noise. A white noise heard in isolation may be described as colorless; it has no pitch and has a neutral sort of timbre. However, when a white noise follows soon after a stimulus with spectral structure, the noise sounds colored [13.175]. A harmonic complex tone with a flat spectrum may be given a speech-like quality if it is preceded by a harmonic complex having a spectrum which is the inverse of that of a speech sound, such as a vowel [13.176].

Another demonstration of the effects of a change in a stimulus can be obtained by listening to a steady complex tone with many harmonics. Usually such a tone is heard with a single pitch corresponding to F_0, and the individual harmonics are not separately perceived. However, if one of the harmonics is changed in some way, by altering either its relative phase or its level, then that harmonic stands out perceptually from the complex as a whole. For a short time after the change is made, a pure-tone quality is perceived. The perception of the harmonic then gradually fades, until it merges with the complex once more.

Change detection is obviously of importance in assigning sound components to their appropriate sources. Normally we listen against a background of sounds which may be relatively unchanging, such as the humming of machinery, traffic noises, and so on. A sudden change in the sound is usually indicative that a new source has been activated, and the change detection mechanisms enable us to isolate the effects of the change and interpret them appropriately.

Correlated Changes in Amplitude or Frequency

Rasch [13.172] also showed that, when the two complex tones start and end synchronously, the detection of the tone with the higher F_0 could be enhanced by frequency modulating it. The modulation was similar to the vibrato which often occurs for musical tones, and it was applied so that all the components in the higher tone moved up and down in synchrony. Rasch found that the modulation could reduce the threshold for detecting the higher tone by 17 dB. A similar effect can be produced by amplitude modulation (AM) of one of the tones. The modulation seems to enhance the salience of the modulated sound, making it appear to stand out from the unmodulated sound.

It is less clear whether the perceptual segregation of simultaneous sounds is affected by the coherence of changes in amplitude or frequency when *both* sounds are modulated. Coherence here refers to whether the changes in amplitude or frequency of the two sounds have the same pattern over time or different patterns over time. Several experiments have been reported suggesting that coherence of amplitude changes plays a role; sounds with coherent changes tend to fuse perceptually, whereas sounds with incoherent changes tend to segregate [13.177–180]. However, *Summerfield* and *Culling* [13.181] found that the coherence of AM did not affect the identification of pairs of simultaneous vowels when the vowels were composed of components placed randomly in frequency (to avoid effects of harmonicity).

Evidence for a role of frequency modulation (FM) coherence in perceptual grouping has been more elu-

Amazon
EX1013 Page 499

sive. While some studies have been interpreted as indicating a weak role for frequency modulation coherence [13.182, 183], the majority of studies have failed to indicate such sensitivity [13.181, 184–188]. *Furukawa* and *Moore* [13.189] have shown that the detectability of FM imposed on two widely separated carrier frequencies is better when the modulation is coherent on the two carriers than when it is incoherent. However, this may occur because the overall pitch evoked by the two carriers fluctuates more when the carriers are modulated coherently than when they are modulated incoherently [13.190–192]. There is at present no clear evidence that the coherence of FM influences perceptual grouping when both sounds are modulated.

Sound Location

The cues used in sound localization may also help in the analysis of complex auditory inputs. A phenomenon that is related to this is called the binaural masking level difference (MLD). The phenomenon can be summarized as follows: whenever the phase or level differences of a signal at the two ears are not the same as those of a masker, the ability to detect the signal is improved relative to the case where the signal and masker have the same phase and level relationships at the two ears. The practical implication is that a signal is easier to detect when it is located in a different position in space from the masker. Although most studies of the MLD have been concerned with threshold measurements, it seems clear that similar advantages of binaural listening can be gained in the identification and discrimination of signals presented against a background of other sound.

An example of the use of binaural cues in separating an auditory object from its background comes from an experiment by *Kubovy* et al. [13.193]. They presented eight continuous sinusoids to each ear via earphones. The sinusoids had frequencies corresponding to the notes in a musical scale, the lowest having a frequency of 300 Hz. The input to one ear, say the left, was presented with a delay of 1 ms relative to the input to the other ear, so the sinusoids were all heard toward the right side of the head. Then, the phase of one of the sinusoids was advanced in the left ear, while its phase in the right ear was delayed, until the input to the left ear led the input to the right ear by 1 ms; this phase-shifting process occurred over a time of 45 ms. The phase remained at the shifted value for a certain time and was then smoothly returned to its original value, again over 45 ms. During the time that the phase shift was present, the phase-shifted sinusoid appeared toward the opposite (left) side of the head, making it stand out perceptually. A sequence of phase shifts in different components was clearly heard as a melody. This melody was completely undetectable when listening to the input to one ear alone. Kubovy et al. interpreted their results as indicating that differences in relative phase at the two ears can allow an auditory object to be isolated in the absence of any other cues.

Culling [13.194] performed a similar experiment to that of *Kubovy* et al. [13.193], but he examined the importance of the phase transitions. He found that, when one component of a complex sound changed rapidly but smoothly in interaural time difference (ITD), it perceptually segregated from the complex. When different components were changed in ITD in succession, a recognizable melody was heard, as reported by Kubovy et al. However, when the transitions were replaced by silent intervals, leaving only static ITDs as a cue, the melody was much less salient. Thus, transitions in ITD seem to be more important than static differences in ITD in producing segregation of one component from a background of other components. Nevertheless, static differences in ITD do seem to be sufficient to produce segregation under some conditions [13.195].

Other experiments suggest that binaural processing often plays a relatively minor role in simultaneous grouping. *Shackleton* and *Meddis* [13.196] investigated the ability to identify each vowel in pairs of concurrent vowels. They found that a difference in F_0 between the two vowels improved identification by about 22%. In contrast, a 400 μs interaural delay in one vowel (which corresponds to an azimuth of about 45 degrees) improved performance by only 7%. *Culling* and *Summerfield* [13.197] investigated the identification of concurrent whispered vowels, synthesized using bands of noise. They showed that listeners were able to identify the vowels accurately when each vowel was presented to a different ear. However, they were unable to identify the vowels when they were presented to both ears but with different ITDs. In other words, listeners could not group the noise bands in different frequency regions with the same ITD and thereby separate them from noise bands in other frequency regions with a different ITD.

In summary, when two simultaneous sounds differ in their interaural level or time, this can contribute to the perceptual segregation of the sounds and enhance their detection and discrimination. However, such binaural processing is not always effective, and in some situations it appears to play little role.

Amazon
EX1013 Page 500

13.8.2 The Perception of Sequences of Sounds

Stream Segregation

When we listen to rapid sequences of sounds, the sounds may be grouped together (i. e., perceived as if they come from a single source, called fusion or coherence), or they may be perceived as different streams (i. e., as coming from more than one source, called fission or stream segregation) [13.158, 199–201]. The term streaming is used to denote the processes determining whether one stream or multiple streams are heard. *Van Noorden* [13.202] investigated this phenomenon using a sequence of pure tones where every second B was omitted from the regular sequence ABABAB . . . , producing a sequence ABA ABA . . . . He found that this could be perceived in two ways, depending on the frequency separation of A and B. For small separations, a single rhythm, resembling a gallop, is heard (fusion). For larger separations, two separate tone sequences can be heard, one of which (A A A) is running twice as fast as the other (B B B) (fission). Components are more likely to be assigned to separate streams if they differ widely in frequency or if there are rapid jumps in frequency between them. The latter point is illustrated by a study of *Bregman* and *Dannenbring* [13.203]. They used tone sequences in which successive tones were connected by frequency glides. They found that these glides reduced the tendency for the sequences to split into high and low streams. Conditions using partial glides also showed decreased stream segregation, although the partial glides were not quite as effective as complete glides. Thus, complete continuity between tones is not required to reduce stream segregation; a frequency change pointing toward the next tone allows the listener to follow the pattern more easily.

The effects of frequency glides and other types of transitions in preventing stream segregation or fission are probably of considerable importance in the perception of speech. Speech sounds may follow one another in very rapid sequences, and the glides and partial glides observed in the acoustic components of speech may be a strong factor in maintaining the percept of speech as a unified stream.

Van Noorden found that, for intermediate frequency separations of the tones A and B in a rapid sequence, either fusion or fission could be heard, according to the instructions given and the attentional set of the subject. When the percept is ambiguous, the tendency for fission to occur increases with increasing exposure time to the tone sequence [13.204]. The auditory system seems to start with the assumption that there is a single sound source, and fission is only perceived when sufficient evidence has built up to contradict this assumption. Sudden changes in a sequence, or in the perception of a sequence, can cause the percept to revert to its initial, default condition, which is fusion [13.205, 206]; for a review, see [13.207].

For rapid sequences of complex tones, strong fission can be produced by differences in spectrum of successive tones, even when all tones have the same F_0 [13.201, 208–210]. However, when successive complex tones are filtered to have the same spectral envelope, stream segregation can also be produced by differences between successive tones in F_0 [13.210, 211], in temporal envelope [13.209, 212], in the relative phases of the components [13.213], or in apparent location [13.214]. *Moore* and *Gockel* [13.207] proposed that any salient perceptual difference between successive tones may lead to stream segregation. Consistent with this idea, *Dowling* [13.215, 216] has shown that stream segregation may also occur when successive pure tones differ in intensity or in spatial location. He presented a melody composed of equal-intensity notes and inserted between each note of the melody a tone of the same intensity, with a frequency randomly selected from the same range. He found that the resulting tone sequence produced a meaningless jumble. Making the interposed notes different from those of the melody, in either intensity, frequency range, or spatial location, caused them to be heard as a separate stream, enabling subjects to pick out the melody.

Darwin and *Hukin* [13.198] have shown that sequential grouping can be strongly influenced by ITD. In one experiment, they simultaneously presented two

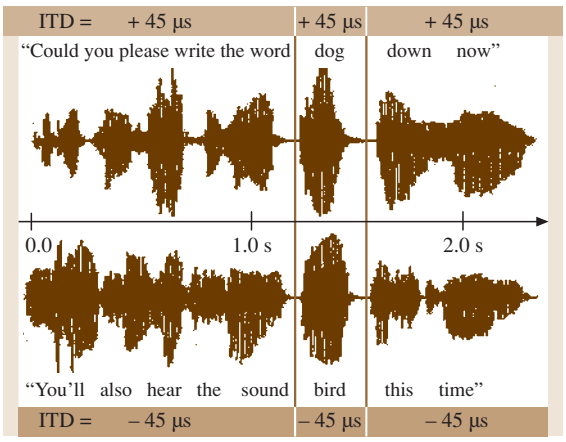


Fig. 13.20 Example of the stimuli used by *Darwin* and *Hukin* [13.198] (After Fig. 1 of [13.198])

Amazon
EX1013 Page 501

sentences (see Fig. 13.20). They varied the ITDs of the two sentences in the range 0 to ±181 μs. For example, one sentence might lead in the left ear by 45 μs, while the other sentence would lead in the right ear by 45 μs (as in Fig. 13.20). The sentences were based on natural speech but were processed so that each was spoken on a monotone, i. e., with constant F_0. The F_0 difference between the two sentences was varied from 0 to 4 semitones. Subjects were instructed to attend to one particular sentence. At a certain point, the two sentences contained two different target words aligned in starting time and duration ("dog" and "bird"). The F_0s and the ITDs of the two target words were varied independently from those of the two sentences. Subjects had to indicate which of the two target words they heard in the attended sentence. They reported the target word that had the same ITD as the attended sentence much more often than the target word with the opposite ITD. In other words, the target word with the same ITD as the attended sentence was grouped with that sentence. This was true even when the target word had the same ITD as the attended sentence but a different F_0. Thus, subjects grouped words across time according to their perceived location, independent of F_0 differences. *Darwin* and *Hukin* [13.198] concluded that listeners who try to track a particular sound source over time direct attention to auditory objects at a particular subjective location. The auditory objects themselves may be formed using cues other than ITD, for example, onset and offset asynchrony and harmonicity.

It should be noted that, for discrete sequences of musical tones, the auditory system does not necessarily form streams according to perceived location, especially when that cue competes with other cues. This is illustrated by an effect, called the scale illusion, reported by *Deutsch* [13.217]. She presented two sequences of tones via headphones, one sequence to each ear. The *n*th tone in the left ear was synchronous with the *n*th tone in the right ear. The sequences were created by repetitive presentation of the C major scale in both ascending and descending form, such that when a component of the ascending scale was in one ear, a component of the descending scale was in the other, and vice versa. However, the tones from each scale alternated between ears. Within each ear there were often large jumps in frequency between successive tones. Most subjects perceived the sounds as two streams, organized by the frequency proximity of successive tones. One stream (which was often heard towards one ear) was heard as a musical scale that started high, descended and then increased again, while the other stream (which was usually heard towards the opposite ear) was heard as a scale that started low, ascended, and then decreased again. Thus, the true location of the tones had little influence on the formation of the perceptual streams.

Another example come from the opening bars of the last movement of Tchaikovsky's sixth symphony. This contains interleaved notes played by the first and second violins, who according to 19th century custom sat on opposite sides of the stage. These notes are perceived as a single stream, despite the difference in location, presumably because of the frequency proximity between successive notes.

A number of composers have exploited the fact that stream segregation occurs for tones that are widely separated in frequency. By playing a sequence of tones in which alternate notes are chosen from separate frequency ranges, an instrument such as the flute, which is only capable of playing one note at a time, can appear to be playing two themes at once. Many fine examples of this are available in the works of Bach, Telemann and Vivaldi.

Judgment of Temporal Order

It is difficult to judge the temporal order of sounds that are perceived in different streams. An example of this comes from the work of *Broadbent* and *Ladefoged* [13.218]. They reported that extraneous sounds in sentences were grossly mislocated. For example, a click might be reported as occurring a word or two away from its actual position. Surprisingly poor performance was also reported by *Warren* et al. [13.219] for judgments of the temporal order of three or four unrelated items, such as a hiss, a tone, and a buzz. Most subjects could not identify the order when each successive item lasted as long as 200 ms. Naive subjects required that each item last at least 700 ms to identify the order of four sounds presented in an uninterrupted repeated sequence. These durations are well above those which are normally considered necessary for temporal resolution.

The poor order discrimination described by Warren et al. is probably a result of stream segregation. The sounds they used do not represent a coherent class. They have different temporal and spectral characteristics, and, as for tones widely differing in frequency, they do not form a single perceptual stream. Items in different streams appear to float about with respect to each other in subjective time. Thus, temporal order judgments are difficult. It should be emphasized that the relatively poor performance reported by *Warren* et al. [13.219] is found only in tasks requiring absolute identification of the order of sounds and not in tasks which simply require the discrimination of different sequences. Also, with

Amazon
EX1013 Page 502

extended training and feedback subjects can learn to distinguish between and identify orders within sequences of unrelated sounds lasting only 10 ms or less [13.220].

To explain these effects, *Divenyi* and *Hirsh* [13.221] suggested that two kinds of perceptual judgments are involved. At longer item durations the listener is able to hear a clear sequence of steady state sounds, whereas at shorter durations a change in the order of items introduces qualitative changes that can be discriminated by trained listeners. Similar explanations have been put forward by *Green* [13.75] and *Warren* [13.220].

Bregman and *Campbell* [13.200] investigated the factors that make temporal order judgments for tone sequences difficult. They used naive subjects, so performance presumably depended on the subjects actually perceiving the sounds as a sequence, rather than on their learning the overall sound pattern. They found that, in a repeating cycle of mixed high and low tones, subjects could discriminate the order of the high tones relative to one another or of the low tones among themselves, but they could not order the high tones relative to the low ones. The authors suggested that this was because the two groups of sounds split into separate perceptual streams and that judgments across streams are difficult. Several more recent studies have used tasks involving the discrimination of changes in timing or rhythm as a tool for studying stream segregation [13.210, 213, 222]. The rationale of these studies is that, if the ability to judge the relative timing of successive sound elements is good, this indicates that the elements are perceived as part of a single stream, while if the ability is poor, this indicates that the elements are perceived in different streams.

13.8.3 General Principles of Perceptual Organization

The Gestalt psychologists [13.223] described many of the factors which govern perceptual organization, and their descriptions and principles apply reasonably well to the way physical cues are used to achieve perceptual grouping of the acoustic input. It seems likely that the rules of perceptual organization have arisen because, on the whole, they tend to give the right answers. That is, use of the rules generally results in a grouping of those parts of the acoustic input that arose from the same source and a segregation of those that did not. No single rule will always work, but it appears that the rules can generally be used together, in a coordinated and probably quite complex way, in order to arrive at a correct interpretation of the input. In the following sections, I outline the major principles or rules of perceptual organization. Many, but not all, of the rules apply to both vision and hearing, and they were mostly described first in relation to vision.

Similarity

This principle is that elements will be grouped if they are similar. In hearing, similarity usually implies closeness of timbre, pitch, loudness, or subjective location. Examples of this principle have already been described. If we listen to a rapid sequence of pure tones, say 10 tones per second, then tones which are closely spaced in frequency, and are therefore similar, form a single perceptual stream, whereas tones which are widely spaced form separate streams.

For pure tones, frequency is the most important factor governing similarity, although differences in level and subjective location between successive tones can also lead to stream segregation. For complex tones, differences in timbre produced by spectral differences seem to be the most important factor. Again, however, other factors may play a role. These include differences in F_0, differences in timbre produced by temporal envelope differences, and differences in perceived location.

Good Continuation

This principle exploits a physical property of sound sources, that changes in frequency, intensity, location, or spectrum tend to be smooth and continuous, rather than abrupt. Hence, a smooth change in any of these aspects indicates a change within a single source, whereas a sudden change indicates that a new source has been activated. One example has already been described; *Bregman* and *Dannenbring* [13.203] showed that the tendency of a sequence of high and low tones to split into two streams was reduced when successive tones were connected by frequency glides.

A second example comes from studies using synthetic speech. In such speech, large fluctuations of an unexpected kind in F_0 (and correspondingly in the pitch) give the impression that a new speaker has stepped in to take over a few syllables from the primary speaker. *Darwin* and *Bethell-Fox* [13.224] synthesized spectral patterns which changed smoothly and repeatedly between two vowel sounds. When the F_0 of the sound patterns was constant they were heard as coming from a single source, and the speech sounds heard included glides ("l" as in let) and semivowels ("w" as in we). When a discontinuous, step-like F_0 contour was imposed on the patterns, they were perceived as two distinct speech streams, and the speech was perceived as containing predominantly stop consonants (e.g., "b" as in be, and "d" as in day). Apparently, a given group of

Amazon
EX1013 Page 503

components is usually only perceived as part of one stream. Thus, the perceptual segregation produces illusory silences in each stream during the portions of the signal attributed to the other stream, and these silences are interpreted, together with the gliding spectral patterns in the vowels, as indicating the presence of stop consonants. It is clear that the perception of speech sounds can be strongly influenced by stream organization.

Common Fate

The different frequency components arising from a single sound source usually vary in a highly coherent way. They tend to start and finish together, change in intensity together, and change in frequency together. This fact is exploited by the perceptual system and gives rise to the principle of common fate: if two or more components in a complex sound undergo the same kinds of changes at the same time, then they are grouped and perceived as part of the same source.

Two examples of common fate were described earlier. The first concerns the role of the onsets and offsets of sounds. Components will be grouped together if they start and stop synchronously; otherwise they will form separate streams. The onset asynchronies necessary to allow the separation of two complex tones are not large, about 30 ms being sufficient. The asynchronies which are observed in performed music are typically as large as or larger than this, so when we listen to polyphonic music we are easily able to hear separately the melodic line of each instrument. Secondly, components which are amplitude modulated in a synchronous way tend to be grouped together. There is at present little evidence that the coherence of modulation in frequency affects perceptual grouping, although frequency modulation of a group of components in a complex sound can promote the perceptual segregation of those components from an unchanging background.

Disjoint Allocation

Broadly speaking, this principle, also known as belongingness, is that a single component in a sound can only be assigned to one source at a time. In other words, once a component has been used in the formation of one stream, it cannot be used in the formation of a second stream. For certain types of stimuli, the perceptual organization may be ambiguous, there being more than one way to interpret the sensory input. When a given component might belong to one of a number of streams, the percept may alter depending on the stream within which that component is included.

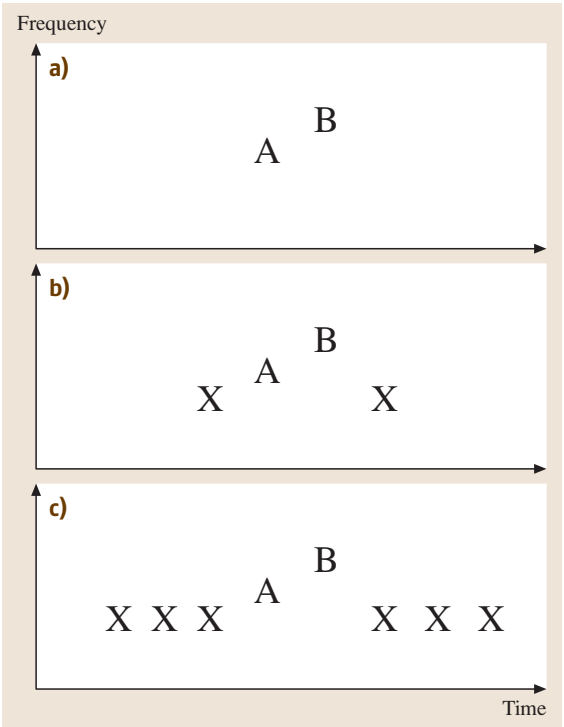


Fig. 13.21 Schematic illustration of the stimuli used by *Bregman* and *Rudnicky* [13.225]. When the tones A and B are presented alone *(panel a)*, it is easy to tell their order. When the tones A and B are presented as part of a four-tone complex XABX *(panel b)*, it is more difficult to tell their order. If the four-tone complex is embedded in a longer sequence of X tones *(panel c)*, the Xs form a separate perceptual stream, and it is easy to tell the order of A and B

An example is provided by the work of *Bregman* and *Rudnicky* [13.225]. They presented a sequence of four brief tones in rapid succession. Two of the tones, labeled X, had the same frequency, but the middle two, A and B, were different. The four-tone sequence was either XABX or XBAX. The listeners had to judge the order of A and B. This was harder than when the tones AB occurred in isolation (Fig. 13.21a) because A and B were perceived as part of a longer four-tone pattern, including the two "distracter" tones, labeled X (Fig. 13.21b). They then embedded the four-tone sequence into a longer sequence of tones, called captor tones (Fig. 13.21c). When the captor tones had frequencies close to those of the distracter tones, they captured the distracters into a separate perceptual stream, leaving the tones AB in a stream of their own. This made the order of A and B easy to

Amazon
EX1013 Page 504

judge. It seems that the tones X could not be perceived as part of both streams. When only one stream is the subject of judgment and hence of attention, the other one may serve to remove distracters from the domain of attention.

It should be noted that the principle of disjoint allocation does not always work, particularly in situations where there are two or more plausible perceptual organizations [13.226]. In such situations, a sound element may sometimes be heard as part of more than one stream.

Closure

In everyday life, the sound from a given source may be temporarily masked by other sounds. While the masking sound is present there may be no sensory evidence which can be used to determine whether the masked sound has continued or not. Under these conditions the masked sound tends to be perceived as continuous. The Gestalt psychologists called this process closure.

A laboratory example of this phenomenon is the continuity effect [13.227–229]. When a sound A is alternated with a sound B, and B is more intense than A, then A may be heard as continuous, even though it is interrupted. The sounds do not have to be steady. For example, if B is noise and A is a tone which is gliding upward in frequency, the glide is heard as continuous even though certain parts of the glide are missing [13.230]. Notice that, for this to be the case, the gaps in the tone must be filled with noise and the noise must be a potential masker of the tone (if they were presented simultaneously). In the absence of noise, discrete jumps in frequency are clearly heard.

The continuity effect also works for speech stimuli alternated with noise. In the absence of noise to fill in the gaps, interrupted speech sounds hoarse and raucous. When noise is presented in the gaps, the speech sounds more natural and continuous [13.231]. For connected speech at moderate interruption rates, the intervening noise actually leads to an improvement in intelligibility [13.232]. This may occur because the abrupt switching of the speech produces misleading cues as to which speech sounds were present. The noise serves to mask these misleading cues.

It is clear from these examples that the perceptual filling in of missing sounds does not take place solely on the basis of evidence in the acoustic waveform. Our past experience with speech, music, and other stimuli must play a role, and the context of surrounding sounds is important [13.233]. However, the filling in only occurs when one source is perceived as masking or occluding another. This percept must be based on acoustic evidence that the occluded sound has been masked. Thus, if a gap is not filled by a noise or other sound, the perceptual closure does not occur; a gap is heard.

13.9 Further Reading and Supplementary Materials

More information about the topics discussed in this chapter can be found in: A. S. Bregman: *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge 1990); W. M. Hartmann: *Signals, Sound, and Sensation* (AIP Press, Woodbury 1997); B. C. J. Moore: *An Introduction to the Psychology of Hearing, 5th Ed.* (Academic, San Diego 2003); R. Plomp: *The Intelligent Ear* (Erlbaum, Mahwah 2002)

A compact disc (CD) of auditory demonstrations has been produced by A. J. M. Houtsma, T. D. Rossing, W. M. Wagenaars (1987). The disc can be obtained through the Acoustical Society of America; contact asapubs@abdintl.com for details.

The following CD has a large variety of demonstrations relevant to perceptual grouping: A. S. Bregman, P. Ahad (1995). *Demonstrations of Auditory Scene Analysis: The Perceptual Organization of Sound*, (Auditory Perception Laboratory, Department of Psychology, McGill University, distributed by MIT Press, Cambridge, MA). It can be ordered from The MIT Press, 55 Hayward St., Cambridge, MA 02142, USA. Further relevant demonstrations can be found at: http://www.kyushu-id.ac.jp/~ynhome/.

A CD simulating the effects of a hearing loss on the perception of speech and music is *Perceptual Consequences of Cochlear Damage*, which may be obtained by writing to B. C. J. Moore, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, England, and enclosing a check for 20 dollars or a cheque for 12 pounds sterling, made payable to B. C. J. Moore.

Amazon
EX1013 Page 505

References

13.1 ISO 389-7: *Acoustics – Reference zero for the calibration of audiometric equipment. Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions* (International Organization for Standardization, Geneva 1996)

13.2 M.C. Killion: Revised estimate of minimal audible pressure: Where is the "missing 6 dB"?, J. Acoust. Soc. Am. **63**, 1501–1510 (1978)

13.3 B.C.J. Moore, B.R. Glasberg, T. Baer: A model for the prediction of thresholds, loudness and partial loudness, J. Audio Eng. Soc. **45**, 224–240 (1997)

13.4 M.A. Cheatham, P. Dallos: Inner hair cell response patterns: implications for low-frequency hearing, J. Acoust. Soc. Am. **110**, 2034–2044 (2001)

13.5 L.J. Sivian, S.D. White: On minimum audible sound fields, J. Acoust. Soc. Am. **4**, 288–321 (1933)

13.6 I. Pollack: Monaural and binaural threshold sensitivity for tones and for white noise, J. Acoust. Soc. Am. **20**, 52–57 (1948)

13.7 J.K. Dierks, L.A. Jeffress: Interaural phase and the absolute threshold for tone, J. Acoust. Soc. Am. **34**, 981–986 (1962)

13.8 K. Krumbholz, R.D. Patterson, D. Pressnitzer: The lower limit of pitch as determined by rate discrimination, J. Acoust. Soc. Am. **108**, 1170–1180 (2000)

13.9 R. Plomp, M.A. Bouman: Relation between hearing threshold and duration for tone pulses, J. Acoust. Soc. Am. **31**, 749–758 (1959)

13.10 ANSI: *ANSI S1.1-1994. American National Standard Acoustical Terminology* (American National Standards Institute, New York 1994)

13.11 R.L. Wegel, C.E. Lane: The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear, Phys. Rev. **23**, 266–285 (1924)

13.12 L.L. Vogten: Low-level pure-tone masking: a comparison of 'tuning curves' obtained with simultaneous and forward masking, J. Acoust. Soc. Am. **63**, 1520–1527 (1978)

13.13 H. Fletcher: Auditory patterns, Rev. Mod. Phys. **12**, 47–65 (1940)

13.14 H.L.F. Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig 1863)

13.15 R.D. Patterson, B.C.J. Moore: Auditory filters and excitation patterns as representations of frequency resolution. In: *Frequency Selectivity in Hearing*, ed. by B.C.J. Moore (Academic, London 1986)

13.16 D. Johnson-Davies, R.D. Patterson: Psychophysical tuning curves: restricting the listening band to the signal region, J. Acoust. Soc. Am. **65**, 765–770 (1979)

13.17 B.J. O'Loughlin, B.C.J. Moore: Off-frequency listening: effects on psychoacoustical tuning curves obtained in simultaneous and forward masking, J. Acoust. Soc. Am. **69**, 1119–1125 (1981)

13.18 B.J. O'Loughlin, B.C.J. Moore: Improving psychoacoustical tuning curves, Hear. Res. **5**, 343–346 (1981)

13.19 R.D. Patterson: Auditory filter shapes derived with noise stimuli, J. Acoust. Soc. Am. **59**, 640–654 (1976)

13.20 B.R. Glasberg, B.C.J. Moore: Derivation of auditory filter shapes from notched-noise data, Hear. Res. **47**, 103–138 (1990)

13.21 J.P. Egan, H.W. Hake: On the masking pattern of a simple auditory stimulus, J. Acoust. Soc. Am. **22**, 622–630 (1950)

13.22 E. Zwicker, E. Terhardt: Analytical expressions for critical band rate and critical bandwidth as a function of frequency, J. Acoust. Soc. Am. **68**, 1523–1525 (1980)

13.23 R.D. Patterson, I. Nimmo-Smith: Off-frequency listening and auditory filter asymmetry, J. Acoust. Soc. Am. **67**, 229–245 (1980)

13.24 B.C.J. Moore, B.R. Glasberg: Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns, Hear. Res. **28**, 209–225 (1987)

13.25 S. Rosen, R.J. Baker, A. Darling: Auditory filter nonlinearity at 2 kHz in normal hearing listeners, J. Acoust. Soc. Am. **103**, 2539–2550 (1998)

13.26 B.R. Glasberg, B.C.J. Moore, R.D. Patterson, I. Nimmo-Smith: Dynamic range and asymmetry of the auditory filter, J. Acoust. Soc. Am. **76**, 419–427 (1984)

13.27 E. Zwicker, H. Fastl: *Psychoacoustics – Facts and Models*, 2nd edn. (Springer, Berlin 1999)

13.28 B.C.J. Moore, J.I. Alcántara, T. Dau: Masking patterns for sinusoidal and narrowband noise maskers, J. Acoust. Soc. Am. **104**, 1023–1038 (1998)

13.29 B.C.J. Moore, B.R. Glasberg: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns, J. Acoust. Soc. Am. **74**, 750–753 (1983)

13.30 K. Miyazaki, T. Sasaki: Pure-tone masking patterns in nonsimultaneous masking conditions, Jap. Psychol. Res. **26**, 110–119 (1984)

13.31 A.J. Oxenham, B.C.J. Moore: Modeling the additivity of nonsimultaneous masking, Hear. Res. **80**, 105–118 (1994)

13.32 B.C.J. Moore, B.R. Glasberg: Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: implications for suppression in noise, J. Acoust. Soc. Am. **73**, 1249–1259 (1983)

13.33 G. Kidd, L.L. Feth: Effects of masker duration in pure-tone forward masking, J. Acoust. Soc. Am. **72**, 1384–1386 (1982)

13.34 E. Zwicker: Dependence of post-masking on masker duration and its relation to temporal ef-

Amazon
EX1013 Page 506

fects in loudness, J. Acoust. Soc. Am. **75**, 219–223 (1984)
13.35 H. Fastl: Temporal masking effects: I. Broad band noise masker, Acustica **35**, 287–302 (1976)
13.36 H. Duifhuis: Audibility of high harmonics in a periodic pulse II. Time effects, J. Acoust. Soc. Am. **49**, 1155–1162 (1971)
13.37 H. Duifhuis: Consequences of peripheral frequency selectivity for nonsimultaneous masking, J. Acoust. Soc. Am. **54**, 1471–1488 (1973)
13.38 C.J. Plack, B.C.J. Moore: Temporal window shape as a function of frequency and level, J. Acoust. Soc. Am. **87**, 2178–2187 (1990)
13.39 W. Jesteadt, S.P. Bacon, J.R. Lehman: Forward masking as a function of frequency, masker level, and signal delay, J. Acoust. Soc. Am. **71**, 950–962 (1982)
13.40 R.L. Smith: Short-term adaptation in single auditory-nerve fibers: Some poststimulatory effects, J. Neurophysiol. **49**, 1098–1112 (1977)
13.41 C.W. Turner, E.M. Relkin, J. Doucet: Psychophysical and physiological forward masking studies: Probe duration and rise-time effects, J. Acoust. Soc. Am. **96**, 795–800 (1994)
13.42 A.J. Oxenham: Forward masking: adaptation or integration?, J. Acoust. Soc. Am. **109**, 732–741 (2001)
13.43 M. Brosch, C.E. Schreiner: Time course of forward masking tuning curves in cat primary auditory cortex, J. Neurophysiol. **77**, 923–943 (1997)
13.44 A.J. Oxenham, B.C.J. Moore: Additivity of masking in normally hearing and hearing-impaired subjects, J. Acoust. Soc. Am. **98**, 1921–1934 (1995)
13.45 R. Plomp: The ear as a frequency analyzer, J. Acoust. Soc. Am. **36**, 1628–1636 (1964)
13.46 R. Plomp, A.M. Mimpen: The ear as a frequency analyzer II, J. Acoust. Soc. Am. **43**, 764–767 (1968)
13.47 B.C.J. Moore, K. Ohgushi: Audibility of partials in inharmonic complex tones, J. Acoust. Soc. Am. **93**, 452–461 (1993)
13.48 D.R. Soderquist: Frequency analysis and the critical band, Psychon. Sci. **21**, 117–119 (1970)
13.49 P.A. Fine, B.C.J. Moore: Frequency analysis and musical ability, Music Percept. **11**, 39–53 (1993)
13.50 B. Gabriel, B. Kollmeier, V. Mellert: Influence of individual listener, measurement room and choice of test-tone levels on the shape of equal-loudness level contours, Acustica Acta Acustica **83**, 670–683 (1997)
13.51 D. Laming: *The Measurement of Sensation* (Oxford University Press, Oxford 1997)
13.52 H. Fletcher, W.A. Munson: Loudness, its definition, measurement and calculation, J. Acoust. Soc. Am. **5**, 82–108 (1933)
13.53 ISO 226: *Acoustics – normal equal-loudness contours* (International Organization for Standardization, Geneva 2003)
13.54 S.S. Stevens: On the psychophysical law, Psych. Rev. **64**, 153–181 (1957)
13.55 R.P. Hellman, J.J. Zwislocki: Some factors affecting the estimation of loudness, J. Acoust. Soc. Am. **35**, 687–694 (1961)
13.56 E.M. Relkin, J.R. Doucet: Is loudness simply proportional to the auditory nerve spike count?, J. Acoust. Soc. Am. **191**, 2735–2740 (1997)
13.57 H. Fletcher, W.A. Munson: Relation between loudness and masking, J. Acoust. Soc. Am. **9**, 1–10 (1937)
13.58 E. Zwicker: Über psychologische und methodische Grundlagen der Lautheit, Acustica **8**, 237–258 (1958)
13.59 E. Zwicker, B. Scharf: A model of loudness summation, Psych. Rev. **72**, 3–26 (1965)
13.60 B.R. Glasberg, B.C.J. Moore: A model of loudness applicable to time-varying sounds, J. Audio Eng. Soc. **50**, 331–342 (2002)
13.61 H. Fastl: Loudness evaluation by subjects and by a loudness meter. In: *Sensory Research – Multimodal Perspectives*, ed. by R.T. Verrillo (Erlbaum, Hillsdale, New Jersey 1993)
13.62 ANSI: *ANSI S3.4-2005. Procedure for the Computation of Loudness of Steady Sounds* (American National Standards Institute, New York 2005)
13.63 E. Zwicker, G. Flottorp, S.S. Stevens: Critical bandwidth in loudness summation, J. Acoust. Soc. Am. **29**, 548–557 (1957)
13.64 B. Scharf: Complex sounds and critical bands, Psychol. Bull. **58**, 205–217 (1961)
13.65 B. Scharf: Critical bands. In: *Foundations of Modern Auditory Theory*, ed. by J.V. Tobias (Academic, New York 1970)
13.66 B.C.J. Moore, B.R. Glasberg: The role of frequency selectivity in the perception of loudness, pitch and time. In: *Frequency Selectivity in Hearing*, ed. by B.C.J. Moore (Academic, London 1986)
13.67 G.A. Miller: Sensitivity to changes in the intensity of white noise and its relation to masking and loudness, J. Acoust. Soc. Am. **191**, 609–619 (1947)
13.68 R.R. Riesz: Differential intensity sensitivity of the ear for pure tones, Phys. Rev. **31**, 867–875 (1928)
13.69 N.F. Viemeister, S.P. Bacon: Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones, J. Acoust. Soc. Am. **84**, 172–178 (1988)
13.70 S.P. Bacon, N.F. Viemeister: Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects, Audiology **24**, 117–134 (1985)
13.71 N.F. Viemeister, C.J. Plack: Time analysis. In: *Human Psychophysics*, ed. by W.A. Yost, A.N. Popper, R.R. Fay (Springer, New York 1993)
13.72 R. Plomp: The rate of decay of auditory sensation, J. Acoust. Soc. Am. **36**, 277–282 (1964)
13.73 M.J. Penner: Detection of temporal gaps in noise as a measure of the decay of auditory sensation, J. Acoust. Soc. Am. **61**, 552–557 (1977)

Amazon
EX1013 Page 507

13.74 D. Ronken: Monaural detection of a phase difference between clicks, J. Acoust. Soc. Am. **47**, 1091–1099 (1970)
13.75 D.M. Green: Temporal acuity as a function of frequency, J. Acoust. Soc. Am. **54**, 373–379 (1973)
13.76 B.C.J. Moore, R.W. Peters, B.R. Glasberg: Detection of temporal gaps in sinusoids: Effects of frequency and level, J. Acoust. Soc. Am. **93**, 1563–1570 (1993)
13.77 A. Kohlrausch, R. Fassel, T. Dau: The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers, J. Acoust. Soc. Am. **108**, 723–734 (2000)
13.78 B.C.J. Moore, B.R. Glasberg: Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners, J. Acoust. Soc. Am. **110**, 1067–1073 (2001)
13.79 H. Fleischer: Modulationsschwellen von Schmalbandrauschen, Acustica **51**, 154–161 (1982)
13.80 T. Dau, B. Kollmeier, A. Kohlrausch: Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers, J. Acoust. Soc. Am. **102**, 2892–2905 (1997)
13.81 T. Dau, B. Kollmeier, A. Kohlrausch: Modeling auditory processing of amplitude modulation: II. Spectral and temporal integration, J. Acoust. Soc. Am. **102**, 2906–2919 (1997)
13.82 T. Dau, J.L. Verhey, A. Kohlrausch: Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers, J. Acoust. Soc. Am. **106**, 2752–2760 (1999)
13.83 N.F. Viemeister: Temporal modulation transfer functions based on modulation thresholds, J. Acoust. Soc. Am. **66**, 1364–1380 (1979)
13.84 B.C.J. Moore, B.R. Glasberg, C.J. Plack, A.K. Biswas: The shape of the ear's temporal window, J. Acoust. Soc. Am. **83**, 1102–1116 (1988)
13.85 R.H. Kay: Hearing of modulation in sounds, Physiol. Rev. **62**, 894–975 (1982)
13.86 T. Houtgast: Frequency selectivity in amplitude-modulation detection, J. Acoust. Soc. Am. **85**, 1676–1680 (1989)
13.87 S.P. Bacon, D.W. Grantham: Modulation masking: effects of modulation frequency, depth and phase, J. Acoust. Soc. Am. **85**, 2575–2580 (1989)
13.88 S.D. Ewert, T. Dau: Characterizing frequency selectivity for envelope fluctuations, J. Acoust. Soc. Am. **108**, 1181–1196 (2000)
13.89 C. Lorenzi, C. Soares, T. Vonner: Second-order temporal modulation transfer functions, J. Acoust. Soc. Am. **110**, 1030–1038 (2001)
13.90 A. Sek, B.C.J. Moore: Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators, J. Acoust. Soc. Am. **113**, 2801–2811 (2003)
13.91 C.D. Creelman: Human discrimination of auditory duration, J. Acoust. Soc. Am. **34**, 582–593 (1962)
13.92 S.M. Abel: Duration discrimination of noise and tone bursts, J. Acoust. Soc. Am. **51**, 1219–1223 (1972)
13.93 S.M. Abel: Discrimination of temporal gaps, J. Acoust. Soc. Am. **52**, 519–524 (1972)
13.94 P.L. Divenyi, W.F. Danner: Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra, Percept. Psychophys. **21**, 125–142 (1977)
13.95 J.H. Patterson, D.M. Green: Discrimination of transient signals having identical energy spectra, J. Acoust. Soc. Am. **48**, 894–905 (1970)
13.96 J. Zera, D.M. Green: Detecting temporal onset and offset asynchrony in multicomponent complexes, J. Acoust. Soc. Am. **93**, 1038–1052 (1993)
13.97 J.C. Risset, D.L. Wessel: Exploration of timbre by analysis and synthesis. In: *The Psychology of Music*, 2nd edn., ed. by D. Deutsch (Academic, San Diego 1999)
13.98 G. von Békésy: *Experiments in Hearing* (McGraw-Hill, New York 1960)
13.99 J.F. Schouten: The residue and the mechanism of hearing, Proc. Kon. Ned. Akad. Wetenschap. **43**, 991–999 (1940)
13.100 W.M. Siebert: Frequency discrimination in the auditory system: place or periodicity mechanisms, Proc. IEEE **58**, 723–730 (1970)
13.101 E. Zwicker: Masking and psychological excitation as consequences of the ear's frequency analysis. In: *Frequency Analysis and Periodicity Detection in Hearing*, ed. by R. Plomp, G.F. Smoorenburg (Sijthoff, Leiden 1970)
13.102 C.C. Wier, W. Jesteadt, D.M. Green: Frequency discrimination as a function of frequency and sensation level, J. Acoust. Soc. Am. **61**, 178–184 (1977)
13.103 A. Sek, B.C.J. Moore: Frequency discrimination as a function of frequency, measured in several ways, J. Acoust. Soc. Am. **97**, 2479–2486 (1995)
13.104 B.C.J. Moore: Relation between the critical bandwidth and the frequency-difference limen, J. Acoust. Soc. Am. **55**, 359 (1974)
13.105 J.L. Goldstein, P. Srulovicz: Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. In: *Psychophysics and Physiology of Hearing*, ed. by E.F. Evans, J.P. Wilson (Academic, London 1977)
13.106 B.C.J. Moore, A. Sek: Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking, J. Acoust. Soc. Am. **100**, 2320–2331 (1996)
13.107 G. Revesz: *Zur Grundlegung der Tonpsychologie* (Veit, Leipzig 1913)
13.108 A. Bachem: Tone height and tone chroma as two different pitch qualities, Acta Psych. **7**, 80–88 (1950)

Amazon
EX1013 Page 508

13.109 W.D. Ward: Subjective musical pitch, J. Acoust. Soc. Am. **26**, 369–380 (1954)

13.110 F. Attneave, R.K. Olson: Pitch as a medium: A new approach to psychophysical scaling, Am. J. Psychol. **84**, 147–166 (1971)

13.111 K. Ohgushi, T. Hatoh: Perception of the musical pitch of high frequency tones. In: *Ninth International Symposium on Hearing: Auditory Physiology and Perception*, ed. by Y. Cazals, L. Demany, K. Horner (Pergamon, Oxford 1991)

13.112 S.S. Stevens: The relation of pitch to intensity, J. Acoust. Soc. Am. **6**, 150–154 (1935)

13.113 J. Verschuure, A.A. van Meeteren: The effect of intensity on pitch, Acustica **32**, 33–44 (1975)

13.114 G.S. Ohm: Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen, Annalen der Physik und Chemie **59**, 513–565 (1843)

13.115 J.F. Schouten: The residue revisited. In: *Frequency Analysis and Periodicity Detection in Hearing*, ed. by R. Plomp, G.F. Smoorenburg (Sijthoff, Leiden, The Netherlands 1970)

13.116 J.C.R. Licklider: Auditory frequency analysis. In: *Information Theorie*, ed. by C. Cherry (Academic, New York 1956)

13.117 E. de Boer: *On the 'residue' in hearing, Ph.D. Thesis* (University of Amsterdam, Amsterdam 1956)

13.118 J.L. Goldstein: An optimum processor theory for the central formation of the pitch of complex tones, J. Acoust. Soc. Am. **54**, 1496–1516 (1973)

13.119 E. Terhardt: Pitch, consonance, and harmony, J. Acoust. Soc. Am. **55**, 1061–1069 (1974)

13.120 G.F. Smoorenburg: Pitch perception of two-frequency stimuli, J. Acoust. Soc. Am. **48**, 924–941 (1970)

13.121 A.J.M. Houtsma, J.F.M. Fleuren: Analytic and synthetic pitch of two-tone complexes, J. Acoust. Soc. Am. **90**, 1674–1676 (1991)

13.122 J.L. Flanagan, M.G. Saslow: Pitch discrimination for synthetic vowels, J. Acoust. Soc. Am. **30**, 435–442 (1958)

13.123 B.C.J. Moore, B.R. Glasberg, M.J. Shailer: Frequency and intensity difference limens for harmonics within complex tones, J. Acoust. Soc. Am. **75**, 550–561 (1984)

13.124 B.C.J. Moore, B.R. Glasberg, R.W. Peters: Relative dominance of individual partials in determining the pitch of complex tones, J. Acoust. Soc. Am. **77**, 1853–1860 (1985)

13.125 R. Plomp: Pitch of complex tones, J. Acoust. Soc. Am. **41**, 1526–1533 (1967)

13.126 R.J. Ritsma: Frequencies dominant in the perception of the pitch of complex sounds, J. Acoust. Soc. Am. **42**, 191–198 (1967)

13.127 R.P. Carlyon, T.M. Shackleton: Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?, J. Acoust. Soc. Am. **95**, 3541–3554 (1994)

13.128 A. Hoekstra, R.J. Ritsma: Perceptive hearing loss and frequency selectivity. In: *Psychophysics and Physiology of Hearing*, ed. by E.F. Evans, J.P. Wilson (Academic, London, England 1977)

13.129 A.J.M. Houtsma, J. Smurzynski: Pitch identification and discrimination for complex tones with many harmonics, J. Acoust. Soc. Am. **87**, 304–310 (1990)

13.130 B.C.J. Moore, S.M. Rosen: Tune recognition with reduced pitch and interval information, Q. J. Exp. Psychol. **31**, 229–240 (1979)

13.131 R. Meddis, M. Hewitt: A computational model of low pitch judgement. In: *Basic Issues in Hearing*, ed. by H. Duifhuis, J.W. Horst, H.P. Wit (Academic, London 1988)

13.132 R. Meddis, M. Hewitt: Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification, J. Acoust. Soc. Am. **89**, 2866–2882 (1991)

13.133 B.C.J. Moore: *An Introduction to the Psychology of Hearing*, 2nd edn. (Academic, London 1982)

13.134 B.C.J. Moore: *An Introduction to the Psychology of Hearing*, 5th edn. (Academic, San Diego 2003)

13.135 P. Srulovicz, J.L. Goldstein: A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum, J. Acoust. Soc. Am. **73**, 1266–1276 (1983)

13.136 R. Plomp: Timbre as a multidimensional attribute of complex tones. In: *Frequency Analysis and Periodicity Detection in Hearing*, ed. by R. Plomp, G.F. Smoorenburg (Sijthoff, Leiden 1970)

13.137 G. von Bismarck: Sharpness as an attribute of the timbre of steady sounds, Acustica **30**, 159–172 (1974)

13.138 R. Plomp: *Aspects of Tone Sensation* (Academic, London 1976)

13.139 R.D. Patterson: A pulse ribbon model of monaural phase perception, J. Acoust. Soc. Am. **82**, 1560–1586 (1987)

13.140 R. Plomp, H.J.M. Steeneken: Effect of phase on the timbre of complex tones, J. Acoust. Soc. Am. **46**, 409–421 (1969)

13.141 A.J. Watkins: Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion, J. Acoust. Soc. Am. **90**, 2942–2955 (1991)

13.142 J.F. Schouten: The perception of timbre, 6th International Conference on Acoustics 1, GP-6-2 (1968)

13.143 R.D. Patterson: The sound of a sinusoid: Spectral models, J. Acoust. Soc. Am. **96**, 1409–1418 (1994)

13.144 R.D. Patterson: The sound of a sinusoid: Time-interval models, J. Acoust. Soc. Am. **96**, 1419–1428 (1994)

13.145 M.A. Akeroyd, R.D. Patterson: Discrimination of wideband noises modulated by a temporally

Amazon
EX1013 Page 509

asymmetric function, J. Acoust. Soc. Am. **98**, 2466–2474 (1995)
13.146 H.F. Pollard, E.V. Jansson: A tristimulus method for the specification of musical timbre, Acustica **51**, 162–171 (1982)
13.147 S. Handel: Timbre perception and auditory object identification. In: *Hearing*, ed. by B.C.J. Moore (Academic, San Diego 1995)
13.148 A.W. Mills: On the minimum audible angle, J. Acoust. Soc. Am. **30**, 237–246 (1958)
13.149 L. Rayleigh: On our perception of sound direction, Phil. Mag. **13**, 214–232 (1907)
13.150 E.R. Hafter: Spatial hearing and the duplex theory: How viable?. In: *Dynamic Aspects of Neocortical Function*, ed. by G.M. Edelman, W.E. Gall, W.M. Cowan (Wiley, New York 1984)
13.151 G.B. Henning: Detectability of interaural delay in high-frequency complex waveforms, J. Acoust. Soc. Am. **55**, 84–90 (1974)
13.152 D.W. Batteau: The role of the pinna in human localization, Proc. Roy. Soc. B. **168**, 158–180 (1967)
13.153 J. Blauert: *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, Mass 1997)
13.154 W.M. Hartmann, A. Wittenberg: On the externalization of sound images, J. Acoust. Soc. Am. **99**, 3678–3688 (1996)
13.155 H. Haas: Über den Einfluss eines Einfachechos an die Hörsamkeit von Sprache, Acustica **1**, 49–58 (1951)
13.156 H. Wallach, E.B. Newman, M.R. Rosenzweig: The precedence effect in sound localization, Am. J. Psychol. **62**, 315–336 (1949)
13.157 R.Y. Litovsky, H.S. Colburn, W.A. Yost, S.J. Guzman: The precedence effect, J. Acoust. Soc. Am. **106**, 1633–1654 (1999)
13.158 A.S. Bregman: *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge, Mass. 1990)
13.159 A.S. Bregman, S. Pinker: Auditory streaming and the building of timbre, Canad. J. Psychol. **32**, 19–31 (1978)
13.160 C.J. Darwin, R.P. Carlyon: Auditory grouping. In: *Hearing*, ed. by B.C.J. Moore (Academic, San Diego 1995)
13.161 D.E. Broadbent, P. Ladefoged: On the fusion of sounds reaching different sense organs, J. Acoust. Soc. Am. **29**, 708–710 (1957)
13.162 M.T.M. Scheffers: *Sifting vowels: auditory pitch analysis and sound segregation, Ph.D. Thesis* (Groningen University, The Netherlands 1983)
13.163 P.F. Assmann, A.Q. Summerfield: Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies, J. Acoust. Soc. Am. **88**, 680–697 (1990)
13.164 J.D. McKeown, R.D. Patterson: The time course of auditory segregation: Concurrent vowels that vary in duration, J. Acoust. Soc. Am. **98**, 1866–1877 (1995)
13.165 B.C.J. Moore, B.R. Glasberg, R.W. Peters: Thresholds for hearing mistuned partials as separate tones in harmonic complexes, J. Acoust. Soc. Am. **80**, 479–483 (1986)
13.166 B. Roberts, J.M. Brunstrom: Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes, J. Acoust. Soc. Am. **104**, 2326–2338 (1998)
13.167 B. Roberts, J.M. Brunstrom: Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch, J. Acoust. Soc. Am. **110**, 2479–2490 (2001)
13.168 R. Meddis, M. Hewitt: Modeling the identification of concurrent vowels with different fundamental frequencies, J. Acoust. Soc. Am. **91**, 233–245 (1992)
13.169 A. de Cheveigné, S. McAdams, C.M.H. Marin: Concurrent vowel identification. II. Effects of phase, harmonicity and task, J. Acoust. Soc. Am. **101**, 2848–2856 (1997)
13.170 A. de Cheveigné, H. Kawahara, M. Tsuzaki, K. Aikawa: Concurrent vowel identification. I. Effects of relative amplitude and F0 difference, J. Acoust. Soc. Am. **101**, 2839–2847 (1997)
13.171 A. de Cheveigné: Concurrent vowel identification. III. A neural model of harmonic interference cancellation, J. Acoust. Soc. Am. **101**, 2857–2865 (1997)
13.172 R.A. Rasch: The perception of simultaneous notes such as in polyphonic music, Acustica **40**, 21–33 (1978)
13.173 C.J. Darwin, N.S. Sutherland: Grouping frequency components of vowels: when is a harmonic not a harmonic?, Q. J. Exp. Psychol. **36A**, 193–208 (1984)
13.174 B. Roberts, B.C.J. Moore: The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony, J. Acoust. Soc. Am. **89**, 2922–2932 (1991)
13.175 E. Zwicker: 'Negative afterimage' in hearing, J. Acoust. Soc. Am. **36**, 2413–2415 (1964)
13.176 A.Q. Summerfield, A.S. Sidwell, T. Nelson: Auditory enhancement of changes in spectral amplitude, J. Acoust. Soc. Am. **81**, 700–708 (1987)
13.177 A.S. Bregman, J. Abramson, P. Doehring, C.J. Darwin: Spectral integration based on common amplitude modulation, Percept. Psychophys. **37**, 483–493 (1985)
13.178 J.W. Hall, J.H. Grose: Comodulation masking release and auditory grouping, J. Acoust. Soc. Am. **88**, 119–125 (1990)
13.179 B.C.J. Moore, M.J. Shailer: Comodulation masking release as a function of level, J. Acoust. Soc. Am. **90**, 829–835 (1991)

Amazon
EX1013 Page 510

13.180 B.C.J. Moore, M.J. Shailer, M.J. Black: Dichotic interference effects in gap detection, J. Acoust. Soc. Am. **93**, 2130–2133 (1993)
13.181 Q. Summerfield, J.F. Culling: Auditory segregation of competing voices: absence of effects of FM or AM coherence, Phil. Trans. R. Soc. Lond. B **336**, 357–366 (1992)
13.182 M.F. Cohen, X. Chen: Dynamic frequency change among stimulus components: Effects of coherence on detectability, J. Acoust. Soc. Am. **92**, 766–772 (1992)
13.183 M.H. Chalikia, A.S. Bregman: The perceptual segregation of simultaneous vowels with harmonic, shifted, and random components, Percept. Psychophys. **53**, 125–133 (1993)
13.184 S. McAdams: Segregation of concurrent sounds. I.: Effects of frequency modulation coherence, J. Acoust. Soc. Am. **86**, 2148–2159 (1989)
13.185 R.P. Carlyon: Discriminating between coherent and incoherent frequency modulation of complex tones, J. Acoust. Soc. Am. **89**, 329–340 (1991)
13.186 R.P. Carlyon: Further evidence against an across-frequency mechanism specific to the detection of frequency modulation (FM) incoherence between resolved frequency components, J. Acoust. Soc. Am. **95**, 949–961 (1994)
13.187 J. Lyzenga, B.C.J. Moore: Effect of FM coherence for inharmonic stimuli: FM-phase discrimination and identification of artificial double vowels, J. Acoust. Soc. Am. **117**, 1314–1325 (2005)
13.188 C.M.H. Marin, S. McAdams: Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width, J. Acoust. Soc. Am. **89**, 341–351 (1991)
13.189 S. Furukawa, B.C.J. Moore: Across-frequency processes in frequency modulation detection, J. Acoust. Soc. Am. **100**, 2299–2312 (1996)
13.190 S. Furukawa, B.C.J. Moore: Dependence of frequency modulation detection on frequency modulation coherence across carriers: Effects of modulation rate, harmonicity and roving of the carrier frequencies, J. Acoust. Soc. Am. **101**, 1632–1643 (1997)
13.191 S. Furukawa, B.C.J. Moore: Effect of the relative phase of amplitude modulation on the detection of modulation on two carriers, J. Acoust. Soc. Am. **102**, 3657–3664 (1997)
13.192 R.P. Carlyon: Detecting coherent and incoherent frequency modulation, Hear. Res. **140**, 173–188 (2000)
13.193 M. Kubovy, J.E. Cutting, R.M. McGuire: Hearing with the third ear: dichotic perception of a melody without monaural familiarity cues, Science **186**, 272–274 (1974)
13.194 J.F. Culling: Auditory motion segregation: a limited analogy with vision, J. Exp. Psychol.: Human Percept. Perf. **26**, 1760–1769 (2000)
13.195 M.A. Akeroyd, B.C.J. Moore, G.A. Moore: Melody recognition using three types of dichotic-pitch stimulus, J. Acoust. Soc. Am. **110**, 1498–1504 (2001)
13.196 T.M. Shackleton, R. Meddis: The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs, J. Acoust. Soc. Am. **91**, 3579–3581 (1992)
13.197 J.F. Culling, Q. Summerfield: Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay, J. Acoust. Soc. Am. **98**, 785–797 (1995)
13.198 C.J. Darwin, R.W. Hukin: Auditory objects of attention: the role of interaural time differences, J. Exp. Psychol.: Human Percept. Perf. **25**, 617–629 (1999)
13.199 G.A. Miller, G.A. Heise: The trill threshold, J. Acoust. Soc. Am. **22**, 637–638 (1950)
13.200 A.S. Bregman, J. Campbell: Primary auditory stream segregation and perception of order in rapid sequences of tones, J. Exp. Psychol. **89**, 244–249 (1971)
13.201 L.P.A.S. van Noorden: *Temporal coherence in the perception of tone sequences*, *Ph.D. Thesis* (Eindhoven University of Technology, Eindhoven 1975)
13.202 L.P.A.S. van Noorden: Rhythmic fission as a function of tone rate, IPO Annual Prog. Rep. **6**, 9–12 (1971)
13.203 A.S. Bregman, G. Dannenbring: The effect of continuity on auditory stream segregation, Percept. Psychophys. **13**, 308–312 (1973)
13.204 A.S. Bregman: Auditory streaming is cumulative, J. Exp. Psychol.: Human Percept. Perf. **4**, 380–387 (1978)
13.205 W.L. Rogers, A.S. Bregman: An experimental evaluation of three theories of auditory stream segregation, Percept. Psychophys. **53**, 179–189 (1993)
13.206 W.L. Rogers, A.S. Bregman: Cumulation of the tendency to segregate auditory streams: resetting by changes in location and loudness, Percept. Psychophys. **60**, 1216–1227 (1998)
13.207 B.C.J. Moore, H. Gockel: Factors influencing sequential stream segregation, Acta Acustica - Acustica **88**, 320–333 (2002)
13.208 W.M. Hartmann, D. Johnson: Stream segregation and peripheral channeling, Music Percept. **9**, 155–184 (1991)
13.209 P.G. Singh, A.S. Bregman: The influence of different timbre attributes on the perceptual segregation of complex-tone sequences, J. Acoust. Soc. Am. **102**, 1943–1952 (1997)
13.210 J. Vliegen, B.C.J. Moore, A.J. Oxenham: The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task, J. Acoust. Soc. Am. **106**, 938–945 (1999)

Amazon
EX1013 Page 511

13.211 J. Vliegen, A.J. Oxenham: Sequential stream segregation in the absence of spectral cues, J. Acoust. Soc. Am. **105**, 339–346 (1999)
13.212 P. Iverson: Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes, J. Exp. Psychol.: Human Percept. Perf. **21**, 751–763 (1995)
13.213 B. Roberts, B.R. Glasberg, B.C.J. Moore: Primitive stream segregation of tone sequences without differences in F0 or passband, J. Acoust. Soc. Am. **112**, 2074–2085 (2002)
13.214 H. Gockel, R.P. Carlyon, C. Micheyl: Context dependence of fundamental-frequency discrimination: Lateralized temporal fringes, J. Acoust. Soc. Am. **106**, 3553–3563 (1999)
13.215 W.J. Dowling: Rhythmic fission and perceptual organization, J. Acoust. Soc. Am. **44**, 369 (1968)
13.216 W.J. Dowling: The perception of interleaved melodies, Cognitive Psychol. **5**, 322–337 (1973)
13.217 D. Deutsch: Two-channel listening to musical scales, J. Acoust. Soc. Am. **57**, 1156–1160 (1975)
13.218 D.E. Broadbent, P. Ladefoged: Auditory perception of temporal order, J. Acoust. Soc. Am. **31**, 151–159 (1959)
13.219 R.M. Warren, C.J. Obusek, R.M. Farmer, R.P. Warren: Auditory sequence: confusion of patterns other than speech or music, Science N.Y. **164**, 586–587 (1969)
13.220 R.M. Warren: Auditory temporal discrimination by trained listeners, Cognitive Psychol. **6**, 237–256 (1974)
13.221 P.L. Divenyi, I.J. Hirsh: Identification of temporal order in three-tone sequences, J. Acoust. Soc. Am. **56**, 144–151 (1974)
13.222 R. Cusack, B. Roberts: Effects of differences in timbre on sequential grouping, Percept. Psychophys. **62**, 1112–1120 (2000)
13.223 K. Koffka: *Principles of Gestalt Psychology* (Harcourt and Brace, New York 1935)
13.224 C.J. Darwin, C.E. Bethell-Fox: Pitch continuity and speech source attribution, J. Exp. Psychol.: Hum. Perc. Perf. **3**, 665–672 (1977)
13.225 A.S. Bregman, A. Rudnicky: Auditory segregation: stream or streams?, J. Exp. Psychol.: Human Percept. Perf. **1**, 263–267 (1975)
13.226 A.S. Bregman: The meaning of duplex perception: sounds as transparent objects. In: *The Psychophysics of Speech Perception*, ed. by M.E.H. Schouten (Martinus Nijhoff, Dordrecht 1987)
13.227 T. Houtgast: Psychophysical evidence for lateral inhibition in hearing, J. Acoust. Soc. Am. **51**, 1885–1894 (1972)
13.228 W.R. Thurlow: An auditory figure-ground effect, Am. J. Psychol. **70**, 653–654 (1957)
13.229 R.M. Warren, C.J. Obusek, J.M. Ackroff: Auditory induction: perceptual synthesis of absent sounds, Science **176**, 1149–1151 (1972)
13.230 V. Ciocca, A.S. Bregman: Perceived continuity of gliding and steady-state tones through interrupting noise, Percept. Psychophys. **42**, 476–484 (1987)
13.231 G.A. Miller, J.C.R. Licklider: The intelligibility of interrupted speech, J. Acoust. Soc. Am. **22**, 167–173 (1950)
13.232 D. Dirks, D. Bower: Effects of forward and backward masking on speech intelligibility, J. Acoust. Soc. Am. **47**, 1003–1008 (1970)
13.233 R.M. Warren: Perceptual restoration of missing speech sounds, Science **167**, 392–393 (1970)

Amazon
EX1013 Page 512

14. Acoustic Signal Processing

Signal processing refers to the acquisition, storage, display, and generation of signals – also to the extraction of information from signals and the re-encoding of information. As such, signal processing in some form is an essential element in the practice of all aspects of acoustics. Signal processing algorithms enable acousticians to separate signals from noise, to perform automatic speech recognition, or to compress information for more efficient storage or transmission. Signal processing concepts are the blocks used to build models of speech and hearing. As we enter the 21st century, all signal processing is effectively digital signal processing. Widespread access to high-speed processing, massive memory, and inexpensive software makes signal processing procedures of enormous sophistication and power available to anyone who wants to use them. Because advanced signal processing is now accessible to everybody, there is a need for primers that introduce basic mathematical concepts that underlie the digital algorithms. The present handbook chapter is intended to serve such a purpose.

The chapter emphasizes careful definition of essential terms used in the description of signals per international standards. It introduces the Fourier series for signals that are periodic and the Fourier transform for signals that are not. Both begin with analog, continuous signals, appropriate for the real acoustical world. Emphasis is placed on the consequences of signal symmetry and on formal relationships. The autocorrelation function is related to the energy and power spectra for finite-duration and infinite-duration signals. The chapter provides careful definitions of statistical terms, moments, and single- and multi-variate distributions. The Hilbert transform is introduced, again in terms of continuous functions. It is applied both to the development of the analytic signal – envelope and phase – and to the dispersion relations for linear, time-invariant systems. The bare essentials of filtering are presented, mostly to provide real-world examples of fundamental concepts – asymptotic responses, group delay, phase delay, etc. There is a brief introduction to cepstrology, with emphasis on acoustical applications. The treatment of the mathematical properties of noise emphasizes the generation of different kinds of noise. Digital signal processing with sampled data is specifically introduced with emphasis on digital-to-analog conversion and analog-to-digital conversion. It continues with the discrete Fourier transform and with the *z*-transform, applied to both signals and linear, time-invariant systems. Digital signal processing continues with an introduction to maximum length sequences as used in acoustical measurements, with an emphasis on formal properties. The chapter ends with a section on information theory including developments of Shannon entropy and mutual information.

14.1 **Definitions** 504

14.2 **Fourier Series** 505
14.2.1 The Spectrum 506
14.2.2 Symmetry 506

14.3 **Fourier Transform** 507
14.3.1 Examples 508
14.3.2 Time-Shifted Function 509
14.3.3 Derivatives and Integrals 509
14.3.4 Products and Convolution 509

14.4 **Power, Energy, and Power Spectrum** 510
14.4.1 Autocorrelation 510
14.4.2 Cross-Correlation 511

14.5 **Statistics** 511
14.5.1 Signals and Processes 512
14.5.2 Distributions 512
14.5.3 Multivariate Distributions 513
14.5.4 Moments 513

14.6 **Hilbert Transform and the Envelope** 514
14.6.1 The Analytic Signal 514

14.7 **Filters** 515
14.7.1 One-Pole Low-Pass Filter 515
14.7.2 Phase Delay and Group Delay 516
14.7.3 Resonant Filters 516
14.7.4 Impulse Response 516
14.7.5 Dispersion Relations 516

Amazon
EX1013 Page 513

14.8 **The Cepstrum** 517
14.9 **Noise** 518
14.9.1 Thermal Noise 518
14.9.2 Gaussian Noise 519
14.9.3 Band-Limited Noise 519
14.9.4 Generating Noise 519
14.9.5 Equal-Amplitude Random-Phase Noise 520
14.9.6 Noise Color 520
14.10 **Sampled data** 520
14.10.1 Quantization and Quantization Noise 520
14.10.2 Binary Representation 520
14.10.3 Sampling Operation 521
14.10.4 Digital-to-Analog Conversion 521
14.10.5 The Sampled Signal 522
14.10.6 Interpolation 522
14.11 **Discrete Fourier Transform** 522
14.11.1 Interpolation for the Spectrum 523
14.12 **The *z*-Transform** 524
14.12.1 Transfer Function 525
14.13 **Maximum Length Sequences** 526
14.13.1 The MLS as a Signal 527
14.13.2 Application of the MLS 527
14.13.3 Long Sequences 527
14.14 **Information Theory** 528
14.14.1 Shannon Entropy 529
14.14.2 Mutual Information 530
References 530

14.1 Definitions

Signal processing begins with signals. The simplest signal is a sine wave with a single spectral component, i. e., with a single frequency, as shown in Fig. 14.1. It is sometimes called a pure tone. A sine wave function of time t with *amplitude* C, *angular frequency* ω, and *starting phase* φ, is given by

$$x(t) = C \sin(\omega t + \varphi) \,. \tag{14.1}$$

The amplitude has the same units as the waveform x, the angular frequency has units of radians per second, and the phase has units of radians.

Because there are 2π radians in one cycle

$$\omega = 2\pi f, \tag{14.2}$$

and (14.1) can be written as

$$x(t) = C \sin(2\pi f t + \varphi) \tag{14.3}$$

or as

$$x(t) = C \sin(2\pi t/T + \varphi) \,, \tag{14.4}$$

where f is the *frequency* in cycles per second (or Hertz) and T is the *period* in units of seconds per cycle, $T = 1/f$.

A *complex wave* is the sum of two or more sine waves, each with its own amplitude, frequency, and phase. For example,

$$x(t) = C_1 \sin(\omega_1 t + \varphi_1) + C_2 \sin(\omega_2 t + \varphi_2) \tag{14.5}$$

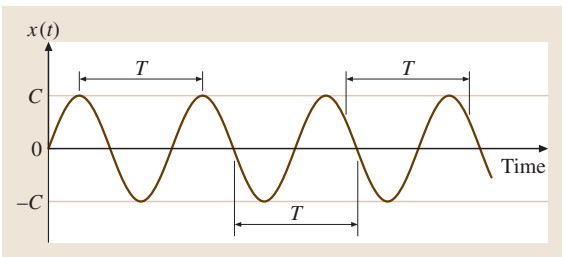


Fig. 14.1 A sine wave with amplitude C and period T. A little more than three and a half cycles are shown. The starting phase is $\varphi = 0$

is a complex wave with two spectral components having frequencies f_1 and f_2. The period of a complex wave is the reciprocal of the greatest common divisor of f_1 and f_2. For instance, if $f_1 = 400\,\text{Hz}$ and $f_2 = 600\,\text{Hz}$, then the period is $1/(200\,\text{Hz})$ or 5 ms. The *fundamental frequency* is the reciprocal of the period.

A general waveform can be written as a sum of N components,

$$x(t) = \sum_{n=1}^{N} C_n \sin(\omega_n t + \varphi_n) \,, \tag{14.6}$$

and the fundamental frequency is the greatest common divisor of the set of frequencies $\{f_n\}$.

An alternative description of the general waveform can be derived by using the trigonometric identity

$$\sin(\theta_1 + \theta_2) = \sin\theta_1 \cos\theta_2 + \sin\theta_2 \cos\theta_1 \tag{14.7}$$

Amazon
EX1013 Page 514

so that

$$x(t) = \sum_{n=1}^{N} A_n \cos(\omega_n t) + B_n \sin(\omega_n t) \,, \tag{14.8}$$

where $A_n = C_n \sin\varphi_n$, and $B_n = C_n \cos\varphi_n$, are the cosine and sine partial amplitudes respectively. Thus the two parameters C_n and φ_n are replaced by two other parameters A_n and B_n.

Because of the trigonometric identity

$$\sin^2\theta + \cos^2\theta = 1 \,, \tag{14.9}$$

the amplitude C_n can be written in terms of the partial amplitudes,

$$C_n^2 = A_n^2 + B_n^2 \,, \tag{14.10}$$

as can the component phase

$$\varphi_n = \mathrm{Arg}(A_n, B_n) \,. \tag{14.11}$$

The Arg function is essentially an inverse tangent, but because the principal value of the arctangent function only runs from $-\pi/2$ to $\pi/2$, an adjustment needs to be made when B_n is negative. In the end,

$$\mathrm{Arg}(A_n, B_n) = \arctan(A_n/B_n) \quad (\text{for } B_n \geq 0) \tag{14.12}$$

and

$$\mathrm{Arg}(A_n, B_n) = \arctan(A_n/B_n) + \pi \quad (\text{for } B_n < 0) \,.$$

The remaining sections of this chapter provide a brief treatment of real signals $x(t)$ – first as continuous functions of time and then as sampled data. Readers who are less familiar with the continuous approach may wish to refer to the more extensive treatment in [14.1].

14.2 Fourier Series

The Fourier series applies to a function $x(t)$ that is periodic. Periodicity means that we can add any integral multiple m of T to the running time variable t and the function will have the same value as at time t, i. e.

$$x(t + mT) = x(t) \,, \quad \text{for all integral } m \,. \tag{14.13}$$

Because m can be either positive or negative and as large as we like, it is clear that x is periodic into the infinite future and past. Then Fourier's theorem says that x can be represented as a Fourier series like

$$x(t) = A_0 + \sum_{n=1}^{\infty} [A_n \cos(\omega_n t) + B_n \sin(\omega_n t)] \,. \tag{14.14}$$

All the cosines and sines have angular frequencies ω_n that are harmonics, i. e., they are integral multiples of a fundamental angular frequency ω_o,

$$\omega_n = n\omega_o = 2\pi n/T \,, \tag{14.15}$$

where n is an integer.

The fundamental frequency f_0 is given by $f_0 = \omega_o/(2\pi)$. The fundamental frequency is the lowest frequency that a sine or cosine wave can have and still fit exactly into one period of the function $x(t)$ because $f_0 = 1/T$. In order to make a function $x(t)$ with period T, the only sines and cosines that are allowed to enter the sum are those that fit exactly into the same period T. These are those sines and cosines with frequencies that are integral multiples of the fundamental.

The factors A_n and B_n in (14.14) are the Fourier coefficients. They can be calculated by projecting the function $x(t)$ onto sine and cosine functions of the harmonic frequencies ω_n. Projecting means to integrate the product of $x(t)$ and a sine or cosine function over a duration of time equal to a period of $x(t)$. Sines and cosines with different harmonic frequencies are orthogonal over a period. Consequently, projecting $x(t)$ onto, for example $\cos(3\omega_o t)$, gives exactly the Fourier coefficient A_3.

It does not matter which time interval is used for integration, as long as it is exactly one period in duration. It is common to use the interval $-T/2$ to $T/2$.

The orthogonality and normality of the sine and cosine functions are described by the following equations:

$$\frac{2}{T} \int_{-T/2}^{T/2} \mathrm{d}t \, \sin(\omega_n t) \, \cos(\omega_m t) = 0 \,, \tag{14.16}$$

for all m and n;

$$\frac{2}{T} \int_{-T/2}^{T/2} \mathrm{d}t \, \cos(\omega_n t) \, \cos(\omega_m t) = \delta_{n,m} \tag{14.17}$$

and

$$\frac{2}{T} \int_{-T/2}^{T/2} \mathrm{d}t \, \sin(\omega_n t) \, \sin(\omega_m t) = \delta_{n,m} \,, \tag{14.18}$$

Amazon
EX1013 Page 515

where $\delta_{n,m}$ is the Kronecker delta, equal to one if $m = n$ and equal to zero otherwise.

It follows that the equations for A_n and B_n are

$$A_n = \frac{2}{T} \int_{-T/2}^{T/2} \mathrm{d}t\, x(t)\, \cos(\omega_n t) \quad \text{for } n > 0\,, \tag{14.19}$$

$$B_n = \frac{2}{T} \int_{-T/2}^{T/2} \mathrm{d}t\, x(t)\, \sin(\omega_n t) \quad \text{for } n > 0. \tag{14.20}$$

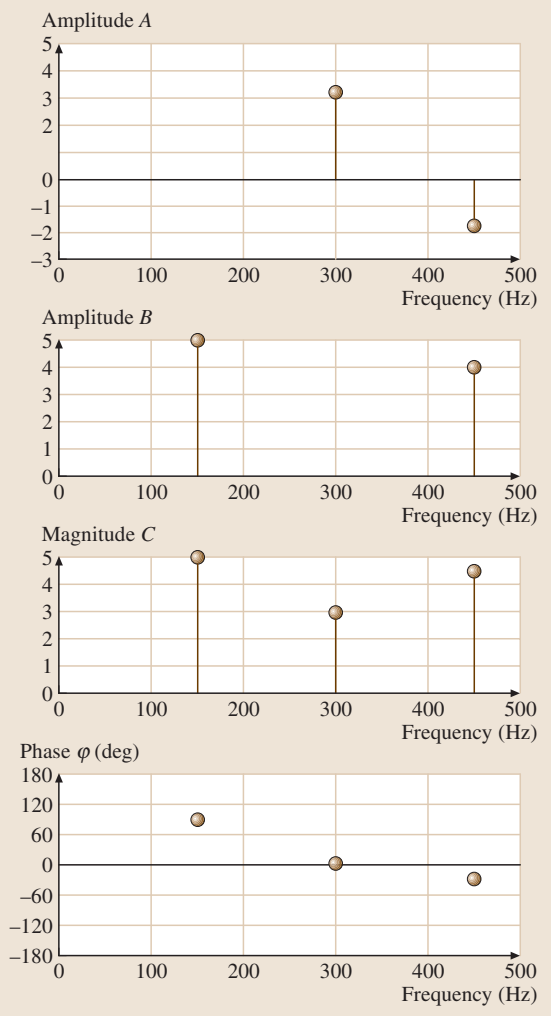


Fig. 14.2 The amplitudes A and B for the signal in (14.22) are shown in the *top* two plots. The corresponding magnitude and phases are shown in the *bottom* two

The coefficient A_0 is simply a constant that shifts the function $x(t)$ up or down. The constant A_0 is the only term in the Fourier series (14.14) that could possibly have a nonzero value when averaged over a period. All the other terms are sines and cosines; they are negative as much as they are positive and average to zero. Therefore, A_0 is the average value of $x(t)$. It is the direct-current (DC) component of x. To find A_0 we project the function $x(t)$ onto a cosine of zero frequency, i. e. onto the number 1, which leads to the average value of x,

$$A_0 = \frac{1}{T} \int_{-T/2}^{T/2} \mathrm{d}t\, x(t)\,. \tag{14.21}$$

14.2.1 The Spectrum

The Fourier series is a function of time, where A_n and B_n are coefficients that weight the cosine and sine contributions to the series. The coefficients A_n and B_n are real numbers that may be positive or negative.

An alternative approach to the function $x(t)$ deemphasizes the time dependence and considers mainly the coefficients themselves. This is the spectral approach. The spectrum simply consists of the values of A_n and B_n, plotted against frequency, or equivalently, plotted against the harmonic number n. For example, if we have a signal given by

$$\begin{aligned} x(t) = {} & 5\, \sin(2\pi\, 150t) + 3\, \cos(2\pi\, 300t) \\ & - 2\cos(2\pi\, 450t) + 4\, \sin(2\pi\, 450t) \end{aligned} \tag{14.22}$$

then the spectrum consists of only a few terms. The period of the signal is $1/150$ s, the fundamental frequency is 150 Hz, and there are two additional harmonics: a second harmonic at 300 Hz and a third at 450 Hz. The spectrum is shown in Fig. 14.2.

14.2.2 Symmetry

Many important periodic functions have symmetries that simplify the Fourier series. If the function $x(t)$ is an even function $[x(-t) = x(t)]$ then the Fourier series for x contains only cosine terms. All coefficients of the sine terms B_n are zero. If $x(t)$ is odd $[x(-t) = -x(t)]$, the the Fourier series contains only sine terms, and all the coefficients A_n are zero. Sometimes it is possible to shift the origin of time to obtain a symmetrical function. Such a time shift is allowed if the physical situation at hand does not require that $x(t)$ be synchronized with some other function of time or with some other time-referenced process. For example, the sawtooth function

Amazon
EX1013 Page 516

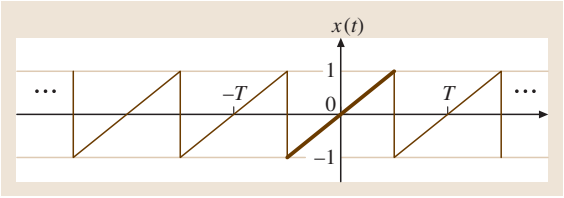


Fig. 14.3 The Fourier series of an odd function like this sawtooth consists of sine terms only. The Fourier coefficients can be computed by an integral over a single period from $-T/2$ to $T/2$

in Fig. 14.3 is an odd function. Therefore, only sine terms are present in the series.

The Fourier coefficients can be calculated by doing the integral over the interval shown by the heavy line. The integral is easy to do analytically because $x(t)$ is just a straight line. The answer is

$$B_n = \frac{2}{\pi}\frac{(-1)^{(n+1)}}{n} \,. \tag{14.23}$$

Consequently, the sawtooth function itself is given by

$$x(t) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{(n+1)}}{n}\,\sin(2\pi nt/T) \,. \tag{14.24}$$

A bridge between the Fourier series and the Fourier transform is the complex form for the spectrum,

$$X_n = A_n + \mathrm{i}B_n \,. \tag{14.25}$$

Because of Euler's formula, namely

$$\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta \,, \tag{14.26}$$

it follows that

$$X_n = \frac{2}{T}\int_{-T/2}^{T/2} \mathrm{d}t\; x(t)\;\mathrm{e}^{\mathrm{i}\omega_n t} \,. \tag{14.27}$$

14.3 Fourier Transform

The Fourier transform of a time-dependent signal is a frequency-dependent representation of the signal, whether or not the time dependence is periodic. Compared to the frequency representation in the Fourier series, the Fourier transform differs in several ways. In general the Fourier transform is a complex function with real and imaginary parts. Whereas the Fourier series representation consists of discrete frequencies, the Fourier transform is a continuous function of frequency. The Fourier transform also requires the concept of negative frequencies. The transformation tends to be symmetrical with respect to the appearance of positive and negative frequencies and so negative frequencies are just as important as positive frequencies. The treatment of the Fourier integral transform that follows mainly states results. For proof and further applications the reader may wish to consult [14.1, mostly Chap. 8].

The Fourier transform of signal $x(t)$ is given by the integral

$$X(\omega) = \mathcal{F}[x(t)] = \int \mathrm{d}t\;\mathrm{e}^{-\mathrm{i}\omega t}\,x(t) \,. \tag{14.28}$$

Here, and eleswhere unless otherwise noted, integrals range over all negative and positive values, i. e. $-\infty$ to $+\infty$.

The inverse Fourier transform expresses the signal as a function of time in terms of the Fourier transform,

$$x(t) = \frac{1}{2\pi}\int \mathrm{d}\omega\;\mathrm{e}^{\mathrm{i}\omega t}\,X(\omega) \,. \tag{14.29}$$

These expressions for the transform and inverse transform can be shown to be self-consistent. A key fact in the proof is that the Dirac delta function can be written as an integral over all time,

$$\delta(\omega) = \frac{1}{2\pi}\int \mathrm{d}t\;\mathrm{e}^{\pm\mathrm{i}\omega t} \,, \tag{14.30}$$

and similarly

$$\delta(t) = \frac{1}{2\pi}\int \mathrm{d}\omega\;\mathrm{e}^{\pm\mathrm{i}\omega t} \,. \tag{14.31}$$

Because a delta function is an even function of its argument, it does not matter if the + or − sign is used in these equations.

Reality and Symmetry

The Fourier transform $X(\omega)$ is generally complex. However, signals like $x(t)$ are real functions of time. In that connection (14.29) would seem to pose a problem, because it expresses the real function x as an integral

Amazon
EX1013 Page 517

involving the complex exponential multiplied by the complex Fourier transform. The requirement that x be real leads to a simple requirement on its Fourier transform X. The requirement is that $X(-\omega)$ must be the complex conjugate of $X(\omega)$, i. e., $X(-\omega) = X^*(\omega)$. That means that

$$\text{Re } X(-\omega) = \text{Re } X(\omega) \tag{14.32}$$

and

$$\text{Im } X(-\omega) = -\text{Im } X(\omega) \, .$$

Similar reasoning leads to special results for signals $x(t)$ that are even or odd functions of time t. If x is even $[x(-t) = x(t)]$ then the Fourier transform X is not complex but is entirely real. If x is odd $[x(-t) = -x(t)]$ then the Fourier transform X is not complex but is entirely imaginary.

The polar form of the Fourier transform is normally a more useful representation than the real and imaginary parts. It is the product of a magnitude, or absolute value, and an exponential phase factor,

$$X(\omega) = |X(\omega)| \, \exp[\mathrm{i}\varphi(\omega)] \, . \tag{14.33}$$

The magnitude is a positive real number. Negative or complex values of X arise from the phase factor. For instance, if X is entirely real then $\varphi(\omega)$ can only be zero or 180°.

14.3.1 Examples

A few example Fourier transforms are insightful.

The Gaussian

The Fourier transform of a Gaussian is a Gaussian. The Gaussian function of time is

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} \, \mathrm{e}^{-t^2/(2\,\sigma^2)} \, . \tag{14.34}$$

The function is normalized to unit area, in the sense that the integral of $g(t)$ over all time is 1.0. The Fourier transform is

$$G(\omega) = \mathrm{e}^{-\omega^2\sigma^2/2} \, . \tag{14.35}$$

The Unit Rectangle Pulse

The unit rectangle pulse, $r(t)$, is a function of time that is zero except on the interval $-T_0/2$ to $T_0/2$. During that interval the function has the value $1/T_0$, so that the function has unit area. The Fourier transform of this pulse is

$$R(\omega) = [\sin(\omega T_0/2)]/(\omega T_0/2) \, , \tag{14.36}$$

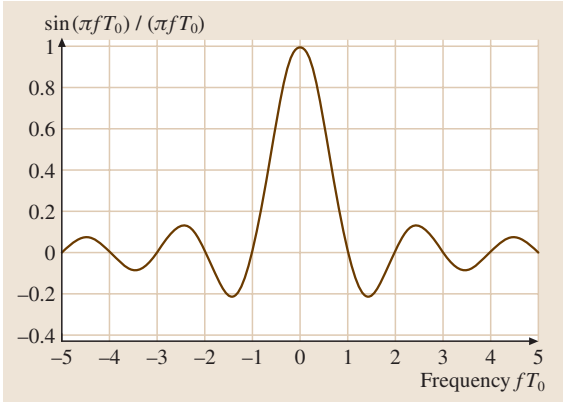


Fig. 14.4 The Fourier transform of a single pulse with duration T_0 as a function of frequency f expressed in dimensionless form fT_0

or, in terms of frequency

$$R(f) = [\sin(\pi f T_0)]/(\pi f T_0) \, ,$$

as shown in Fig. 14.4.

The function of the form $(\sin x)/x$ is sometimes called the sinc function. However, $(\sin \pi x)/(\pi x)$ is also called the sinc function. Therefore, whenever the sinc function is used by name it must be defined.

Both the Gaussian and the unit rectangle illustrate a reciprocal effect sometimes called the *uncertainty principle*. The Gaussian function of time $g(t)$ is narrow if σ is small because σ appears in the denominator of the exponential in $g(t)$. Then the Fourier transform $G(\omega)$ is wide because σ appears in the numerator of the exponential in $G(\omega)$. Similarly, the unit rectangle is narrow if T_0 is small. Then the Fourier transform $R(\omega)$ is broad because $R(\omega)$ depends on the product ωT_0. The general statement of the principle is that, if a function of one variable (e.g. time) is compact, then the transform representation, that is the function of the conjugate variable (e.g. frequency), is broad, and vice versa. The extreme expression of the uncertainty principle appears in the Fourier transform of a function that is constant for all time. According to (14.30), that transform is a delta function of frequency. Conversely, the Fourier transform of a delta function is a constant for all frequency. That means that the spectrum of an ideal impulse contains all frequencies equally.

A contrast between the Fourier transforms of Gaussian and rectangle pulses is also revealing. Because the Gaussian is a smooth function of time, the transform has a single peak. Because the rectangle has sharp edges,

Amazon
EX1013 Page 518

there are oscillations in the transform. If the rectangle is given sloping or rounded edges, the amplitude of the oscillations is reduced.

14.3.2 Time-Shifted Function

If $y(t)$ is a time-shifted version of $x(t)$, i. e.

$$y(t) = x(t - t_0) \,, \tag{14.37}$$

then the Fourier transform of y is related to the Fourier transform of x by the equation

$$Y(\omega) = \exp(-\mathrm{i}\omega t_0) X(\omega) \,. \tag{14.38}$$

The transform Y is the same as X except for a phase shift that increases linearly with frequency. There are two important implications of this equation. First, because the magnitude of the exponential with imaginary argument is 1.0, the magnitude of Y is the same as the magnitude of X for all values of ω. Second, reversing the logic of the equation shows that, if the phase of a signal is changed in such a way that the phase shift is a linear function of frequency, then the change corresponds only to a shift along the time axis for the function of time, and not to a distortion of the shape of the wave. A general phase-shift function of frequency can be separated into the best-fitting straight line and a residual. Only the residual distorts the shape of the signal as a function of time.

14.3.3 Derivatives and Integrals

If $v(t)$ is the derivative of $x(t)$, i. e., $v(t) = \mathrm{d}x/\mathrm{d}t$, then the Fourier transform of v is related to the transform of x by the equation

$$V(\omega) = \mathrm{i}\omega X(\omega) \,. \tag{14.39}$$

Thus, differentiating a signal is equivalent to ideal high-pass filtering with a boost of 6 dB per octave, i. e., doubling the frequency doubles the ratio of the output to the input, as processed by the differentiator. Differentiating also leads to a simple phase shift of 90° ($\pi/2$ radians) in the sense that the new factor of i equals $\exp(\mathrm{i}\pi/2)$. The differentiation equation can be iterated. The Fourier transform of the n-th derivative of $x(t)$ is $(\mathrm{i}\omega)^n X(\omega)$.

Integration is the inverse of differentiation, and that fact becomes apparent in the Fourier transforms. If $w(t)$ is the integral of $x(t)$, i. e.,

$$w(t) = \int_{-\infty}^{t} \mathrm{d}t' \, x(t') \,, \tag{14.40}$$

then the Fourier transform of w is related to the Fourier transform of x by the equation

$$W(\omega) = X(\omega)/(\mathrm{i}\omega) + X(0)\delta(\omega) \,. \tag{14.41}$$

The first term above could have been anticipated based on the transform of the derivative. The second term corresponds to the additive constant of integration that always appears in the context of an integral. The number $X(0)$ is the average (DC) value of the signal $x(t)$, and if this average value is zero then the second term can be neglected.

14.3.4 Products and Convolution

If the signal x is the product of two functions y and w, i. e. $x(t) = y(t)w(t)$ then, according to the convolution theorem, the Fourier transform of x is the convolution of the Fourier transforms of y and w, i. e.

$$X(\omega) = \frac{1}{2\pi} Y(\omega) * W(\omega) \,. \tag{14.42}$$

The convolution, indicated by the symbol *, is defined by the integral

$$X(\omega) = \frac{1}{2\pi} \int \mathrm{d}\omega' \, Y(\omega') \, W(\omega - \omega') \,. \tag{14.43}$$

The convolution theorem works in reverse as well. If X is the product of Y and W, i. e.

$$X(\omega) = Y(\omega) \, W(\omega) \tag{14.44}$$

then the functions of time, x, y, and w are related by a convolution,

$$x(t) = \int_{-\infty}^{\infty} \mathrm{d}t' \, y(t') \, w(t - t') \tag{14.45}$$

or

$$x(t) = y(t) * w(t) \,.$$

The symmetry of the convolution equations for multiplication of functions of frequency and multiplication of functions of time is misleading. Multiplication of frequency functions, e.g. $X(\omega) = Y(\omega)W(\omega)$, corresponds to a linear operation on signals generally known as filtering. Multiplication of signal functions of time, e.g. $y(t)w(t)$, is a nonlinear operation such as modulation.

Amazon
EX1013 Page 519

14.4 Power, Energy, and Power Spectrum

The instantaneous power in a signal is defined as $P(t) = x^2(t)$. This definition corresponds to the power that would be transferred by a signal to a unit load that is purely resistive, or dissipative. Such a load is not at all reactive, wherein energy is stored for some fraction of a cycle.

The energy in a signal is the accumulation of power over time,

$$E = \int \mathrm{d}t\; P(t) = \int \mathrm{d}t\; x^2(t)\,. \tag{14.46}$$

At this point, a distinction must be made between finite-duration signals and infinite-duration signals. For a finite-duration signal, the above integral exits. By substituting the Fourier transform for $x(t)$, one finds that

$$E = \frac{1}{2\pi}\int \mathrm{d}\omega\; X(\omega)\; X(-\omega) \quad \text{or} \quad \int \mathrm{d}\omega\; E(\omega)\,. \tag{14.47}$$

Thus the energy in the signal is written as the accumulation of of the energy spectral density,

$$E(\omega) = \frac{1}{2\pi} X(\omega)\; X(-\omega) = \frac{1}{2\pi}|X(\omega)|^2\,. \tag{14.48}$$

The symmetry between (14.46) and (14.47) is known as Parseval's theorem. It says that one can compute the energy in a signal by either a time or a frequency integral.

The power spectral density is obtained by dividing the energy spectral density by the duration of the signal, T_D,

$$P(\omega) = E(\omega)/T_\mathrm{D}\,. \tag{14.49}$$

For white noise, the power density is constant on average, $P(\omega) = P_0$. From (14.47) it is evident that a signal cannot be white over the entire range of frequencies out to infinite frequency without having infinite energy. One is therefore limited to noise that is white over a finite frequency band.

For pink noise the power density is inversely proportional to frequency, $P(\omega) = c/\omega$, where c is a constant. The energy integral in (14.47) for pink noise also diverges. Therefore, pink noise must be limited to a finite frequency band.

Turning now to infinite-duration signals, for an infinite-duration signal the energy is not well defined. It is likely that one would never even think about an infinite-duration signal if it were not for the useful concept of a *periodic* signal. Although the energy is undefined, the power P is well defined, and so is the power spectrum, or power spectral density $P(\omega)$. As expected, the power is the integral of the power spectral density,

$$P = \int \mathrm{d}\omega\; P(\omega)\,, \tag{14.50}$$

where $P(\omega)$ is given in terms of X from (14.27),

$$P(\omega) = \frac{\pi}{2}\sum_{n=0}^{\infty}|X_n|^2[\delta(\omega-\omega_n)+\delta(\omega+\omega_n)]\,. \tag{14.51}$$

It is not hard to convert densities to different units. For instance, the power spectral density can be written in terms of frequency f instead of ω ($\omega = 2\pi f$). By the definition of a density we must have that

$$P = \int \mathrm{d}f P(f)\,. \tag{14.52}$$

This definition is consistent with the fact that a delta function has dimensions that are the inverse of its argument dimensions. Therefore, $\delta(\omega) = \delta(2\pi f) = \delta(f)/(2\pi)$, and

$$P(f) = \frac{1}{4}\sum_{n=0}^{\infty}|X_n|^2[\delta(f-f_n)+\delta(f+f_n)]\,. \tag{14.53}$$

14.4.1 Autocorrelation

The autocorrelation function a_f of a signal $x(t)$ provides a measure of the similarity between the signal at time t and the same signal at a different time, $t+\tau$. The variable τ is called the *lag*, and the autocorrelation function is given by

$$a_\mathrm{f}(\tau) = \int_{-\infty}^{\infty} \mathrm{d}t\; x(t)\; x(t+\tau)\,. \tag{14.54}$$

When τ is zero then the integral is just the square of $x(t)$, and this leads to the largest possible value for the autocorrelation, namely E. For a signal of finite duration, the autocorrelation must always be strictly less than its value at $\tau = 0$. Consequently, the normalized autocorrelation function $a(\tau)/a(0)$ is always less than 1.0 ($\tau \neq 0$).

By substituting (14.29) for $x(t)$ one finds a frequency integral for the autocorrelation function,

$$a_\mathrm{f}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\; \mathrm{e}^{\mathrm{i}\omega t}|X(\omega)|^2\,, \tag{14.55}$$

Amazon
EX1013 Page 520

or, from (14.47),

$$a_{\mathrm{f}}(\tau) = \int_{-\infty}^{\infty} \mathrm{d}\omega \; \mathrm{e}^{\mathrm{i}\omega\tau} E(\omega) \,. \tag{14.56}$$

Equation (14.56) says that the autocorrelation function is the Fourier transform of the energy spectral density. This relationship is known as the Wiener–Khintchine relation. Because $E(-\omega) = E(\omega)$, one can write a_{f} in a way that proves that it is a real function with no imaginary part,

$$a_{\mathrm{f}}(\tau) = 2 \int_{0}^{\infty} \mathrm{d}\omega \; \cos(\omega\tau) E(\omega) \,. \tag{14.57}$$

Furthermore, because the cosine is an even function of its argument $[a_{\mathrm{f}}(-\tau) = a_{\mathrm{f}}(\tau)]$, the autocorrelation function only needs to be given for positive values of the lag.

A signal does not have finite duration if it is periodic. Then the autocorrelation function is defined as

$$a(\tau) = \lim_{T_{\mathrm{D}} \to \infty} \frac{1}{2T_{\mathrm{D}}} \int_{-T_{\mathrm{D}}}^{T_{\mathrm{D}}} \mathrm{d}t \; x(t) x(t+\tau) \,. \tag{14.58}$$

If the period is T then $a(\tau) = a(\tau + nT)$ for all integer n, and the maximum value occurs at $a(0)$ or $a(nT)$. Because of the factor of time in the denominator of (14.58), the function $a(\tau)$ is the Fourier transform of the power spectral density and not of the energy spectral density.

A critical point for both $a_{\mathrm{f}}(\tau)$ and $a(\tau)$ is that autocorrelation functions are independent of the phases of spectral components. This point seems counterintuitive because waveforms depend on phases and it seems only natural that the correlation of a waveform with itself at some later time should reflect this phase dependence. However, the fact that autocorrelation is the Fourier transform of the energy or power spectrum proves that the autocorrelation function must be independent of phases because the spectra are independent of phases.

For example, if $x(t)$ is a periodic function with zero average value, it is defined by (14.6). Then it is not hard to show that the autocorrelation function is given by

$$a(\tau) = \frac{1}{2} \sum_{n=1}^{N} C_n^2 \; \cos(\omega_n \tau) \,. \tag{14.59}$$

The autocorrelation function is only a sum of cosines with none of the phase information. Only the harmonic frequencies and amplitudes play a role.

14.4.2 Cross-Correlation

Parallel to the autocorrelation function, the cross-correlation function is a measure of the similarity of the signal $x(t)$ to the signal $y(t)$ at a different time, i. e. the similarity to $y(t+\tau)$. The cross-correlation function is

$$\rho_o(\tau) = \int \mathrm{d}t \; x(t) \; y(t+\tau) \,. \tag{14.60}$$

In practice, the cross-correlation is usually normalized,

$$\rho(\tau) = \frac{\int \mathrm{d}t \; x(t) \; y(t+\tau)}{[\int \mathrm{d}t_1 \; x^2(t_1) \;\; \int \mathrm{d}t_2 \; y^2(t_2)]^{1/2}} \,, \tag{14.61}$$

so that the maximum value of $\rho(\tau)$ is equal to 1.0. Unlike the autocorrelation function, the maximum of $\rho(\tau)$ does not necessarily occur at $\tau = 0$. For example, if signal $y(t)$ is the same as signal $x(t)$ except that $y(t)$ has been delayed by T_{del} then $\rho(\tau)$ has its maximum value 1.0 when $\tau = T_{\mathrm{del}}$.

14.5 Statistics

Measured signals are always finite in length. Definitions of statistical terms for measured signals, together with their continuum limits are given in this section.

The number of samples in a measurement is N. The *duration* of the measured signal is T_{D}, and $T_{\mathrm{D}} = N\delta t$, where δt is the inverse of the sample rate.

The sampled signal has values x_i, $(1 \le i \le N)$, and the continuum analog is the signal $x(t)$, $(0 \le t \le T_{\mathrm{D}})$.

The *average* value, or mean, is

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \text{or} \quad \frac{1}{T_{\mathrm{D}}} \int_{0}^{T_{\mathrm{D}}} \mathrm{d}t \; x(t) \,. \tag{14.62}$$

The *variance* is

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2$$

Amazon
EX1013 Page 521

or

$$\frac{1}{T_D}\int_0^{T_D} dt\,[x(t)-\overline{x}]^2\,. \tag{14.63}$$

The *standard deviation* is the square root of the variance, $\sigma=\sqrt{\sigma^2}$.

The *energy* is

$$E=\delta t\sum_{i=1}^{N}x_i^2 \quad \text{or} \quad \int_0^{T_D} dt\,x^2(t)\,. \tag{14.64}$$

The *average power* is

$$\overline{P}=\frac{1}{N}\sum_{i=1}^{N}x_i^2 \quad \text{or} \quad \frac{E}{T_D}\,. \tag{14.65}$$

The *root-mean-square* (RMS) value is the square root of the average power, $x_{\mathrm{RMS}}=\sqrt{\overline{P}}$.

14.5.1 Signals and Processes

Signals are the observed results of processes. A process is *stationary* if its stochastic properties, such as mean and standard deviation, do not change during the time for which a signal is observed. Signals provide incomplete glimpses into processes.

The best estimate of the mean of the underlying process is equal to the mean of an observed signal. The expected error in the estimate of the mean of the underlying process, the so-called *standard deviation of the mean*, is

$$s=\sigma/\sqrt{N}\,, \tag{14.66}$$

where N is the number of data points contributing to the mean of the observed signal.

14.5.2 Distributions

Digitized signals are often regarded as sampled data $\{x\}$. If the data are integers or are put into bins j then the probability that the signal has value x_j is the probability mass function $\mathrm{PMF}(j)=N_j/N$, the ratio of the number of samples in bin j to the total number of samples. If data are continuous floating-point numbers, the analogous distribution is the probability density function $\mathrm{PDF}(x)$. In terms of these distributions, the mean is given by

$$\overline{x}=\sum x_j\,\mathrm{PMF}(j) \quad \text{or} \quad \int_{-\infty}^{\infty} dx\,x\,\mathrm{PDF}(x)\,. \tag{14.67}$$

The most important PDF is the normal (Gaussian) density $G(x)$,

$$G(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp[(x-\overline{x})^2/2\sigma^2]\,. \tag{14.68}$$

Like all PDFs, $G(x)$ is normalized to unit area, i. e.

$$\int_{-\infty}^{\infty} dx\,G(x)=1\,. \tag{14.69}$$

The probability that x lies between some value x_1 and x_1+dx is $\mathrm{PDF}(x_1)\,dx$, and normalization reflects the simple fact that x must have *some* value.

The probability that variable x is less than some value x_1 is the cumulative distribution function (CDF),

$$\mathrm{CDF}(x_1)=\int_{-\infty}^{x_1} dx'\,\mathrm{PDF}(x')\,. \tag{14.70}$$

If the density is normal, the integral is the cumulative normal distribution (CND),

$$\mathrm{CND}(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} dx'\exp(x'^2/2\sigma^2)\,. \tag{14.71}$$

It is convenient to think of the CND as a function of x compared to the standard deviation, i. e., as a function of $y=(x-\overline{x})/\sigma$, as shown in Fig. 14.5.

$$C(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} dy'\exp(y'^2/2)\,. \tag{14.72}$$

Because of the symmetry of the normal density,

$$C(-y)=1-C(y)\,. \tag{14.73}$$

Therefore, it is enough to know $C(y)$ for $y>0$. A few important values follow.

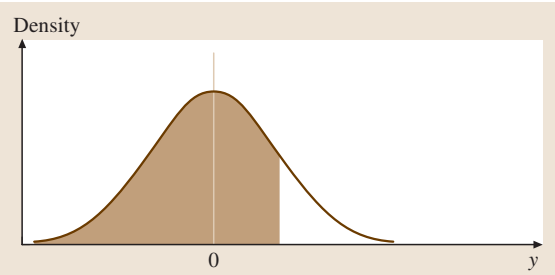


Fig. 14.5 The *area* under the normal density is the cumulative normal. Here the area is the function $C(y)$

Amazon
EX1013 Page 522

Table 14.1 Selected values of the cumulative normal distribution

C(0)	0.5000
C(0.675)	0.7500
C(1)	0.8413
C(2)	0.9773
C(3)	0.9987
C(∞)	1.0000

Table 14.1 can be used to find probabilities. For example, the probability that a normally distributed variable lies between its mean and its mean plus a standard deviation, i. e., between $\overline{x}$ and $\overline{x}+\sigma$, is $C(1)-0.5=0.3413$. The probability that it lies within plus or minus two standard deviations ($\pm 2\sigma$) of the mean is $2[C(2)-0.5]=0.9546$.

The importance of the normal density lies in the central limit theorem, which says that the distribution for a sum of random variables approaches a normal distribution as the number of variables becomes large. In other words, if the variable x is a sum

$$x = x_1 + x_2 + x_3 + \ldots x_N = \sum_{i=1}^{N} x_i \,, \tag{14.74}$$

then no matter how the individual x_i are distributed, x will be normally distributed in the limit of large N.

14.5.3 Multivariate Distributions

A multivariate distribution is described by a *joint probability density* PDF(x, y), where the probability that variable x has a value between x_1 and $x_1 + \mathrm{d}x$ and simultaneously variable y has a value between y_1 and $y_1 + \mathrm{d}y$ is

$$P(x_1, y_1) = \mathrm{PDF}(x_1, y_1)\mathrm{d}x\ \mathrm{d}y \,. \tag{14.75}$$

The normalization requirement is

$$\int \mathrm{d}x \int \mathrm{d}y\, \mathrm{PDF}(x, y) = 1 \,. \tag{14.76}$$

The *marginal probability density* for x, PDF(x), is the probability density for x itself, regardless of the value of y. Hence,

$$\mathrm{PDF}(x) = \int \mathrm{d}y\, \mathrm{PDF}(x, y) \,. \tag{14.77}$$

The y dependence has been integrated out.

The *conditional probability density* PDF$(x|y)$ describes the probability of a value x, given a specific value of y, for instance, if $y = y_1$, then

$$\mathrm{PDF}(x|y_1) = \mathrm{PDF}(x, y_1) / \int \mathrm{d}x\, \mathrm{PDF}(x, y_1) \,. \tag{14.78}$$

or

$$\mathrm{PDF}(x|y_1) = \mathrm{PDF}(x, y_1) / \mathrm{PDF}(y_1) \,. \tag{14.79}$$

The probability that $x = x_1$ and $y = y_1$ is equal to the probability that $y = y_1$ multiplied by the conditional probability that if $y = y_1$ then $x = x_1$, i. e.,

$$P(x_1, y_1) = P(x_1|y_1)P(y_1) \,. \tag{14.80}$$

Similarly, the probability that $x = x_1$ and $y = y_1$ is equal to the probability that $x = x_1$ multiplied by the conditional probability that if $x = x_1$ then $y = y_1$, i. e.

$$P(x_1, y_1) = P(y_1|x_1)P(x_1) \,. \tag{14.81}$$

The two expressions for $P(x_1, y_1)$ must be the same, and that leads to Bayes's Theorem,

$$P(x_1|y_1) = P(y_1|x_1)P(x_1)/P(y_1) \,. \tag{14.82}$$

14.5.4 Moments

The m-th moment of a signal is defined as

$$\overline{x^m} = \frac{1}{N}\sum_{i=1}^{N} x_i^m \quad \text{or} \quad \frac{1}{T_\mathrm{D}}\int_0^{T_\mathrm{D}} \mathrm{d}t\, x^m(t) \,. \tag{14.83}$$

Hence the first moment is the mean (14.62) and the second moment is the average power (14.65).

The m-th central moment is

$$\mu_m = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^m \quad \text{or} \quad \frac{1}{T_\mathrm{D}}\int_0^{T_\mathrm{D}} \mathrm{d}t\, [x(t) - \overline{x}]^m \,. \tag{14.84}$$

The first central moment is zero by definition. The second central moment is the alternating-current (AC) power, which is equal to the average power (14.65) minus the time-independent (or DC) component of the power.

The third central moment is zero if the signal probability density function is symmetrical about the mean. Otherwise, the third moment is a simple way to measure how the PDF is skewed. The *skewness* is the normalized

Amazon
EX1013 Page 523

third moment,

$$\text{skewness} = \mu_3/\mu_2^{3/2} . \tag{14.85}$$

The fourth central moment leads to an impression about how much strength there is in the wings of a probability density compared to the standard deviation. The normalized fourth moment is the *kurtosis*,

$$\text{kurtosis} = \mu_4/\mu_2^2 . \tag{14.86}$$

For instance, the kurtosis of a normal density, which has significant wings, is 3. But the kurtosis of a rectangular density, which is sharply cut off, is only 9/5.

14.6 Hilbert Transform and the Envelope

The Hilbert transform of a signal $x(t)$ is $\mathcal{H}[x(t)]$ or function $x_\mathrm{I}(t)$, where

$$x_\mathrm{I}(t) = \mathcal{H}[x(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \mathrm{d}t' \, \frac{x(t')}{t-t'} . \tag{14.87}$$

Some facts about the Hilbert transform are stated here without proof. Proofs and further applications may be found in appendices to [14.1].

First, the Hilbert transform is its own inverse, except for a minus sign,

$$x(t) = \mathcal{H}[x_\mathrm{I}(t)] = -\frac{1}{\pi} \int_{-\infty}^{\infty} \mathrm{d}t' \, \frac{x_\mathrm{I}(t')}{t-t'} . \tag{14.88}$$

Second, a signal and its Hilbert transform are orthogonal in the sense that

$$\int \mathrm{d}t \, x(t) \, x_\mathrm{I}(t) = 0 . \tag{14.89}$$

Third, the Hilbert transform of $\sin(\omega t + \varphi)$ is $-\cos(\omega t + \varphi)$, and the Hilbert transform of $\cos(\omega t + \varphi)$ is $\sin(\omega t + \varphi)$.

Further the Hilbert transform is linear. Consequently, for any function for which a Fourier transform exists,

$$\begin{aligned} &\mathcal{H}\left[\sum_n A_n \cos(\omega_n t) + B_n \sin(\omega_n t)\right] \\ &= \sum_n A_n \sin(\omega_n t) - B_n \cos(\omega_n t) \end{aligned} \tag{14.90}$$

or

$$\begin{aligned} &\mathcal{H}\left[\sum_n C_n \sin(\omega_n t + \varphi_n)\right] \\ &= -\sum_n C_n \cos(\omega_n t + \varphi_n) \\ &= \sum_n C_n \sin(\omega_n t + \varphi_n - \pi/2) . \end{aligned} \tag{14.91}$$

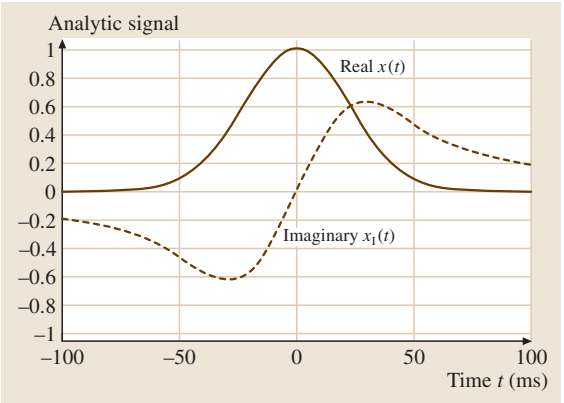


Fig. 14.6 A Gaussian pulse $x(t)$ and its Hilbert transform $x_\mathrm{I}(t)$ are the real and imaginary parts of the analytic signal corresponding to the Gaussian pulse

Comparing the two sine functions above makes it clear why a Hilbert transform is sometimes called a 90-degree rotation of the signal.

Figure 14.6 shows a Gaussian pulse, $x(t)$, and its Hilbert transform, $x_\mathrm{I}(t)$. The Gaussian pulse was made by adding up 100 cosine harmonics with amplitudes given by a Gaussian spectrum per (14.35). The Hilbert transform was computed by using the same amplitude spectrum and replacing all the cosine functions by sine functions.

Figure 14.6 illustrates the difficulty often encountered in computing the Hilbert transform using the time integrals that define the transform and its inverse. If we had to calculate $x(t)$ by transforming $x_\mathrm{I}(t)$ using (14.88) we would be troubled by the fact that $x_\mathrm{I}(t)$ goes to zero so slowly. An accurate calculation of $x(t)$ would require a longer time span than that shown in the figure.

14.6.1 The Analytic Signal

The *analytic signal* $\tilde{x}(t)$ for $x(t)$ is given by the complex sum of the original signal and an imaginary part equal

Amazon
EX1013 Page 524

to the Hilbert transform of $x(t)$,

$$\tilde{x}(t) = x(t) + \mathrm{i}\, x_\mathrm{I}(t)\,. \tag{14.92}$$

The analytic signal, in turn, can be used to calculate the envelope of signal $x(t)$. The envelope $e(t)$ is the absolute value – or magnitude – of the analytic signal

$$e(t) = |\tilde{x}(t)|\,. \tag{14.93}$$

For instance, if $x(t) = A\cos(\omega t + \varphi)$, then $x_\mathrm{I}(t) = A\sin(\omega t + \varphi)$ and

$$\tilde{x}(t) = A[\cos(\omega t + \varphi) + \mathrm{i}\sin(\omega t + \varphi)]\,. \tag{14.94}$$

By Euler's theorem

$$\tilde{x}(t) = A\exp[\mathrm{i}(\omega t + \varphi)]\,, \tag{14.95}$$

and the absolute value is

$$\begin{aligned} e(t) &= \{A\exp[\mathrm{i}(\omega t + \varphi)]A\exp[-\mathrm{i}(\omega t + \varphi)]\}^{1/2} \\ &= A\,. \end{aligned} \tag{14.96}$$

14.7 Filters

Filtering is an operation on a signal that is typically defined in frequency space. If $x(t)$ is the input to a filter and $y(t)$ is the output then the Fourier transforms of x and y are related by

$$Y(\omega) = H(\omega)X(\omega)\,, \tag{14.97}$$

where $H(\omega)$ is the transfer function of the filter. The transfer function has a magnitude and a phase

$$H(\omega) = |H(\omega)|\exp[\mathrm{i}\Phi(\omega)]\,. \tag{14.98}$$

The frequency-dependent magnitude is the amplitude response, and it characterizes the filter type – low pass, high pass, bandpass, band-reject, etc. The phase $\Phi(\omega)$ is the phase shift for a spectral component with frequency ω. The amplitude and phase responses of a filter are explicitly separated by taking the natural logarithm of the transfer function

$$\ln H(\omega) = \ln[|H(\omega)|] + \mathrm{i}\Phi(\omega)\,. \tag{14.99}$$

Because $\ln|H| = \ln(10)\log|H|$,

$$\ln H(\omega) = 0.1151\Gamma(\omega) + \mathrm{i}\Phi(\omega)\,, \tag{14.100}$$

where G is the filter gain in decibels, and Φ is the phase shift in radians.

14.7.1 One-Pole Low-Pass Filter

The one-pole low-pass filter serves as an example to illustrate filter concepts. This filter can be made from a single resistor (R) and a single capacitor (C) with a time constant $\tau = RC$. The transfer function of this filter is

$$H(\omega) = \frac{1}{1 + \mathrm{i}\omega\tau} = \frac{1 - \mathrm{i}\omega\tau}{1 + \omega^2\tau^2}\,. \tag{14.101}$$

The filter is called *one-pole* because there is a single value of ω for which the denominator of the transfer function is zero, namely $\omega = 1/(\mathrm{i}\tau) = -\mathrm{i}/\tau$.

The magnitude (or amplitude) response is

$$|H(\omega)| = \sqrt{\frac{1}{1 + \omega^2\tau^2}}\,. \tag{14.102}$$

The filter cut-off frequency is the *half-power* point (or 3-dB-down point), where the magnitude of the transfer function is $1/\sqrt{2}$ compared to its maximum value. For the one-pole low-pass filter, the half-power point occurs when $\omega = 1/\tau$.

Filters are often described by their asymptotic frequency response. For a low-pass filter asymptotic behavior occurs at high frequency, where, for the one-pole filter $|H(\omega)| \propto 1/\omega$. The $1/\omega$ dependence is equivalent to a high-frequency slope of $-6\,\mathrm{dB/octave}$, i. e., for octave frequencies,

$$L_2 - L_1 = 20\log\left(\frac{\omega_1}{2\omega_1}\right) = 20\log\frac{1}{2} = -6\,. \tag{14.103}$$

A filter with an asymptotic dependence of $1/\omega^2$ has a slope of $-12\,\mathrm{dB/octave}$, etc.

The phase shift of the low-pass filter is the arctangent of the ratio of the imaginary and real parts of the transfer function,

$$\Phi(\omega) = \tan^{-1}\left(\frac{\mathrm{Im}[H(\omega)]}{\mathrm{Re}[H(\omega)]}\right)\,, \tag{14.104}$$

Amazon
EX1013 Page 525

which, for the one-pole filter, is $\Phi(\omega) = \tan^{-1}(-\omega\tau)$. The phase shift is zero at zero frequency, and approaches 90° at high frequency. This phase behavior is typical of simple filters in that important phase shifts occur in frequency regions where the magnitude shows large attenuation.

14.7.2 Phase Delay and Group Delay

The phase shifts introduced by filters can be interpreted as delays, whereby the output is delayed in time compared to the input. In general, the delay is different for different frequencies, and therefore, a complex signal composed of several frequencies is bent out of shape by the filtering process. Systems in which the delay is different for different frequencies are said to be *dispersive*.

Two kinds of delay are of interest. The *phase delay* simply reinterprets the phase shift as a delay. The phase delay T_φ is given by $T_\varphi = -\Phi(\omega)/\omega$. The *group delay* T_g is given by the derivative $T_g = -\mathrm{d}\Phi(\omega)/\mathrm{d}\omega$. Phase and group delays for a one-pole low-pass filter are shown in Fig. 14.7 together with the phase shift.

14.7.3 Resonant Filters

Resonant filters, or *tuned* systems, have an amplitude response that has a peak at some frequency where $\omega = \omega_o$. Such filters are characterized by the resonant frequency, ω_o, and by the bandwidth, $2\Delta\omega$. The bandwidth is specified by half-power points such that $|H(\omega_o + \Delta\omega)|^2 \approx |H(\omega_o)|^2/2$ and $|H(\omega_o - \Delta\omega)|^2 \approx |H(\omega_o)|^2/2$. The sharpness of a tuned system is often quoted as a Q value, where Q is a dimensionless number given by

$$Q = \omega_o/(2\Delta\omega)\,. \tag{14.105}$$

As an example, a two-pole low-pass filter with a resonant peak near the angular frequency ω_o is described by the transfer function

$$H(\omega) = \frac{\omega_o^2}{\omega_o^2 - \omega^2 + j\omega\omega_o/Q}\,. \tag{14.106}$$

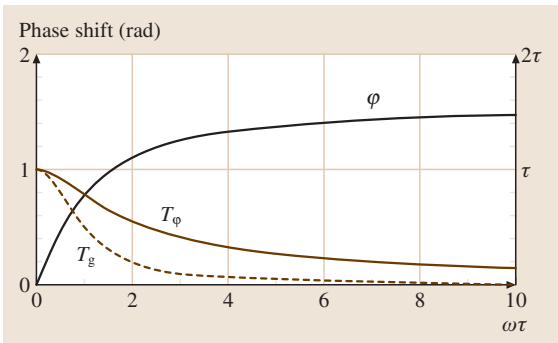


Fig. 14.7 The phase shift Φ for a one-pole low-pass filter can be read on the *left ordinate*. The phase and group delays can be read on the *right ordinate*

14.7.4 Impulse Response

Because filtering is described as a product of Fourier transforms, i. e., in frequency space, the temporal representation of filtering is a convolution

$$y(t) = \int \mathrm{d}t'\, h(t-t')x(t') = \int \mathrm{d}t'\, h(t')x(t-t')\,. \tag{14.107}$$

The two integrals on the right are equivalent.

Equation (14.107) is a special form of linear processor. A more general linear processor is described by the equation

$$y(t) = \int \mathrm{d}t'\, h(t,t')x(t')\,, \tag{14.108}$$

where $h(t, t')$ permits a perfectly general dependence on t and t'. The special system in which only the difference in time values is important, i. e. $h(t, t') = h(t - t')$, is a linear *time-invariant* system. Filters are time invariant.

A system that operates in real time obeys a further filter condition, namely *causality*. A system is causal if the output $y(t)$ depends on the input $x(t')$ only for $t' < t$. In words, this says that the present output cannot depend on the future input. Causality requires that $h(t) = 0$ for $t < 0$. For the one-pole corona, low-pass filter of (14.101) the impulse response is

$$\begin{aligned} h(t) &= \frac{1}{\tau}\mathrm{e}^{-t/\tau} && \text{for } t > 0\,,\\ h(t) &= 0 && \text{for } t < 0\,,\\ h(t) &= \frac{1}{2\tau} && \text{for } t = 0\,. \end{aligned} \tag{14.109}$$

For the two-pole low-pass resonant filter of (14.106), the impulse response is

$$\begin{aligned} h(t) = {} & \frac{\omega_o}{\sqrt{1-[1/(2Q)]^2}}\,\mathrm{e}^{-\frac{\omega_o}{2Q}t} \\ & \times \sin\left\{\omega_o t\sqrt{1-[1/(2Q)]^2}\right\}\,, \quad t \geq 0\,,\\ h(t) = {} & 0\,, \quad t < 0\,. \end{aligned} \tag{14.110}$$

Amazon
EX1013 Page 526

14.7.5 Dispersion Relations

The causality requirement on the impulse response, $h(t) = 0$ for $t < 0$, has implications for the transfer function. Causality means that the real and imaginary parts of the transfer function are Hilbert transforms of one another. Specifically, if the real and imaginary parts of H are defined as $H(\omega) = H_R(\omega) + iH_I(\omega)$ then

$$H_R(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{H_I(\omega')}{\omega - \omega'}, \tag{14.111}$$

and

$$H_I(\omega) = \frac{-1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{H_R(\omega')}{\omega - \omega'}.$$

The symbol $\mathcal{P}$ signifies that the principal value of a divergent integral should be taken. In many cases, this requires no special steps, and definite integrals from integral tables give the correct answers.

These equations are known as *dispersion relations*. They arise from doing an integral in frequency space to calculate the impulse response for $t < 0$. The fact that this calculation must return zero means that $H(\omega)$ must have no singularities in the complex frequency plane for frequencies with a negative imaginary part. Similar dispersion relations apply to the natural log of the transfer function, relating the filter gain to the phase shift as in (14.100)

$$\Gamma(\omega) = \Gamma(0) - \frac{\omega^2}{0.1151\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{\Phi(\omega')}{\omega'(\omega'^2 - \omega^2)} \tag{14.112}$$

and

$$\Phi(\omega) = \frac{0.1151\,\omega}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{\Gamma(\omega')}{\omega'^2 - \omega^2}.$$

Because $\Gamma(\omega)$ is even and $\Phi(\omega)$ is odd, both integrands are even in ω', and these integrals can be replaced by twice the integral from zero to infinity. The second equation above is particularly powerful. It says that, if we want to find the phase shift of a system, we only have to measure the gain of the system in decibels, multiply by 0.1151, and do the integral. Of course, it is in the nature of the integral that in order to find the phase shift at any given frequency we need to know the gain over a wide frequency range.

The dispersion relations for gain and phase shift also arise from a contour integral over frequencies with a negative imaginary part, but now the conditions on $H(\omega)$ are more stringent. Not only must $H(\omega)$ have no poles for $\mathrm{Im}(\omega) < 0$, but $\ln H(\omega)$ must also have no poles. Consequently $H(\omega)$ must have no zeros for $\mathrm{Im}(\omega) < 0$. A system that has neither poles nor zeros for $\mathrm{Im}(\omega) < 0$ is said to be *minimum phase*. The dispersion relations in (14.112) only apply to a system that is minimum phase.

14.8 The Cepstrum

The cepstrum (pronounced *kepstrum*) is the inverse Fourier transform of the natural logarithm of the spectrum. Because it is the inverse transform of a function of frequency, the cepstrum is a function of a time-like variable. But just as the word *cepstrum* is an anagram of the word *spectrum*, the time-like coordinate is called the *quefrency*, an anagram of *frequency*. The field of cepstrology is full of word fun like this.

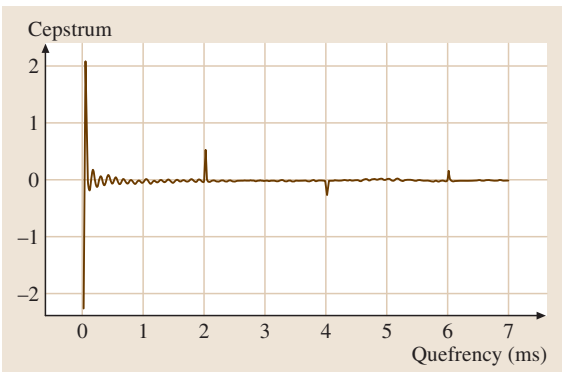


Fig. 14.8 The cepstrum of an original signal to which is added a delayed version of the same signal, with a delay of 2 ms ($a = 1$). The original signal is the sum of two female talkers

Amazon
EX1013 Page 527

The complex cepstrum of complex spectrum $Y(\omega)$ is

$$q(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \, \mathrm{e}^{\mathrm{i}\omega\tau} \ln[Y(\omega)] \,, \tag{14.113}$$

where τ is the quefrency. Because $Y(\omega) = |Y(\omega)| \, \mathrm{e}^{\mathrm{i}\varphi(\omega)}$,

$$\begin{aligned} q(\tau) = & \frac{1}{2\pi} \int_{0}^{\infty} \mathrm{d}\omega \, \mathrm{e}^{\mathrm{i}\omega\tau} [\ln |Y(\omega)| + \mathrm{i}\varphi(\omega)] \\ & + \frac{1}{2\pi} \int_{0}^{\infty} \mathrm{d}\omega \, \mathrm{e}^{-\mathrm{i}\omega\tau} [\ln |Y(-\omega)| + \mathrm{i}\varphi(-\omega)] \,. \end{aligned} \tag{14.114}$$

For a real signal $y(t)$, the magnitude $|Y(\omega)|$ is an even function of ω, and $\varphi(\omega)$ is odd. Therefore,

$$\begin{aligned} q(\tau) = & \frac{1}{\pi} \int_{0}^{\infty} \mathrm{d}\omega \, [\ln |Y(\omega)|] \cos(\omega\tau) \\ & + \frac{\mathrm{i}}{\pi} \int_{0}^{\infty} \mathrm{d}\omega \, \varphi(\omega) \sin(\omega\tau) \,. \end{aligned} \tag{14.115}$$

The real part of q comes from the magnitude, the imaginary part from the phase. The phase must be unwrapped; it cannot be artificially restricted to a 2π range.

It is common to deal only with the real part of the cepstrum q_R. It is evident that the calculation will fail if $|Y(\omega)|$ is zero. The cepstrum is not applied to theoretical objects such as periodic functions of time that have delta function spectra – hence zeros. The cepstrum is applied to measured data, where it can lead to insight into features of the underlying processes.

The cepstrum is used in the acoustical and vibrational monitoring of machinery. Bearings and other rotating parts tend to produce sounds with interleaved periodic spectra. These periodicities lead to peaks at the corresponding quefrencies, revealing features that may not be apparent in the spectrum.

The cepstrum is particularly suited to the separation of source and filter functions. If Y is a filtered version of X, where the transfer function is H, then

$$|Y(\omega)| = |H(\omega)| \, |X(\omega)| \,. \tag{14.116}$$

The logarithm operation turns the product on the right-hand side into a sum, so that

$$\begin{aligned} q_\mathrm{R}(\tau) = & \frac{1}{\pi} \int_{0}^{\infty} \mathrm{d}\omega \, [\ln |H(\omega)|] \cos(\omega\tau) \\ & + \frac{1}{\pi} \int_{0}^{\infty} \mathrm{d}\omega \, [\ln |X(\omega)|] \cos(\omega\tau) \,. \end{aligned} \tag{14.117}$$

For instance, if $|Y|$ is the spectrum of a spoken vowel, then the term involving the formant filter $|H|$ leads to a low-quefrency structure, and the term involving source spectrum $|X|$ leads to a high-quefrency peak characteristic of the glottal pulse period.

The cepstrum can reveal reflections. As a simple example, we consider a direct sound X plus its reflection with relative amplitude a and delay T_D. The sum then has a spectrum Y,

$$|Y(\omega)| = [1 + a\cos(\omega T_\mathrm{D})] |X(\omega)| \quad (a < 1) \,. \tag{14.118}$$

The logarithm of the factor in square brackets is periodic in ω with period $2\pi/T_\mathrm{D}$. The corresponding term in the cepstrum leads to a peak at quefrency $\tau = T_\mathrm{D}$, as shown in Fig. 14.8. The addition of more reflections with other delays will lead to additional peaks. Maintaining the anagram game, the separation of peaks along the quefrency axis is sometimes called *liftering*.

14.9 Noise

Noise has many definitions in acoustics. Commonly, noise is any unwanted signal. In the context of communications, it is an excitation that competes with the information that one wishes to transmit. In signal processing, noise is defined as a random signal that can only be defined in statistical terms with no long-term predictability.

14.9.1 Thermal Noise

Thermal noise, or Johnson noise, is generated in a resistor. An electrical circuit that describes this source of noise is a resistor R in series with a voltage source that depends on R, such that the RMS voltage is given by the equation

$$V = \sqrt{4Rk_\mathrm{B}T\Delta f} \,, \tag{14.119}$$

Amazon
EX1013 Page 528

where R is the resistance in ohms, k_B is Boltzmann's constant, T is the absolute temperature, and Δf is the bandwidth over which the noise is measured.

The corresponding noise power can be defined by measuring the maximum power that is transferred to a load resistor connected across the series circuit above. Maximum power occurs when the load resistor also has a resistance R and has zero temperature so that the load resistor produces no Johnson noise of its own. Then the thermal noise power is given by

$$P = k_B T \Delta f \,. \tag{14.120}$$

Because $k_B T$ has dimensions of Joules and Δf has dimensions of inverse seconds, the quantity P has dimensions of watts, as expected. Boltzmann's constant is 1.38×10^{-23} J/K, and room temperature is 293 K. Therefore, the noise power density is 4×10^{-21} W/Hz. Because the power is proportional to the first power of the bandwidth, the noise is white. Johnson noise is also Gaussian.

14.9.2 Gaussian Noise

A noise is Gaussian if its instantaneous values form a Gaussian (normal) distribution. A noise distribution is illustrated in an experiment wherein an observer makes hundreds of instantaneous measurements of a noise voltage and plots these instantaneous values as a histogram. Unless there is some form of bias, the measured values are equally often positive and negative, and so the mean of the distribution is zero. The noise is Gaussian if the histogram derived in this way is a Gaussian function. The more intense the noise, the larger is the standard deviation of the Gaussian function. Because of the central limit theorem, there is a tendency for noise to be Gaussian. However, non-Gaussian noises are easily generated. Random telegraph noise, where instantaneous values can only be $+1$ or -1, is an example.

14.9.3 Band-Limited Noise

Band-limited noise can be written in terms of Fourier components,

$$x(t) = \sum_{n=1}^{N} A_n \cos(\omega_n t) + B_n \sin(\omega_n t) \,. \tag{14.121}$$

The amplitudes A_n and B_n are defined only statistically. According to a famous paper by *Einstein* and *Hopf* [14.2], these amplitudes are normally distributed with zero mean, and the distributions of A_n and B_n have the same variance σ_n^2. The distributions themselves can be thought of as representative of an ensemble of noises, all of which are intended by the creator to be the same: same duration and power, same frequency range and bandwidth.

Because the average power in a sine or cosine is 0.5, the average power in band-limited noise is

$$P = \sum_{n=1}^{N} \sigma_n^2 \,. \tag{14.122}$$

An alternative description of band-limited noise is the amplitude and phase form

$$x(t) = \sum_{n=1}^{N} C_n \, \cos(\omega_n t + \varphi_n) \,, \tag{14.123}$$

where φ_n are random variables with a rectangular distribution from 0 to 2π, and $C_n = \sqrt{A_n^2 + B_n^2}$.

Given that A_n and B_n follow a Gaussian distribution with variance σ_n, the amplitude C_n follows a Rayleigh distribution f_{Rayl}

$$f_{\text{Rayl}}(C_n) = \frac{C_n}{\sigma_n^2} \, \mathrm{e}^{-C_n^2/(2\sigma_n^2)} \quad (C_n > 0) \,. \tag{14.124}$$

The peak of the Rayleigh distribution occurs at $C_n = \sigma$. The zeroth moment is 1.0 because the distribution is normalized. The first moment, or $\overline{C_n}$, is $\sigma_n \sqrt{\pi/2}$. The second moment is $2\sigma_n^2$, and the fourth moment is $8\sigma_n^4$.

The cumulative Rayleigh distribution can be calculated in closed form,

$$F_{\text{Rayl}}(C_n) = \int_0^{C_n} \mathrm{d}C_n' \, f_{\text{Rayl}}(C_n') = 1 - \mathrm{e}^{-C_n^2/(2\sigma_n^2)} \,. \tag{14.125}$$

14.9.4 Generating Noise

To generate the amplitudes A_n and B_n with normal distributions using a computer random-number generator, one can add up twelve random numbers and subtract 6. On the average, the amplitudes will have a normal distribution, because of the central limit theorem, with a mean of zero and a variance of 1.0.

To generate the amplitudes C_n with a Rayleigh distribution, one can transform the random numbers r_n that come from a computer random-number generator, according to the formula

$$C_n = \sigma \sqrt{-2 \ln(1 - r_n)} \,. \tag{14.126}$$

Amazon
EX1013 Page 529

14.9.5 Equal-Amplitude Random-Phase Noise

Equal-amplitude random-phase (EARP) noise is of the form

$$x(t) = C \sum_{n=1}^{N} \cos(\omega_n t + \varphi_n) \,, \tag{14.127}$$

where φ_n is again a random variable over the range 0 to 2π.

The advantage of EARP noise is that every noise sample has the same power spectrum. A possible disadvantage is that the amplitudes A_n and B_n are no longer normally distributed. Instead, they are distributed like the probability density functions for the sine or cosine functions, with square-root singularities at $A_n = \pm C$ and $B_n = \pm C$. However, the actual values of noise are normally distributed as long as the number of noise components is more than about five.

14.9.6 Noise Color

White noise has a constant spectral density, which means that the power in white noise is proportional to the bandwidth. On the average, every band with given bandwidth Δf has the same amount of power. Pink noise has a spectral density that decreases inversely with the frequency. Consequently, pink noise decreases at a rate of 6 dB per octave. On the average, every octave band has the same amount of power.

14.10 Sampled data

Converting an analog signal, such as a time-dependent voltage, into a digital representation dices the signal in two dimensions, the dimension of the signal voltage and the dimension of time. Dicing the signal voltage is known as quantization, dicing with respect to time is known as sampling.

14.10.1 Quantization and Quantization Noise

It is common for an analog-to-digital converter (ADC) to represent the values of input voltages as integers. The range of the integers is determined by the number of bits per sample in the conversion process. A conversion into an M-bit sample (or word) allows the voltage value to be represented by 2^M bits. For instance, a 10-bit ADC that is restricted to converting positive voltages would represent 0 V by the number 0 and +10 V by $2^{10} - 1$ or 1023.

A 16-bit ADC would allow 2^{16} or 65 536 different values. A 16-bit ADC that converts voltages between −10 and +10 V would represent −10 V by −32 768 and +10 V by +32 767. Conversion is linear. Thus 0.3052 V would be converted to the sample value 1000 and 0.3055 V to the value 1001. A voltage of 0.3053 would also be converted to a value of 1000, no different from 0.3052. The discrepancy is an error known as the quantization error or *quantization noise*.

Quantization noise referenced to the signal is a signal-to-noise ratio. Standard practice makes this ratio as large as possible by assuming a signal with the maximum possible power. For the positive and negative ADC described above, maximum power occurs for a square wave between a sampled waveform value of $-2^{(M-1)}$ and $+2^{(M-1)}$. The power is the square of the waveform or $\frac{1}{4} \times 2^{2M}$.

For its part, the noise is a random variable that represents the difference between an accurately converted voltage and the actual converted value as limited by the number of bits in the sample word. This error is never more than 0.5 and never less than −0.5. The power in noise that fluctuates randomly over the range −0.5 to +0.5 is 1/12. Consequently the signal-to-noise (S/N) ratio is 3×2^{2M}. Expressed in decibels, this value is $10 \log(3 \times 2^{2M})$, or $20M \log(2) + 4.8$ dB, or $6M + 4.8$ dB. For a 16-bit word, this would be 96 + 4.8 or about 101 dB. An alternative calculation would assume that the maximum power is the power for the largest sine wave that can be reproduced by such a system. This sine has half the power of the square, and the S/N ratio is then about $6M$ dB.

14.10.2 Binary Representation

Digitized data, like a sampled waveform are represented in binary form by numbers (or words) consisting of digits 0 and 1. For example, an eight-bit word consisting of two four-bit bytes and representing the decimal number 7, would be written as

0 0 0 0 0 1 1 1 .

This number has 1 in the ones column, 1 in the twos column, 1 in the fours column, and nothing in any other

Amazon
EX1013 Page 530

column. One plus two plus four is equal to 7, which is what was desired.

An eight-bit word ($M = 8$) could represent decimal integers from 0 to 255. It cannot represent 2^M, which is decimal 256. If one starts with the decimal number 255 and adds 1, the binary representation becomes all zeros, i. e. $255 + 1 = 0$. It is like the 100 000-mile odometer on an automobile. If the odometer reads 99 999 and the car goes one more mile, the odometer reads 00 000.

Signals are generally negative as often as they are positive, and that leads to a need for a binary representation of negative numbers. The usual standard is a representation known as *twos-complement*. In twos-complement representation, any number that begins with a 1 is negative. Thus, the leading digit serves as a *sign bit*.

In order to represent the number $-x$ in an M-bit system one computes $2^M - x$. That way, if one adds x and $-x$ one ends up with 2^M, which is zero.

A convenient algorithm for calculating the twos-complement of a binary number is to reverse each bit, 0 for 1 and 1 for 0, and then add 1. Thus, in an eight-bit system the number -7 is given by

1 1 1 1 1 0 0 1 .

14.10.3 Sampling Operation

The sampling process replaces an analog signal, which is a continuous function of time, by a sequence of points. The operation is equivalent to the process shown in Fig. 14.9, where the analog signal $x(t)$ is multiplied by a train of evenly spaced delta functions to create a sequence of sampled values $y(t)$.

Intuitively, it seems evident that this operation is a sensible thing to do if the delta functions come along rapidly enough – rapid compared to the speed of the temporal changes in the waveform. That concept is most clearly seen by studying the Fourier transforms of functions x, s and y.

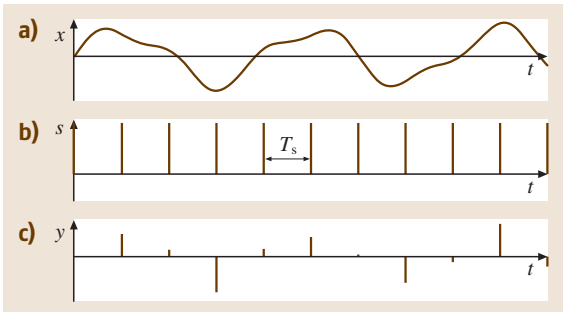


Fig. 14.9 An analog signal $x(t)$ is multiplied by a train of delta functions $s(t)$ to produce a sampled signal $y(t)$

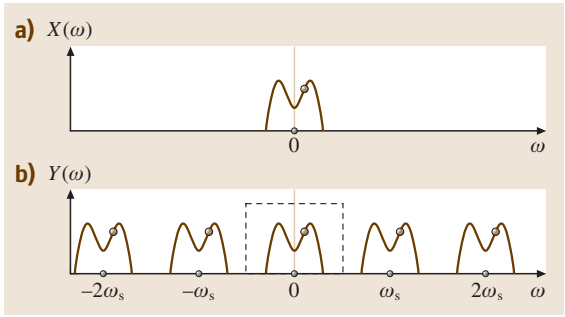


Fig. 14.10 **(a)** The spectrum of the analog signal $X(\omega)$ is bounded in frequency. **(b)** The spectrum of the sampled signal, $Y(\omega)$, is the convolution of $X(\omega)$ and the Fourier transform of the sampling train of delta functions. Consequently, multiple images of $X(\omega)$ appear. Frequencies that are allowed by the sampling theorem are included in the *dashed box*. A particular frequency (*circled*) is followed through the multiple imaging

The Fourier transform of the analog signal is $X(\omega)$, with a spectrum that is limited to some highest frequency ω_{max}. By contrast, the Fourier transform of the train of delta functions is, itself, a train of delta functions, $S(\omega)$ that extends over the entire frequency axis. Because the delta functions in time have period T_s, the delta functions in $S(\omega)$ are separated by ω_s, equal to $2\pi/T_s$. Because $y(t)$ is the product of the time-dependent analog signal and the train of delta functions, the Fourier transform $Y(\omega)$ is the convolution of $X(\omega)$ and $S(\omega)$, as shown in part (b) of Fig. 14.10. Because of the convolution operation, $Y(\omega)$ includes multiple images of the original spectrum.

It is evident from Fig. 14.10b that, if the ω_{max} is less than half of ω_s, the multiple images will not overlap. That observation has the status of a theorem known as the *sampling theorem*, which says that the sampled signal is an adequate representation of an analog signal if the sample rate is more than twice the highest frequency in the analog signal, i. e., $\omega_s > 2\omega_{max}$.

As an example of a failure to apply the sampling theorem, suppose that a 600 Hz sine tone is sampled at a rate of 1000 Hz. The spectrum of the sampled signal will contain 600 Hz as expected, and it will also contain a component at $1000 - 600 = 400$ Hz. The 400 Hz component was not present in the original spectrum; it is an alias, an unwanted image of the 600 Hz tone.

Amazon
EX1013 Page 531

14.10.4 Digital-to-Analog Conversion

In converting a signal from digital to analog form, one can begin with the train of delta functions that is signal $y(t)$ as shown in Fig. 14.9c. An electronic device to do that is a digital-to-analog converter (DAC). However, as shown in Fig. 14.10b, this signal includes many high frequencies that are unwanted byproducts of the sampling process. Consequently, one needs to low-pass filter the signal so as to pass only the frequencies less than half the sample rate, i. e., the frequencies in the dashed box. Such a low-pass filter is called a *reconstruction* filter.

Practical DACs do not produce delta-function voltage spikes. Instead, they produce rectangular functions with durations pT_s, where p is a fraction of a sample period $0 < p \leq 1$. If $p = 1$, the output of the DAC resembles a staircase function. Mathematically, replacing the delta function train of Fig. 14.9c by the train of rectangles is equivalent to convolving the function $y(t)$ with a rectangular function. The consequence of this convolution is that the output is filtered, and the transfer function of the filter is the Fourier transform of the rectangle. The magnitude of the transfer function is

$$|H(\omega)| = \frac{\sin(\omega p T_s/2)}{\omega p T_s/2} \,. \tag{14.128}$$

The phase shift of the filter is a pure delay and consequently unimportant. The effective filtering that results from the rectangles, known as *sin(x)-over-x* filtering, can be corrected by the reconstruction filter.

14.10.5 The Sampled Signal

This brief section will introduce a notation that will be useful in later discussions of sampled signals. It is supposed at the outset that one begins with a total of N samples, equally spaced in time by the sampling period T_s. By convention, the first sample occurs at time $t = 0$ and the last sample occurs at time $t = (N-1)T_s$. Consequently, the signal duration is $T_D = (N-1)T_s$.

In dealing with sampled signals, it is common to replace the time variable with a discrete index k. Thus,

$$x(t) = x(kT_s) = x_k \,, \tag{14.129}$$

where the equation on the left indicates that the original data exist only at discrete time values.

14.10.6 Interpolation

The discrete-time values of a sampled waveform can be used to compute an approximate Fourier transform of the original signal. This Fourier transform is valid up to a frequency as high as half the sample rate, i. e., as high as $\omega_s/2$, or π/T_s. The Fourier transform can then be used to estimate the values of the original signal $x(t)$ at times other than the sample times. In this way, the Fourier transform computed from the samples serves to interpolate between the samples. Such an interpolation scheme proceeds as follows.

First, the Fourier transform is

$$X(\omega) = T_s \sum_k x_k \, \exp(-\mathrm{i}\omega T_s k) \,, \tag{14.130}$$

where, as noted above, x_k is the signal $x(t)$ at the times $t = T_s k$, and the leading factor of T_s gets the dimensions right.

Then the inverse Fourier transform is

$$x(t) = \frac{T_s}{2\pi} \int_{-\omega_s/2}^{\omega_s/2} \mathrm{d}\omega \, \mathrm{e}^{\mathrm{i}\omega t} \sum_k x_k \, \mathrm{e}^{-\mathrm{i}\omega T_s k} \,. \tag{14.131}$$

Reversing the order of sum and integral and using the fact that $T_s\omega_s/2 = \pi$, we find that

$$x(t) = \sum_k x_k \, \frac{\sin \pi(t/T_s - k)}{\pi(t/T_s - k)} \,. \tag{14.132}$$

The sinc function is 1.0 whenever $t = T_s k$, and is zero whenever t is some other integer multiple of T_s. Therefore, the sum on the right only interpolates; it does not change the values of $x(t)$ when t is equal to a sample time.

14.11 Discrete Fourier Transform

The Fourier transform of a signal with finite duration is well defined in principle. The finite signal itself can be regarded as some base function that is multiplied by a rectangular window to limit the duration. Then the Fourier transform proceeds by convolving with the transform of the window. For example, a truncated exponentially decaying sine function can be regarded as a decaying sine, with the usual infinite duration, multiplied by a rectangular window. Then the Fourier transform of the truncated function is the Fourier trans-

Amazon
EX1013 Page 532

form of the decaying sine convolved with a sinc function – the Fourier transform of the rectangular window. Such a Fourier transform is a function of a continuous frequency, and it shows the broad spectrum associated with the abrupt truncation.

In digital signal processing the frequency axis is not continuous. Instead, the Fourier transform of a signal is defined at discrete frequencies, just as the signal itself is defined at discrete time points. This kind of Fourier transform is known as the discrete Fourier transform (DFT).

To compute the DFT of a function, one begins by periodically repeating the function over the entire time axis. For example, the truncated decaying sine in Fig. 14.11a is repeated in Fig. 14.11b where it should be imagined that the repetition precedes indefinitely to the left and right.

Then the Fourier transform of the periodically repeated signal becomes a Fourier series. The fundamental frequency of the Fourier series is the reciprocal of the duration, $f_0 = 1/T_\mathrm{D}$, and the spectrum becomes a set of discrete frequencies, which are the harmonics of f_0. For instance, if the signal is one second in duration, the spectrum consists of the harmonics of 1 Hz, and if the duration is two seconds then the spectrum has all the harmonics of 0.5 Hz. As expected, the highest harmonic is limited to half the sample rate. That Fourier series is the DFT. Using $\underline{x_k}$ to define the periodic repetition of the original discrete function, x_k, the DFT $\underline{X}(\omega)$ is defined for $\omega = 2\pi n/T_\mathrm{D}$, where n indicates the n-th harmonic. In terms of the fundamental angular frequency $\omega_o = 2\pi/T_\mathrm{D}$, the DFT is

$$\underline{X}(n\omega_o) = T_\mathrm{s} \sum_{k=0}^{N-1} x_k \, \mathrm{e}^{-\mathrm{i}n\omega_o k T_\mathrm{s}} \,, \qquad (14.133)$$

where the prefactor T_s keeps the dimensions right. The product $\omega_o T_\mathrm{s}$ is equal to $\omega_o T_\mathrm{D}/(N-1)$ or $2\pi/(N-1)$, and so

$$\underline{X}(n\omega_o) = T_\mathrm{s} \sum_{k=0}^{N-1} x_k \, \mathrm{e}^{-\mathrm{i}2\pi nk/(N-1)} \,. \qquad (14.134)$$

Both positive and negative frequencies occur in the Fourier transform. Because the maximum frequency is equal to $[1/(2T_\mathrm{s})]/(1/T_\mathrm{D})$ times the fundamental frequency, the number of discrete positive frequencies is $(N-1)/2$, and the number of discrete negative frequencies is the same. Consequently the inverse DFT can be

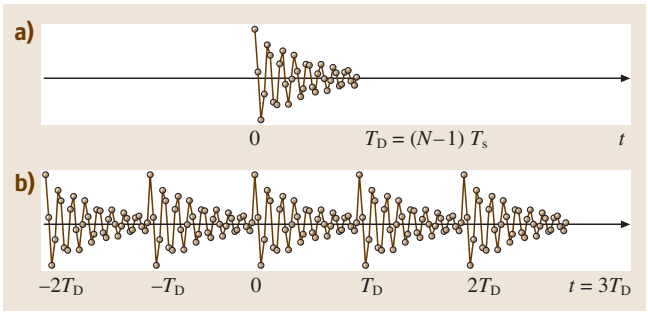


Fig. 14.11a,b A decaying function in part **(a)** is periodically repeated in part **(b)** to create a periodic signal with period T_D

written

$$\underline{x_k} = \frac{1}{T_\mathrm{D}} \sum_{n=-(N-1)/2}^{(N-1)/2} X(n\omega_o) \, \mathrm{e}^{\mathrm{i}2\pi nk/(N-1)} \,, \qquad (14.135)$$

or

$$\underline{x}(t) = \frac{1}{T_\mathrm{D}} \sum_{n=-(N-1)/2}^{(N-1)/2} X(n\omega_o) \, \mathrm{e}^{\mathrm{i}n\omega_o t} \,. \qquad (14.136)$$

A virtue of the DFT is that the information in the DFT is exactly what is needed to create the original truncated function $x(t)$ – no more and no less. The fact that the DFT spectrum actually creates the periodically repeated function $\underline{x_k}$ and not the original x_k is not a problem if we agree in advance to ignore $\underline{x_k}$ for k outside the range of the original time-limited signal. However, it should be noted that certain operations, such as time translations, products, and convolution, that have familiar characteristics in the context of the Fourier transform, retain those characteristics only for the periodically extended signal $\underline{x_k}$ and its Fourier transform $\underline{X}(n\omega_o)$ and not for the finite-duration signal.

14.11.1 Interpolation for the Spectrum

It is possible to estimate the Fourier transform at values of frequency between the harmonics of ω_o. The procedure begins with the definition of the Fourier transform of a finite function,

$$X(\omega) = \int_0^{T_\mathrm{D}} \mathrm{d}t \; x(t) \mathrm{e}^{-\mathrm{i}\omega t} . \qquad (14.137)$$

Next, the function $x(t)$ is replaced by the inverse DFT from (14.136), and the variable of integration t is replaced by t', which has symmetrical upper and lower

Amazon
EX1013 Page 533

limits,

$$X(\omega) = \frac{1}{T_D}\int_{-T_D/2}^{T_D/2} dt'\, e^{-i\omega t'} \times \sum_{n=-(N-1)/2}^{(N-1)/2} X(n\omega_o)\, e^{in\omega_o t'}\, e^{-i\omega T_D/2}\, e^{in\omega_o T_D/2}\,, \tag{14.138}$$

which reduces to

$$X(\omega) = \sum_{n=-(N-1)/2}^{(N-1)/2} X(n\omega_o)\frac{\sin[(\omega - n\omega_o)T_D/2]}{(\omega - n\omega_o)T_D/2} \times e^{-i\omega T_D/2}\, e^{i\pi n}\,. \tag{14.139}$$

14.12 The *z*-Transform

Like the discrete Fourier transform, the z-transform is well suited to describing sampled signals. We consider $x(t)$ to be a sampled signal so that it is defined at discrete time points $t = t_k = kT_s$, where T_s is the sampling period. Then the time dependence of x can be described by an index, $x_k = x(t_k)$. The z-transform of x is

$$X(z) = \sum_{k=-\infty}^{\infty} x_k z^{-k}\,. \tag{14.140}$$

The quantity z is complex, with amplitude A and phase φ,

$$z = A\, e^{i\varphi} = A\, e^{i\omega T_s}\,, \tag{14.141}$$

where φ is the phase advance in radians per sample.

In the special case where $A = 1$, all values of z lie on a circle of radius 1 (the *unit circle*) in the complex z plane. In that case the z-transform is equivalent to the discrete Fourier transform. An often-overlooked alternative view is that the z-transform is an extension of the Fourier transform wherein the angular frequency ω becomes complex,

$$\omega = \omega_R + i\omega_I\,, \tag{14.142}$$

so that

$$z = e^{-\omega_I T_s}\, e^{i\omega_R T_s}\,. \tag{14.143}$$

The extended Fourier transform will not be pursued further in this chapter.

A well-defined z-transform naturally includes a function of variable z, but the function itself is not enough. In order for the inverse transform to be unique, the definition also requires that the region of the complex plane in which the transform converges must also be specified.

Table 14.2 z-Transform pairs

x_k	$X(z)$	Radius of convergence
δ_{k,k_0}	z^{-k_0}	all z
$a^k u_k$	$z/(z-a)$	$\lvert z\rvert > a$
$ka^k u_k$	$az/(z-a)^2$	$\lvert z\rvert > a$
$a^k \cos(\omega_o T_s k)u_k$	$\frac{z^2 - az\cos(\omega_o T_s)}{z^2 - 2az\cos(\omega_o T_s) + a^2}$	$\lvert z\rvert > a$
$a^k \sin(\omega_o T_s k)u_k$	$\frac{az\sin(\omega_o T_s)}{z^2 - 2az\cos(\omega_o T_s) + a^2}$	$\lvert z\rvert > a$

To illustrate that point, one can consider two different functions x_k that have the same z-transform function, but different regions of convergence.

Consider first the function

$$x_k = 2^k \quad \text{for} \quad k \geq 0\,, \qquad x_k = 0 \quad \text{for} \quad k < 0\,. \tag{14.144}$$

This two-line function can be written as a single line by using the discrete Heaviside function u_k. The function u_k is defined as zero when k is a negative integer and as $+1$ when k is any other integer, including zero. Then x_k above becomes

$$x_k = 2^k u_k\,. \tag{14.145}$$

The z-transform of x_k is

$$X(z) = \sum_{k=0}^{\infty} (2/z)^k\,. \tag{14.146}$$

The sum is a geometric series, which converges to

$$X(z) = \frac{1}{1 - 2/z} = z/(z-2) \tag{14.147}$$

Amazon
EX1013 Page 534

if $|z| > 2$. The region of convergence is therefore the entire complex plane except for the portion inside and on a circle of radius 2.

Next consider the function

$$x_k = -2^k u_{-k-1} \,. \tag{14.148}$$

The z-transform of x_k is

$$X(z) = -\sum_{k=-\infty}^{-1} (2/z)^k \quad \text{or} \quad -\sum_{k=1}^{\infty} (z/2)^k \,. \tag{14.149}$$

The sum converges to

$$X(z) = -\frac{(z/2)}{1-z/2} = \frac{z}{z-2} \tag{14.150}$$

if $|z| < 2$. The function is identical to the function in (14.147), but the region of convergence is now the portion of the complex plane entirely inside the circle of radius 2.

The inverse z-transform is given by a counterclockwise contour integral circling the origin

$$x_k = \frac{1}{2\pi \mathrm{i}} \oint_C \mathrm{d}z X(z) z^{k-1} \,. \tag{14.151}$$

The contour C must lie entirely within the region of convergence of x and must enclose all the poles of $X(z)$.

The regions of convergence when the functions x and y are combined in some way are at least the intersection of the regions of convergence for x and y separately. Scaling and time reversal lead to regions of convergence that are scaled and inverted, respectively. For instance, if $X(z)$ converges in the region between radii r_1 and r_2, them $X(1/z)$ converges in the region between $1/r_2$ and $1/r_1$.

14.12.1 Transfer Function

The output of a process at time point k, namely y_k, may depend on the inputs x at earlier times and also on the outputs at earlier times. In equation form,

$$y_k = \sum_{q=0}^{Nq} \beta_q x_{k-q} - \sum_{p=1}^{Np} \alpha_p y_{k-p} \,. \tag{14.152}$$

This equation can be z-transformed using the time-shift property in Table 14.3,

$$\sum_{q=0}^{Nq} \beta_q z^{-q} X(z) = \sum_{p=0}^{Np} \alpha_p z^{-p} Y(z) \,, \tag{14.153}$$

where $\alpha_0 = 1$. The transfer function is the ratio of the transformed output over the transformed input,

$$H(z) = Y(z)/X(z) \,, \tag{14.154}$$

which is

$$H(z) = \frac{\sum_{q=0}^{Nq} \beta_q z^{-q}}{\sum_{p=0}^{Np} \alpha_p z^{-p}} \,. \tag{14.155}$$

From the fundamental theorem of algebra, the numerator of the fraction above has Nq roots and the denominator has Np roots, so that $H(z)$ can be written as

$$H(z) = \frac{(1-q_1 z^{-1})(1-q_2 z^{-1}) \ldots (1-q_{Nq} z^{-1})}{(1-p_1 z^{-1})(1-p_2 z^{-1}) \ldots (1-p_{Np} z^{-1})} \,. \tag{14.156}$$

This equation and its development are of central importance to digital filters, also known as linear time-invariant systems. If the system is recursive, outputs from a previous point in time are sent back to the input. Therefore, some values of the coefficients α_p are finite for $p > 1$ and so are the values of some poles, such as p_2. Such filters are called *infinite impulse response* (IIR) filters because it is possible that the response of the system to an impulse put in at time zero will never entirely die out. Some of the output is always fed back into the input. A similar conclusion is reached by recognizing that the expansion of $1/(1-pz^{-1})$ in powers of z^{-1} goes on forever. Because the system has poles, there are concerns about stability.

If the system is nonrecursive, no values of the output are ever sent back to the input. Therefore, the denominator of $H(z)$ is simply the number 1. Such filters are called *finite impulse response* (FIR) filters because their response to a delta function input will always die out eventually as long as Nq is finite. The system is said to be an *all-zero* system. The order of the filter is estab-

Table 14.3 Properties of the z-transform

Property	Signal	z-transform
Definition	x_k	$X(z)$
Linearity	$ax_k + by_k$	$aX(z) + bY(z)$
Time shift	x_{k-k_o}	$z^{-k_o} X(z)$
Scaling z	$a^k x_k$	$X(z/a)$
Time reversal	x_{-k}	$X(1/z)$
Derivative w.r.t. z	kx_k	$-z\,\mathrm{d}X(z)/\mathrm{d}z$
Convolution	$x_k * y_k$	$X(z)Y(z)$
Multiplication	$x_k y_k$	$\frac{1}{2\pi \mathrm{i}} \oint_C \mathrm{d}z'/z'\ X(z')Y(z/z')$

Amazon
EX1013 Page 535

lished by Nq or Np, the number of time points back to the earliest input or output that contribute to the current output value.

The formal z-transform,

$$H(z) = \sum_{k=-\infty}^{\infty} h_k z^{-k} \tag{14.157}$$

leads to conclusions about causality and stability.

A filter is causal if the current value of the output does not depend on future inputs. For a causal filter h_k is zero for $k < 0$. Then this sum has no terms with positive powers of z, and the region of convergence of $H(z)$ includes $|z| = \infty$.

A filter is stable if

$$S = \sum_{k=-\infty}^{\infty} |h_k| \tag{14.158}$$

is finite. It follows, that $H(z)$ is finite for $|z| = 1$, i. e., for z on the unit circle. Thus, if the region of convergence includes the unit circle, the filter is stable.

14.13 Maximum Length Sequences

A maximum length sequence (MLS) is a train of ones and zeros that makes a useful signal for measuring the impulse response of a linear system. An MLS can be generated by a bit-shift register, which resembles a bucket brigade. To make an N-bit MLS, one needs an N-stage shift register. Each stage can hold either a one or a zero. The register is imagined to have a clock which synchronizes the transfer of bits from each stage to the next. On every clock tick the content of each stage of the register is transferred to the next stage down the line. The content of the last stage is regarded as the output of the register, and it is also fed back into the first stage. In addition, the output can be fed back into one or more of the other stages, and when that occurs the stage receiving the output, in addition to the content of the previous stage, performs an exclusive OR (XOR) operation on those two inputs. The XOR operation obeys the truth table shown in Table 14.4. In words, the XOR of inputs A and B is zero if A and B are the same and is 1 if A and B are different.

Table 14.4 Truth table for the exclusive or (XOR) operation

A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

A shift register with three stages is shown in Fig. 14.12. With three stages and feedback *taps* to stages 1 and 2, it is defined as [3: 1,2].

At the instant shown in the figure, the register holds the value 1,1,1. The subsequent development of the register values is given in Table 14.5. The sequence repeats after seven steps. The table shows that every possible pattern of *ones* and *zeros* occurs once, and only once, before the pattern repeats. There are $2^N - 1 = 2^3 - 1 = 7$ such patterns. There is one exception, namely the pattern 0,0,0. If this pattern should ever appear in the register then the process gets stuck forever. Therefore, this pattern is not allowed. The output sequence is the contents of the stage on the right, here, 1,1,0,0,1,0,1. Because all seven register patterns appear before repetition, this

Table 14.5 Successive values in the shift register of Fig. 14.12

Step			
0	1	1	1
1	1	0	1
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	0	1	1
7	1	1	1
8	1	0	1
9	1	0	0

Table 14.6 Successive values in the shift register of Fig. 14.13

Step			
0	1	1	1
1	1	0	0
2	0	1	0
3	0	0	1
4	1	1	1
5	1	0	0
6	0	1	0

Amazon
EX1013 Page 536

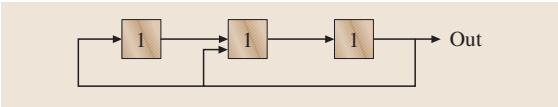


Fig. 14.12 A three-stage shift register [3: 1, 2] in which the output is fed back into the first and second stages

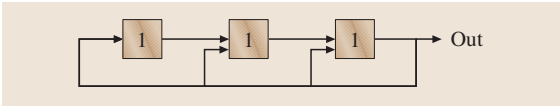


Fig. 14.13 A three-stage shift with feedback into all the stages does not produce an MLS

output is a maximum length sequence. There is nothing special about the starting register value, 1,1,1. Therefore, any cyclic permutation of the MLS is also an MLS. For instance, the sequence, 1,0,0,1,0,1,1 is the same sequence.

An example of a three-bit shift register that does not produce an MLS is [3: 1,2,3], shown in Fig. 14.13. The pattern for this shift register is shown in Table 14.6. The pattern of register values begins to repeat after only four steps. Therefore, the sequence of output values, namely 1,0,0,1,1,0,0,1,1,0,0,1, is not an MLS.

14.13.1 The MLS as a Signal

To make a signal from an MLS requires only one step: every 0 in the sequence is replaced by -1. Therefore, the MLS for the shift register in Fig. 14.12 becomes: $1, 1, -1, -1, 1, -1, 1$. For this three-stage register (N = 3) the MLS has a length of seven; there are four $+1$ values and three -1 values. These results can be generalized to an N-stage register which has $2^N - 1$ values; $2^{(N-1)}$ are $+1$ values and $2^{(N-1)} - 1$ are -1 values. The average value is therefore $1/(2^N - 1)$.

14.13.2 Application of the MLS

The key fact about an MLS is that its autocorrelation function is very nearly a delta function. To express that idea, one can write the autocorrelation function in the form appropriate for discrete samples,

$$c_k = \frac{1}{2^N - 1} \sum_{k1} x_{k1}\, x_{k1+k} \,. \tag{14.159}$$

This sum, and all sums to follow, are over the $2^N - 1$ values of the MLS sequence x. Because the sequence is cyclical, it does not matter where one starts the sum.

An MLS has the property that

$$c_k = \left(1 + \frac{1}{2^N - 1}\right) \delta_{k,0} - \frac{1}{2^N - 1} \,. \tag{14.160}$$

Therefore, c_k is approximately a Kronnecker delta function

$$c_k \approx \delta_{k,0} \,. \tag{14.161}$$

If we would like to know the impulse response h of a linear system, we can excite the system with the MLS x, and record the output y. As for filters, the linear response y is the convolution of x and h, i. e.,

$$y_k = x * h = \sum_{k1} x_{k1+k}\, h_{k1} \,. \tag{14.162}$$

To find the impulse response, one can form the quantity d, by convolving the recording y with the original MLS x, i. e.,

$$d_k = \frac{1}{2^N - 1} \sum_{k2} x_{k2+k}\, y_{k2} \tag{14.163}$$

or from (14.162)

$$d_k = \frac{1}{2^N - 1} \sum_{k1,k2} x_{k2+k}\, x_{k1+k2}\, h_{k1} \,. \tag{14.164}$$

Only $x * x$ involves the index $k2$, and doing the sum over $k2$ leads to

$$d_k = \sum_{k1} \delta_{k,k1}\, h_{k1} \tag{14.165}$$

so that $d_k = h_k$. In this way, we have found the desired impulse response.

As applied in architectural acoustics, the MLS is an alternative to recording the response to a popping balloon or gun shot. Because the MLS is continuous, it avoids the dynamic-range problem associated with an impulsive test signal, and by repeating the sequence one can achieve remarkable noise immunity.

Similarly, the MLS is an alternative to recording the response to white noise (the MLS is white). However, digital white noise, such as random telegraph noise, has an autocorrelation function that is zero only for a long-term or ensemble average. In practice, the white-noise response of a linear system is much noisier than the MLS response.

14.13.3 Long Sequences

Table 14.7 gives the taps for some MLSs generated by shift registers with 2–20 stages, i. e., orders 2–20. The

Amazon
EX1013 Page 537

Table 14.7 Taps for maximum length sequences

Number of stages	Length (bits)	Number of taps	Number of sets	Set
2	3	2	1	[2: 1,2]
3	7	2	1	[3: 1,3]
4	15	2	1	[4: 1,4]
5	31	2	1	[5: 1,4]
6	63	2	1	[6: 1,6]
7	127	2	2	[7: 1,7], [7: 1,5]
8	255	4	6	[8: 1,2,7,8]
9	511	2	1	[9: 1,6]
10	1023	2	1	[10: 1,8]
11	2047	2	1	[11: 1,10]
12	4095	4	9	[12: 1,6,7,9]
13	8191	4	33	[13: 1,7,8,9]
14	16383	4	21	[14: 1,6,9,10]
15	32767	2	3	[15: 1,9], [15: 1,12], [15: 1,15]
16	65535	4	26	[16: 1,7,10,11]
17	131071	2	3	[17: 1,12], [17: 1,13], [17: 1,15]
18	262143	2	1	[18: 1,12]
19	524287	4	79	[19: 1,10,11,14]
20	1048575	2	1	[20: 1,18]

longest sequence has a length of more than one million bits,

For orders 2, 3, and 4, there is only one possible set of taps. These sets have two taps, including the feedback to stage 1. For orders 7, 15, and 17 there is more than one set with two taps, and all of them are shown in the table.

Beginning with order 5 there are four-tap sets as well as two-tap sets, except that for some orders, such as 8, there are no two-tap sets. For every order the table gives a set with the smallest possible number of taps.

Beginning with order 7 there are six-tap sets. As the order increases the number of sets also increases. For order 19, there are 79 four-tap sets.

14.14 Information Theory

Information theory provides a way to quantify information by computing information content. The information content of a message depends on the context, and the context determines the initial uncertainty about the message. Suppose, for example, that we receive one character, but we know in advance that the context is one in which the character must be a digit between 0 and 9. Our uncertainty before receiving that actual character is described by the number of possible outcomes, which is $\Omega = 10$ in this case. Suppose instead, that the context is one in which the character must be a letter of the alphabet. Then our initial uncertainty is greater because the number of possible outcomes is now $\Omega = 26$. The first step of information theory is to recognize that, when we actually receive and identify a character, the information content of that character is greater in the second context than in the first because in the second context the character has eliminated a greater number of a priori possibilities.

The second step in information theory is to consider a message with two characters. If the context of the message is decimal digits then the number of possibilities is the product of 10 for the first digit and 10 for the second, namely $\Omega = 100$ possibilities. Compared to a message with one character, the number of possibilities has been multiplied by 10. However, it is only logical to expect that two characters will give twice as much information as one, not 10 times as much. The logical problem can be solved by quantifying information in terms of entropy, which is the logarithm of the

Amazon
EX1013 Page 538

number of possibilities

$$H = \log \Omega \,. \tag{14.166}$$

Because log 100 is just twice log 10, the logical problem is solved. The information measured in bits is obtained by using a base 2 logarithm.

A few simple features follow immediately. If the number of possible messages is $\Omega = 1$ then the message provides no information, which agrees with $\log 1 = 0$. If the context is binary, where a character can be only 1 or 0 ($\Omega = 2$), then receiving a character provides 1 bit of information, which agrees with $\log_2 2 = 1$.

If the context is an alphabet with M possible symbols, and all of the symbols are equally probable, then a message with N characters has $\Omega = M^N$ possible outcomes and the information entropy is

$$H = \log M^N = N \log M \,, \tag{14.167}$$

illustrating the additivity of information over the characters of the message.

14.14.1 Shannon Entropy

Information theory becomes interesting when the probabilities of different symbols are different. *Shannon* [14.3, 4] showed that the information content per character is given by

$$H_c = -\sum_{i=1}^{M} p_i \log p_i \,, \tag{14.168}$$

where p_i is the probability of symbol i in the given context.

The rest of this section proves Shannon's formula. The proof begins with the plausible assumption that, if the probability of symbol i is p_i, then in a very long message of N characters, the number of occurrences of character i, m_i will be exactly $m_i = Np_i$.

The number of possibilities for a message of N characters in which the set of $\{m_i\}$ is fixed by the corresponding $\{p_i\}$ is

$$\Omega = \frac{N!}{m_1!\, m_2! \,\ldots\, m_M!} \,. \tag{14.169}$$

Therefore,

$$H = \log N! - \log m_1! - \log m_2! - \ldots \log m_M! \,. \tag{14.170}$$

One can write $\log N!$ as a sum

$$\log N! = \sum_{k=1}^{N} \log k \tag{14.171}$$

and similarly for $\log m_i!$.

For a long message one can replace the sum by an integral,

$$\log N! = \int_1^N \mathrm{d}x \log x = N \log N - N + 1 \tag{14.172}$$

and similarly for $\log m_i!$.

Therefore,

$$H = N \log N - N + 1 - \sum_{i=1}^{M} m_i \log m_i + \sum_{i=1}^{M} m_i - \sum_{i=1}^{M} 1 \,. \tag{14.173}$$

Because $\sum_{i=1}^{M} m_i = N$, this reduces to

$$H = N \log N + 1 - \sum_{i=1}^{M} m_i \log m_i - M \,. \tag{14.174}$$

The information per character is obtained by dividing the message entropy by the number of characters in the message,

$$H_c = \log N - \sum_{i=1}^{M} p_i \log m_i + (1 - M)/N \,, \tag{14.175}$$

where we have used the fact that $m_i/N = p_i$.

In a long message, the last term can be ignored as small. Then because the sum of probabilities p_i is equal to 1,

$$H_c = -\sum_{i=1}^{M} p_i (\log m_i - \log N) \,, \tag{14.176}$$

or

$$H_c = -\sum_{i=1}^{M} p_i \log p_i \,, \tag{14.177}$$

which is (14.168) as advertised.

If the context of written English consists of 27 symbols (26 letters and a space), and if all symbols are equally probable, then the information content of a single character is

$$H_c = -1.443 \sum_{i=1}^{27} \frac{1}{27} \ln \frac{1}{27} = 4.75 \text{ (bits)} \,, \tag{14.178}$$

where the factor $1/\ln(2) = 1.443$ converts the natural log to a base 2 log. However, in written English all symbols are not equally probable. For example, the most

Amazon
EX1013 Page 539

common letter, 'E', is more than 100 times more likely to occur than the letter 'J'. Because equal probability of symbols always leads to the highest entropy, the unequal probability in written English is bound to reduce the information content – to about 4 bits per character. An even greater reduction comes from the redundancy in larger units, such as words, so that the information content of written English is no more than 1.6 bits per character.

The concept of information entropy can be extended to continuous distributions defined by a probability density function

$$h = -\int_{-\infty}^{\infty} \mathrm{d}x\ \mathrm{PDF}(x)\ \log[\mathrm{PDF}(x)]\ . \tag{14.179}$$

14.14.2 Mutual Information

The mutual information between sets of variables $\{i\}$ and $\{j\}$ is a measure of the amount of uncertainty about one of these variables that is eliminated by knowing the other variable. Mutual information H_m is given in terms of the joint probability mass function $p(i, j)$

$$H_\mathrm{m} = \sum_{i=1}^{M}\sum_{j=1}^{M} p(i, j) \log \frac{p(i, j)}{p(i)p(j)}\ . \tag{14.180}$$

Using written English as an example again, $p(i)$ might describe the probability for the first letter of a word and $p(j)$ might describe the probability for the second. It is convenient to let the indices i and j be integers, e.g., $p(i = 1)$ is the probability that the first letter is an 'A', and $p(j = 2)$ is the probability that the second letter is a 'B'. Then $p(1, 2)$ is the probability that the word starts with the two letters 'AB'. It is evident that in a context where the first two letters are completely independent of one another so that $p(i, j) = p(i)p(j)$ then the amount of mutual information is zero because $\log(1) = 0$. In the opposite limit the context is one in which the second letter is completely determined by the first. For instance, if the second letter is always the letter of the alphabet that immediately follows the first letter then $p(i, j) = p(j) = p(i)\delta(j, i + 1)$, and

$$H_\mathrm{m} = \sum_{i=1}^{M} p(i) \log \frac{p(i)}{p(i)p(i)} \tag{14.181}$$

which simply reduces to (14.168) for H_c, the information content of the first letter of the word.

In the general case, the mutual information is a difference in information content. It is equal to the information provided by the second letter of the word given no prior knowledge at all, minus the information provided by the second letter of the word given knowledge of the first letter. Mathematically, $p(i, j) = p(i)p(j|i)$, where $p(j|i)$ is the probability that the second letter is j given that the first letter is i. Then

$$H_\mathrm{m} = \sum_{j=1}^{M} p(j) \log \frac{1}{p(j)} - \sum_{i=1}^{M}\sum_{j=1}^{M} p(i, j) \log \frac{1}{p(j|i)}\ . \tag{14.182}$$

The information transfer ratio T is the degree to which the information in the first letter predicts the information in the second. Equivalently, it describes the transfer of information from an input to an output

$$T = \frac{-H_\mathrm{m}}{\sum_{i=1}^{M} p(i) \log p(i)}\ . \tag{14.183}$$

This ratio ranges between 0 and 1, where 1 indicates that the second letter, or output, can be predicted from the first letter, or input, with perfect reliability. The mutual information is the basis for the calculation of the information capacity of a noisy communications channel.

References

14.1 W.M. Hartmann: *Signals, Sound, and Sensation* (Springer, Berlin, New York 1998)

14.2 A. Einstein, L. Hopf: A principle of the calculus of probabilities and its application to radiation theory, Annal. Phys. **33**, 1096–1115 (1910)

14.3 C.E. Shannon: A mathematical theory of communication, Part I, Bell Syst. Tech. J. **27**, 379–423 (1948)

14.4 C.E. Shannon: A mathematical theory of communication, Part II, Bell Syst. Tech. J. **27**, 623–656 (1948)

Amazon
EX1013 Page 540

Part E Music, Speech, Electroacoustics

15 Musical Acoustics
Colin Gough, Birmingham, UK

16 The Human Voice in Speech and Singing
Björn Lindblom, Stockholm, Sweden
Johan Sundberg, Stockholm, Sweden

17 Computer Music
Perry R. Cook, Princeton, USA

18 Audio and Electroacoustics
Mark F. Davis, San Francisco, USA

Amazon
EX1013 Page 541

15. Musical Acoustics

This chapter provides an introduction to the physical and psycho-acoustic principles underlying the production and perception of the sounds of musical instruments. The first section introduces generic aspects of musical acoustics and the perception of musical sounds, followed by separate sections on string, wind and percussion instruments.

In all sections, we start by considering the vibrations of simple systems – like stretched strings, simple air columns, stretched membranes, thin plates and shells. We show that, for almost all musical instruments, the usual text-book description of such systems is strongly perturbed by material properties, geometrical factors and acoustical coupling between the drive mechanism, vibrating system and radiated sound.

For stringed, woodwind and brass instruments, we discuss excitation by the bow, reed and vibrating lips, which all involve strongly non-linear processes, even though the vibrations of the excited system usually remains well within the linear regime. However, the amplitudes of vibration of very strongly excited strings, air columns, thin plates and membranes can sometimes exceed the linear approximation limit, resulting in a number of interesting non-linear phenomena, often of significant musical importance.

Musical acoustics therefore provides an excellent introduction to the physics of both linear and non-linear acoustical systems, in a context of rather general interest to professional acousticians, teachers and students, at both school and college levels.

The subject continues its long tradition in advancing the frontiers of experimental, computational and theoretical acoustics, in an area of wide general appeal and contemporary relevance.

By discussing the theoretical models and experimental methods used to investigate the acoustics of many musical instruments, we have aimed to provide a useful background for professional acousticians, students and their teachers, for whom musical acoustics provides an exceedingly rich area for original research projects at all educational levels.

15.1 **Vibrational Modes of Instruments** 535
15.1.1 Normal Modes 535
15.1.2 Radiation from Instruments 537
15.1.3 The Anatomy of Musical Sounds .. 540
15.1.4 Perception and Psychoacoustics .. 551

15.2 **Stringed Instruments** 554
15.2.1 String Vibrations 555
15.2.2 Nonlinear String Vibrations 563
15.2.3 The Bowed String 566
15.2.4 Bridge and Soundpost 570
15.2.5 String–Bridge–Body Coupling 575
15.2.6 Body Modes 581
15.2.7 Measurements 594
15.2.8 Radiation and Sound Quality 598

15.3 **Wind Instruments** 601
15.3.1 Resonances in Cylindrical Tubes .. 602
15.3.2 Non-Cylindrical Tubes 606
15.3.3 Reed Excitation 619
15.3.4 Brass-Mouthpiece Excitation 628
15.3.5 Air-Jet Excitation 633
15.3.6 Woodwind and Brass Instruments 637

15.4 **Percussion Instruments** 641
15.4.1 Membranes 642
15.4.2 Bars 648
15.4.3 Plates 652
15.4.4 Shells 658

References 661

Because the subject is ultimately about the sounds produced by musical instruments, a large number of audio illustrations have been provided on a CD accompanying this volume, which can also be accessed by the electronic version of the Handbook on springerlink.com. The extensive list of references is intended as a useful starting point for entry to the current research literature, but makes no attempt to provide a comprehensive list of all important research.

This chapter highlights the acoustics of musical instruments. Other related topics, such as the human voice, the perception and psychology of sound, architectural acoustics, sound recording and reproduction, and many experimental, computational and analytic techniques are described in more detail elsewhere in this volume.

Amazon
EX1013 Page 542

Musical acoustics is one the oldest of all the experimental Sciences (see *Levenson* [15.1] for an informative account of the interactions between Music and Science over the ages). The observation of the relationship between the notes produced by the exact fraction divisions of a stretched string and consonant musical intervals like the octave (2:1), perfect fifth (3:2) and fourth (4:3), resulted in the first physical law to be expressed in mathematical terms. It also led to the idea of a divinely created cosmos based on exact fractions, filled with the *music of the spheres* (see, for example, Kepler's account of the ellipticity of the planetary orbits described as notes on a musical scale, in *Harmonies of the World* (1618) [15.2]). Ultimately, such observations led to Newton's discovery of celestial dynamics and the laws of gravity leading to the modern view of the universe subject to physical laws rather than numerical relationships.

In the nineteenth century, musical acoustics continued to occupy a central scientific role. This culminated in Lord Rayleigh's monumental two volumes on the *Theory of Sound* [15.3], which still provide the foundations for almost every branch of modern acoustics. The 19th century advances in understanding waves in acoustics also laid the mathematical framework for quantum wave mechanics in the early part of the 20th century. More recently, the physics of vibrating strings can be said to have come full circle, with the suggestion that string-like vibrations of the quantum field equations account for the mass to the elementary particles (*Hawkins* [15.4]).

Musical acoustics still remains a challenging and exciting field of research and continues to advance mainstream acoustics in many ways. Examples include nonlinear physics and the use of laser holography and both modal and finite-element analysis to investigate complicated vibrating systems. Such developments are described in this chapter and in more detail in other chapters of this Handbook and in the *Physics of Musical Instruments* by *Fletcher* and *Rossing* [15.5], which will often be cited, as an authoritative text and source of additional references for most topics discussed. The *Science of Sound* by *Rossing* et al. [15.6] covers an even wider range of topics at a somewhat less mathematical level. An informative overview of the history, technology and performance of western musical instruments has recently been published by *Campbell*, *Greated* and *Myers* [15.7].

The first section of this chapter deals with the generic properties of the vibrations and sounds of musical instruments. A brief description is first given of the properties of both simple and coupled resonators, typifying the vibrational modes of stringed, wind and percussion instruments, where the sound is generated by vibrating strings, air columns, plates, membranes and shells. The radiation of sound by such structures is then described in terms of multipole sources. This is followed by a brief description of the envelopes, waveforms and spectra of the sounds that characterize the sound of individual instruments. The section ends with a consideration of the way the listener perceives such sounds.

The section on stringed instruments first considers the general properties of string vibrations and their excitation by plucking, bowing and striking. Large amplitude vibrations are shown to provide a particular interesting illustration of non-linearity of much wider applicability than to musical acoustics alone. The coupling of the vibrating string via the bridge to the acoustically radiating surfaces of the instrument is then discussed in some detail, followed by a more detailed discussion of excitation of a string via the bowed slip-stick mechanism. The vibrational modes of the main shell of the instrument and the importance of the bridge and soundpost in determining the efficiency of energy transfer to the radiating surfaces of the instrument are then discussed. The section ends with a description of some of the experiment and computational techniques used to describe the vibrational modes, followed by a brief description of the radiated sound and the subjective assessment of the quality of stringed instruments.

The section on woodwind and brass instruments starts with a consideration of oscillating air columns and sound radiation from cylindrical and conical tubes and the more complicated shapes used for woodwind and brass instruments. This is followed by sections on the highly non-linear processes involved in the excitation of such vibrations by reed and lip vibrations and air-jets. The section concludes with a brief description of the acoustical properties of various wind instruments.

The final section on percussion instruments describes the acoustical properties of a range of instruments based on the vibrations of stretched membranes, bars, plates and shells. Typical waveform envelopes and time-dependent spectra are used to illustrate the relationship between the vibrational modes of such instruments and the radiated sound. Non-linearity at large amplitude excitation is again shown to be important and accounts for the characteristic sounds of certain instruments like gongs and cymbals.

Amazon
EX1013 Page 543

15.1 Vibrational Modes of Instruments

15.1.1 Normal Modes

All musical instruments produce sound via the excitation of a vibrating structure. Woodwind, brass and percussion instruments radiate sound directly. However, stringed instruments radiate sound indirectly, because the vibrating string itself radiates an insignificant amount of energy. Energy from the vibrating string therefore has to be transferred to the much larger area, acoustically efficient, radiating surfaces of the body of the instrument. The resultant modes of vibration are complex and involve the interactions and vibrations of all the component parts, such as the strings, bridge, front and back plates, soundpost, neck, and even the air inside the volume of the violin body.

Any vibrating structure, however complicated, will have a number of what are called normal modes of vibration (see Chap. 22). The important influence of damping on the nature of the normal modes will be described in the section on stringed instruments. The normal modes satisfy exactly the same equations of motion as a simple damped mass–spring resonator. The displacement ξ_n of a given excited mode measured at any chosen point p on the structure is given by

$$m_n\left(\frac{\partial^2\xi_n}{\partial t^2}+\frac{\omega_n}{Q_n}\frac{\partial\xi_n}{\partial t}+\omega_n^2\xi_n\right)=F(t)\,, \qquad (15.1)$$

where the effective mass m_n at the point p is defined in terms of the kinetic energy of the excited mode, $\frac{1}{2}m_n(\partial\xi_n/\partial t)_p^2$, $\omega_n=2\pi f_n$ is the eigenfrequency (the angular frequency) of free vibration of the excited mode in the absence of damping and Q_n is the quality factor describing its damping. Initially, we consider a local driving force $F(t)$ at the point p, though it could be applied at any chosen point on the structure or distributed over the whole surface.

The effective mass of a one-dimensional string, solid bar or air column, at the point of maximum displacement, is half the mass of the vibrating system, the factor half resulting from averaging the kinetic energy over the sinusoidal spatial displacement. Likewise, the effective mass of a two-dimensional vibrating object at maximum displacement, like a violin plate or drum skin, is of order $1/4$ its mass. The effective mass is very large close to nodal positions, where the displacement is small, and is small at antinodes, where the displacement is large.

Typical driving forces are those acting on the bridge of a bowed or plucked string instrument and the pressure fluctuations at the input end of the air column of a blown woodwind or brass instrument. Such forces are generated by highly nonlinear excitation mechanisms. In contrast, the vibrations of the vibrating structure are generally linear with displacements proportional to the driving force. However, there are important exceptions for almost all types of instruments, when nonlinearity becomes significant at sufficiently strong excitation, as discussed later.

In any continuously bowed or blown musical instruments, feedback from the vibrating system results in a periodic driving force, which will not in general be sinusoidal. Nevertheless, by the Fourier theorem, any periodic force can always be represented as a superposition of sinusoidally varying, harmonic, partials with frequencies that are integer multiple of the periodic repetition frequency. We can therefore consider the induced vibrations of any musical instruments in terms of the induced response of its vibrational modes to a harmonic series of sinusoidal driving forces.

Resonance and Admittance

In the harmonic approach, the applied forces and induced motions are assumed to vary sinusoidally as $e^{i\omega t}$. We will generally use this complex notation for notational and algebraic simplicity, where $\mathrm{Re}(e^{i\omega t})=cos\omega t$ and $\mathrm{Im}(e^{i\omega t})=\sin\omega t$. The resonant response, with displacement $\xi_n e^{i\omega t}$ and velocity $i\omega\xi_n e^{i\omega t}$ at the driving

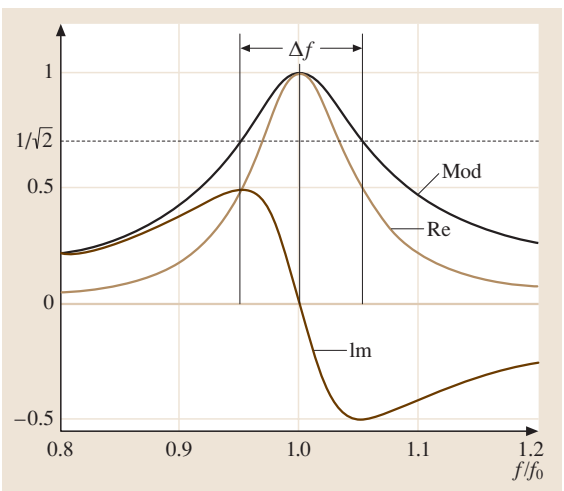


Fig. 15.1 Normalised real (Re) and imaginary (Im) components and the modulus (Mod) of the induced velocity of a simple harmonic resonator driven by a constant amplitude sinusoidal force for a Q-factor of 10

Amazon
EX1013 Page 544

point p, for an applied sinusoidal force $F\,e^{i\omega t}$, is then given by

$$\frac{\partial \xi_n}{\partial t} = i\omega\xi_n = \frac{F}{m_n}\frac{i\omega}{\left(\omega_n^2 - \omega^2 + i\omega\omega_n/Q_n\right)}. \quad (15.2)$$

The ratio of induced velocity to driving force is known as the local admittance at the driving point p, and is plotted in normalised form in Fig. 15.1 for a Q-value of 10. The admittance has both real and imaginary components. The real part describes the component of the induced velocity in phase with the driving force, while the imaginary part describes the component in phase quadrature (with phase leading that of the force by 90° degrees).

Well below resonance the induced displacement is in phase with the driving force, while at resonance the phase lags behind the driving force by 90°, and well above resonance lags by 180°. The velocity $v(t)$ leads the displacement by 90° degree and is thus in phase with the driving force at resonance, which corresponds to the maximum rate of energy transfer $i\omega F\xi_n$ to the excited mode. The 180° change in the phase of the response, as the excitation frequency passes from well below to well above resonance, is especially important in interpreting the multiple resonances of any musical instrument.

Provided the damping of an excited mode is not too strong (i. e. Q_n is significantly larger than unity), the peak in the modulus or real part of the admittance occurs at $\omega_n\left(1 - 1/8Q^2\right)$, which is very close to the natural resonant frequency ω_n. The width of the resonance is $\Delta f = f_n/Q$, where Δf is defined as the difference in frequency between the points on the resonance curve when the modulus of the induced displacement has fallen to $1/\sqrt{2}$ of its maximum value (i. e. the stored energy is half that at resonance). The displacement at resonance is $Q\times$ the static displacement.

Multi-Mode Systems

For any musical instrument having a number of vibrational modes, the admittance at the driving point p can be written as

$$A_{pp} = \sum_n \frac{1}{m_n}\frac{i\omega}{\omega_n^2 - \omega^2 + i\omega\omega_n/Q_n}, \quad (15.3)$$

with admittances A_{pp} of individual modes adding in series, equivalent to impedances in parallel. The vibrational response of a multi-resonant mode musical instrument can therefore be characterised by fitting the measured admittance to such a function giving the effective mass at the point of excitation, resonant frequency and Q-value for each of the excited modes. Using such a procedure, *Bissinger* [15.8] typically identifies up to around 40 vibrational modes for the violin below 4 kHz. However, at high frequencies, the width of individual resonances exceeds the spacing between them, making it increasingly difficult to identify individual modes.

It is important to recognise that damping is only important in a relatively narrow frequency range $\sim f_n/Q$ around the individual resonance peaks. Outside such regions, the reactive component associated with each vibrational mode continues to contribute significantly to the overall response. For example, well below resonance, each mode acts as a spring with effective spring constant $m_n(\omega_n^2 - \omega^2)$, while well above resonance it acts as a mass with effective mass $m_n(1 - \omega_n^2/\omega^2)$. The static displacement (at $\omega = 0$) is given by $\xi = F/K_o = \sum_n 1/(m_n\omega_n^2)$. Note that this involves contributions from all the vibrational modes of the structure, which is an important global property describing the low-frequency response of a multi-resonant structure such as the violin or guitar. If displacements are measured at a point p for an applied force at q, a *nonlocal admittance* can be expressed as

$$A_{pq} = \sum_n \frac{1}{m_{n,p}}\frac{i\omega}{\left(\omega_n^2 - \omega^2 + i\omega\omega_n/Q_n\right)}\frac{\xi_{n,p}\xi_{n,q}}{\xi_{n,p}^2}, \quad (15.4)$$

where $\xi_{n,p}$ and $\xi_{n,q}$ are the simultaneous displacements of the nth mode at the points p and q, with identical stored modal energy $1/2m_{n,p}\omega^2\xi_{n,p}^2 = 1/2m_{n,q}\omega^2\xi_{n,q}^2$.

Equation (15.4) illustrates the principle of reciprocity in acoustics, which states that the motion at a point p induced by a force at q is identical to the motion at q induced by the same force at p. Equation (15.4) also shows that, by applying a force at a particular position and measuring the induced motion (amplitude and phase) at a large number of different points $p(x, y, z)$ on the structure, it is possible to map out the amplitude of the modal vibrations $\xi_n(x, y, z)$ over the whole of any excited structure. Alternatively, the measurement point can be fixed and the excitation point moved across the structure. This is the basis of the powerful technique of modal analysis, which has been widely used to investigate the vibrational modes of many stringed and percussion instruments, as described by Rossing in Chap. 28 of this Handbook.

It also follows from (15.4) that a particular mode of vibration will never be excited if the driving force is located at a node of its vibrational state. This has important consequences for the spectrum of sound produced

Amazon
EX1013 Page 545

by bowed, plucked and struck stringed instruments and all percussion instruments.

Time-Domain Measurements

The vibrational characteristics of an instrument can also be investigated in the time domain. For example, by striking a stringed instrument with a light hammer or exciting the vibrational modes of a woodwind or brass instrument with a short puff of air, the frequencies of free vibration of the vibrational modes and their damping can be determined from their time-dependent decay. Provided the damping is not too strong ($Q \gg 1$), the modes will decay with time as

$$\xi_n(t) = \xi_0 \, \mathrm{e}^{-t/\tau_n} \, \mathrm{e}^{\mathrm{i}\omega_n t}, \tag{15.5}$$

where $\tau_n = 2Q_n/\omega_n = Q_n/\pi f_n$. The frequency f_n of a given mode can be determined from its inverse period and Q_n from $\pi\times$ the number of periods for the amplitude to fall by the factor exponential e. The Q-value of strongly excited modes of a musical instrument can be estimated from $\tau_{60\mathrm{dB}} = 13.6\tau$, the Sabine decay time (Chap. 10 *Concert Hall Acoustics*). This is the perceptually significant time taken for the sound pressure to fall by a factor of 10^3 – from a very loud level to just being detectable. Hence, $Q_n = \pi f_n \tau_{60}/13.6 \sim 0.23 f_n \tau_{60}$. For example, the sound of a strongly plucked cello open A-string (220 Hz) can be heard for at least ~ 2 s, corresponding to a Q-value of ≈ 100 or more.

Damping results in a loss of stored energy given by

$$\frac{\mathrm{d}E_n}{\mathrm{d}t} = -\frac{\omega_n}{Q_n} E_n = -2\frac{E_n}{\tau_n} \, . \tag{15.6}$$

Hence, the power P required to maintain a constant amplitude at resonance is $\frac{\omega_n}{Q_n} E_n$, where E_n is the energy stored. This tends to be the way that Q is defined and measured by physicists, whereas in acoustic spectroscopy it is more usual to define and measure Q-values from either the width of resonances in spectroscopic measurements or from decay times after transient excitation. As illustrated above, all such definitions are equivalent.

15.1.2 Radiation from Instruments

Although a large number of vibrational modes of a musical instrument may be excited simultaneously, they will not be equally important in radiating sound, which has important consequences for the quality of the sound. This section therefore provides a brief introduction to the radiation of sound from the vibrational modes of musical instruments.

Sound Waves in Air

In free space, the longitudinal displacement $\xi(x,t) = \xi_0 \, \mathrm{e}^{\mathrm{i}(\omega t - kx)}$ of plane sound waves satisfies the wave equation

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{1}{c_0^2} \frac{\partial^2 \xi}{\partial t^2} \, . \tag{15.7}$$

The dispersionless (independent of frequency) velocity of sound $c_0 = \sqrt{\gamma P_0/\rho}$, where $\gamma (\approx 1.4)$ is the ratio of specific heats at constant pressure and volume, P_0 ($\approx 10^5$ Pa or N/m^2) is the ambient pressure and ρ (≈ 1 kg/m^3) is the density (the brackets give the values for air at ambient pressure and temperature). The ratio of acoustic pressure $p = -\gamma P_0 \partial\xi/\partial x$ to the particle velocity $v = \partial\xi/\partial t$ is referred to as the specific impedance, $z_0 = p/v = \rho c_0$.

The appearance of γ in the expression for the velocity of sound reflects the adiabatic nature of acoustic waves. This arises because acoustic wavelengths are far too long to allow any significant equalisation of the longitudinal temperature fluctuations arising from the compressions and rarefactions of a sound wave. In free space longitudinal heat flow between the fluctuating regions is only important at very high ultrasonic frequencies (MHz), where it leads to significant attenuation. The major source of attenuation of freely propagating acoustic sound waves arises from the water vapour present. However, both viscous and transverse thermal losses to the side walls of woodwind and brass instruments can result in significant attenuation, as described later.

The above expressions neglect first-order, nonlinear, corrections to the compressibility, proportional to $\partial\xi/\partial x$, and other inertial correction terms in the nonlinear Navier–Stokes equation. This approximation breaks down at the very high intensities in the bores of the trumpet and trombone when played very loudly [15.9], which results in shockwave propagation, with a transition from a relatively smooth to a very brassy sound (*son cuivré* in French). For the present, such corrections will be neglected.

The speed of sound in air depends on the temperature θ (degrees centigrade) and, to a lesser extent, on the humidity. For 50% humidity,

$$c_0(\theta) = 332 \, (1 + \theta/273)^{1/2} \approx 332(1 + 1.710^{-3}\theta) \, , \tag{15.8}$$

giving a value of 343 m/s at 20 °C. Note that the air inside a woodwind or brass instrument, once the instrument is *warmed up*, will always be warm and humid, which will affect the playing pitch.

Amazon
EX1013 Page 546

Pressure and Intensity

The intensity I of sound radiated by a musical instrument is given by the flow of acoustic energy ($1/2\rho v^2$ per unit volume) crossing unit area per unit time,

$$I = \frac{1}{2}\rho c_0 v_{\max}^2 = \frac{1}{2} z_0 v_{\max}^2 = \frac{1}{2z_0} p_{\max}^2 \,. \qquad (15.9)$$

Sound pressure levels (SPL) are measured in dB relative to a reference sound pressure p_0 of 2×10^{-5} Pa or $\mathrm{N/m^2}$, so that $\mathrm{SPL(dB)} = 20\log_{10}(p/p_0)$. The reference pressure is approximately equal to the lowest level of sound that can be heard at around 1–3 kHz in a noise-free environment. Relative changes in sound pressure levels are given by $20\log_{10}(p_1/p_2)$ dB. A sound pressure of 2×10^{-5} Pa is very close to an intensity of $I_0 = 10^{-12}$ $\mathrm{W/m^2}$, which is used to define the almost identical intensity level, $\mathrm{IL(dB)} = 10\log_{10}(I/I_0)$. The difference between the factor 10 and 20 arises because sound intensity is proportional to the square of the sound pressure.

Spherical Waves

In free space, sound from a localised source will propagate as a spherical wave satisfying the three-dimensional wave equation (*Fletcher* and *Rossing* [15.5], Sect. 6.2), with pressure

$$p(r) = A\frac{\mathrm{e}^{\mathrm{i}(\omega t - kr)}}{r} \qquad (15.10)$$

and particle velocity

$$v(r) = \frac{A}{z_0}\left(1 + \frac{1}{\mathrm{i}kr}\right)\frac{\mathrm{e}^{\mathrm{i}(\omega t - kr)}}{r} \,. \qquad (15.11)$$

Near and Far Fields

Note that, unlike plane-wave solutions, the velocity and pressure differ in phase by an amount that depends on the distance from the source and the wavelength. Close to the source, in the *near field* ($kr \ll 1$), the pressure and induced velocity are in phase quadrature. Such terms therefore involve no work being done (proportional to $\int pv\,\mathrm{d}t$) and hence no radiation of sound. The near-field term describes the motion of the air that is forced to vibrate backwards and forwards with the vibrating surface of the source, which simply adds inertial mass to the vibrating mode. This term is responsible for the *end correction* ($\Delta L \sim a$, where a is the pipe radius), which extends the effective length of an open-ended vibrating air column. The additional inertial mass also lowers the vibrational frequency of the relatively light vibrating membranes of a stretched drum skin.

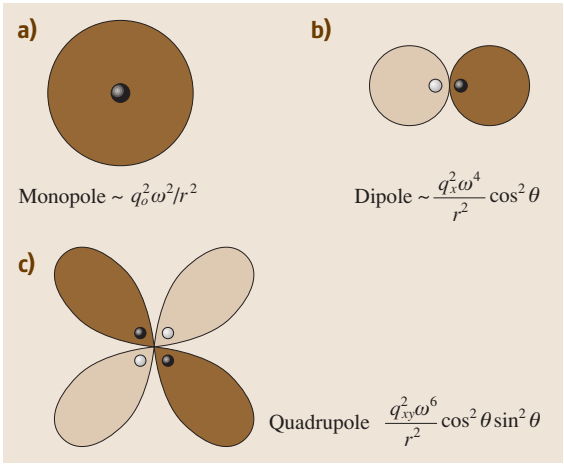


Fig. 15.2 Typical radiation patterns and intensities for monopole, dipole and quadrupole sources. The *two colours* represent monopole sources and sound pressures of opposite signs

In contrast, in the *far field* ($kr \gg 1$), the pressure is in phase with the velocity, so that work is done on the surrounding gas. This accounts for the fact that sound radiation varies in intensity as $1/r^2$.

The transition from the near- to far-field regions occurs when $r \sim \lambda/2\pi$, where λ is the acoustic wavelength of the radiated sound. At 340 Hz, this corresponds to a distance of only ≈15 cm. The difference in the frequency dependencies of the near- and far-field sound means that a violinist or piccolo player, with their ears relatively close to the instrument, experiences a rather different sound from that heard by the listener in the far field. However, for most musical instruments, the distance between the source of radiated sound and the player's head is already at least $\lambda/2$, so that even the player is in the *far field* ($kr > 1$), at least for the high-frequency partials of a musical tone.

Directionality and Multipole Sources

At very low frequencies, the acoustic wavelength λ is often considerably larger than the physical size of the radiating source (e.g. the open ends of woodwind and brass instrument bores and the body of most stringed instruments), which can then be considered as a point source radiating isotropically into space. However, as soon as the wavelength becomes comparable with the size of the radiating source, the radiated sound will acquire directional properties determined by the geometry of the instrument and the vibrational characteristics of the excited modes. The directional properties can then

Amazon
EX1013 Page 547

be described by treating the instrument as a superposition of monopole, dipole, quadrupole and higher-order multipole acoustic sources, with the directional radiating properties shown schematically in Fig. 15.2.

A monopole source can be considered as a pulsating sphere of radius a with surface velocity $v\,e^{i\omega t}$ resulting in a pulsating volume source $4\pi a^2 v\,e^{i\omega t} = Q\,e^{i\omega t}$. Equations (15.10) and (15.11) describe the sound field generated by such a source. Equating the velocities on the surface of the sphere to that of the induced air motion gives, at low frequencies such that $ka \ll 1$,

$$p(r,t) = \frac{i\omega\rho}{4\pi r} Q\, e^{i(\omega t - kr)} \,. \tag{15.12}$$

The radiated power P is then given by $\frac{1}{2}p^2/\rho c_0$ integrated over the surface of a sphere at radius r, so that

$$P(ka \ll 1) = \frac{\omega^2 \rho Q^2}{8\pi c_0} \,. \tag{15.13}$$

In the high-frequency limit ($ka \gg 1$), when the acoustic wavelength is much less than the size of the sphere,

$$P(ka \gg 1) = \frac{\rho c_0}{8\pi a^2} Q^2 = 4\pi a^2 \frac{1}{2} z_0 v^2 \,. \tag{15.14}$$

Equation (15.12) is a special case of the general result that, at sufficiently high frequencies such that the size of the radiating object $\gg \lambda$, the radiated sound is simply $\frac{1}{2}z_0 v^2$ per unit area, though the sound at a distance also has to take into account the phase differences from different parts of the vibrating surface. Note that

$$\frac{P(ka \ll 1)}{P(ka \gg 1)} = (ka)^2 \,. \tag{15.15}$$

The radiated sound intensity from a monopole source therefore initially increases with the square of the frequency but becomes independent of frequency above the crossover frequency when $ka > 1$. Hence members of the violin family and guitar families are rather poor acoustic radiators for the fundamental component of notes played on their lowest strings, as are wind and brass instruments, which radiate sound from the relatively small open ends and side holes. However, it is only because of such low radiation efficiencies, that strong resonances can be excited in the air columns of brass and woodwind instruments.

A dipole source can be formed by displacing two oppositely signed monopoles $\pm Q$ a short distance along the x-, y- or z-directions. For a dipole aligned along the x-axis of strength $q_x = Q\Delta x$. The sound pressure is simply the difference in pressure from monopoles of opposite sign a distance Δx apart, so that in the far field ($kr \gg 1$)

$$p(\theta)_{\text{dipole}} = p(\theta)_{\text{monopole}} \times (ik\Delta x)\cos\theta \,. \tag{15.16}$$

A polar plot of the sound pressure from a dipole is illustrated schematically in Fig. 15.2, with intensity and radiated power now proportional to ω^4 and q_x^2. In general, any radiating three-dimensional object will involve three dipole components (p_x, p_y and p_z), with radiation lobes along the three directions.

A quadrupole source is generated by two oppositely signed dipole sources displaced a small distance along the x- ,y-, or z-directions (e.g. of the general form $q_{xy} = Q\Delta x\Delta y$). The pressure is now given by the differential of the dipole radiation in the newly displaced direction, so that, for example, the pressure from a quadrupole source q_{xy} in the xy-plane is given by

$$p_{\text{dipole}} = p_{\text{monopole}} \times \left(-k^2 \Delta x \Delta y\right) \cos\theta \sin\theta \,, \tag{15.17}$$

as illustrated in Fig. 15.2. Note that each time the order of the multipole source increases, the radiated pressure depends on one higher power of frequency, while the intensity increases by two powers of the frequency. The radiated power from multipole sources therefore decreases dramatically at low frequencies relative to that of a monopole source. At low frequencies, radiation from most musical instruments is dominated by monopole components.

In general, six quadrupole sources ($q_{xx}, \ldots, q_{yz}$) would be required to describe radiation from a three-dimensional source. However, because the acoustic power radiated by a quadrupole source at low frequencies is proportional to ω^6, one need often only consider the monopole and three dipole components to describe the low-frequency radiation pattern of instruments like the violin and guitar family, as described in a recent study of the low-frequency radiativity of a number of quality guitars by *Hill* et al. [15.10]. However, at high frequencies, when λ is comparable with or less than the size of an instrument, the above simplifications break down. The directionality of the radiated sound then has to be computed from the known velocities over the whole surface, taking into account phase differences and baffling effects from the body of the instrument.

Radiation from Surfaces

Many musical instruments produce sound from the vibrations of two-dimensional surfaces – like the plates of a violin or the stretched membrane of a drum. Imagine first a standing wave set up in the two-dimensional

Amazon
EX1013 Page 548

xy-plane with displacements in the z-direction varying as $\sin(k_x x)e^{i\omega t}$. We look for propagating sound waves solutions radiating from the surface of the form $\sin(k_x x)e^{i(\omega t - k_z z)}$, which must satisfy the wave equation and hence the relationship,

$$k_z^2 = \frac{\omega^2}{c_0^2} - k_x^2 = \omega^2\left(\frac{1}{c_0^2} - \frac{1}{c_m^2}\right), \qquad (15.18)$$

where c_m is the phase velocity of transverse waves on the membrane or plate in the xy-plane. Sound will therefore only propagate away from the surface ($k_z^2 > 0$) when $c_m > c_0$. If the sound velocity is greater than the phase velocity in the plate or membrane, energy will flow from regions of positive to negative vertical displacements and vice versa, with an exponentially decaying sound field, varying as $e^{-z/\delta}$ where $\delta = |k_z|^{-1}$.

Typical dispersionless wave velocities for the stretched drum heads of timpani are around 100 m/s (*Fletcher* and *Rossing* [15.5], Sect. 18.1.2), so that they are not very efficient radiators of sound. This is particular relevant for asymmetrical modes, when sound energy can flow from the regions of positive to negative displacement and vice versa. However, for even modes, the cancellation between adjacent regions moving out of phase with each other can never be complete, so that such modes will radiate more effectively.

A particularly interesting case occurs for stringed instruments, where the phase velocity of the transverse vibrations of the thin front and back plates increases with frequency as $\omega^{1/2}$ (Sect. 15.2.6). Hence, below a critical crossover or coincidence frequency, when the phase velocity in the plates is less than the speed of sound in air, standing waves on the vibrating plates are relatively inefficient radiators of sound, while above the crossover frequency the plates radiate sound rather efficiently. *Cremer* [15.11] estimates the critical frequency for a 4 mm-thick cello plate as 2.8 kHz; for a 2.5 mm violin plate the equivalent frequency would be ≈ 2 kHz.

Radiation from Wind Instruments

The holes at the ends or in the side walls of wind instruments can be considered as piston-like radiation sources. At high frequencies, such that $ka \gg 1$, where a is their radius, the holes will be very efficient radiators radiating acoustic energy $\sim 1/2 z_0 v^2$ per unit hole area. However, over most of the playing range $ka \ll 1$, so that the radiation efficiency drops off as $(ka)^2$, just like the spherical monopole source. Most of the sound impinging on the end of the instrument is therefore reflected, so that strong acoustic resonances can be excited, as discussed in the later section on woodwind and brass instruments.

15.1.3 The Anatomy of Musical Sounds

The singing voice, bowed string, and blown wind instruments produce continuous sounds with repetitive waveforms giving musical notes with a well-defined sense of musical pitch. In contrast, many percussion instruments produce sounds with non-repetitive waveforms composed of a large number of unrelated frequencies with no definite sense of pitch, such as the side drum, cymbal or rattle. There are also other stringed instruments and percussion instruments, such as the guitar, piano, harp, xylophone, bells and gongs, which produce relatively long sounds, where the slowly decaying vibrations produce a definite sense of pitch.

In all such cases, the complexity of the waveforms of real musical instruments distinguishes their sound from the highly predictable sounds of simple electronic synthesisers. This is why the sounds of computer-generated synthesised instruments lack realism and are musically unsatisfying. In this section, we introduce the way that sound waveforms are analysed and described.

Sinusoidal Waves

The most important, but musically least interesting, waveform is the pure sinusoid. This can be expressed in several alternative forms,

$$a\cos(2\pi ft + \phi) = a\cos(\omega t + \phi) = \mathrm{Re}(\boldsymbol{a}\, e^{i\omega t}), \qquad (15.19)$$

where $\boldsymbol{a}$ is in general complex to account for phase, f is the frequency measured in Hz and equal to the inverse of the period T, $\omega = 2\pi f$ is the angular frequency measured in radians per second, t is time, and ϕ is the phase, which depends on the origin taken for time.

Any sound, however complex, can be described in terms of a superposition of sinusoidal waveforms with a spectrum of frequencies. Figure 15.3 contrasts the envelopes, waveforms and spectra of a synthesised sawtooth waveform and the much more complex and musically interesting waveform of a note played on the oboe (CD-ROM provides an audio comparison). Note the much more complex fluctuating envelope and less predictable amplitudes of the frequency components in the spectrum of the oboe.

As we will show later, in defining the sound and quality of any musical instrument, the shape and fluctuations in amplitude of the overall envelope are just as important as the waveform and spectrum.

Amazon
EX1013 Page 549

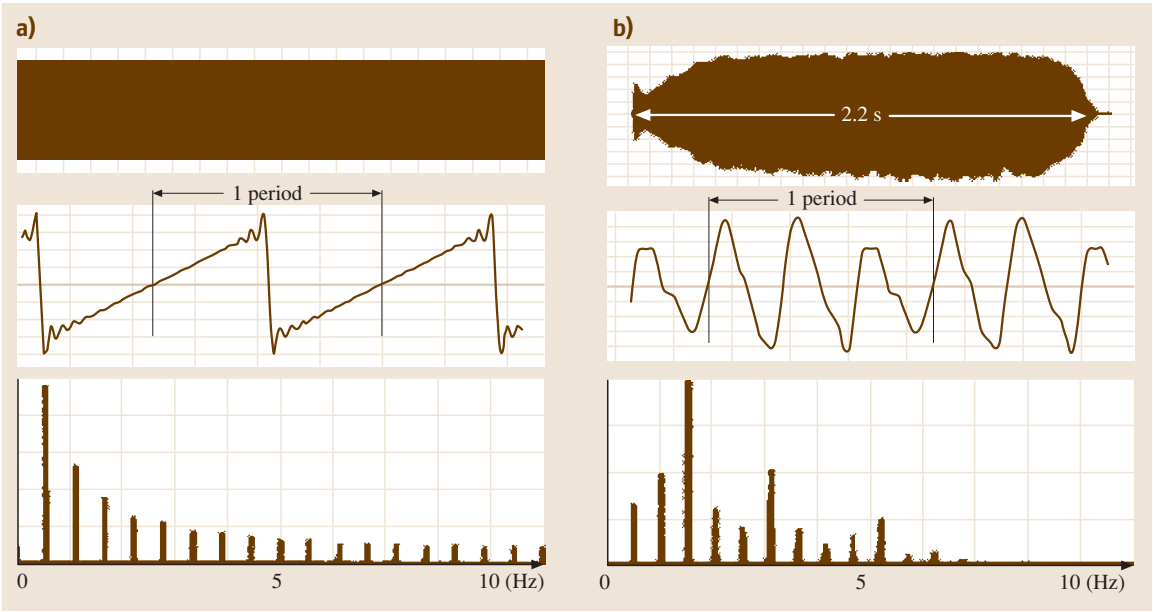


Fig. 15.3a,b Comparison of the envelope, repetitive waveform and spectrum of **(a)** a synthesised sawtooth and **(b)** a note played on an oboe

Range of Hearing and Musical Instruments

A young adult can usually hear musical sounds from around 20 Hz to 16 kHz, with the high-frequency response decreasing somewhat with age (typically down to between 10–12 kHz for 60-year olds). Audio CD-ROM provides a sequence of 1 s-long sine waves starting from 25 Hz to 12.8 kHz, doubling in frequency each time. Doubling the frequency of a sinusoidal wave is equivalent in musical terms of increasing the pitch of the note by an octave. Audio CD-ROM is a similar sequence of pure sine waves from 8 kHz to 18 kHz in 2 kHz steps. Any loss of sound at the low frequencies in CD-ROM will almost certainly be due to the limitations of the reproduction system used, which is particularly poor below ≈ 200 Hz on most PC laptops and notebooks, while the decrease in intensity at high frequencies in CD-ROM simply reflects the loss of high-frequency sensitivity of the ear (see Fig. 15.16 and Chap. 13 for more details on the amplitude and frequency response of the human ear).

The above sounds should be compared with the much smaller range of notes on a concert grand piano, typically from the lowest note A0 at 27.5 Hz to the highest note C9 at 4.18 kHz, as illustrated in Fig. 15.4. The nomenclature for musical notes is based on octave sequences of C-major scales with, for example, the note C1 followed by the white keys D1, E1, F1, G1, A1, B1, C2, D2, ... Alternatively, the octave is indicated by a subscript (e.g. A_4 is concert A). Where the white notes are a tone apart, a black key is inserted to play the semitone between the adjacent white keys. This is indicated by the symbol # from the note below or ♭ from the note above. Figure 15.4 also illustrates the playing range of many of the instruments to be considered in this chapter.

Frequency and Pitch

It is important to distinguish between the terms frequency and pitch. The frequency of a waveform is strictly only defined in terms of a continuous sinusoidal waveform. In contrast, the waveforms of real musical instruments are in general complex, as illustrated by the oboe waveform in Fig. 15.3. However, despite such complexity, the repetition frequency and period T can still be defined provided the waveform does not vary too rapidly with time. The periodicity of a note (measured in Hz) can then be defined as the inverse of T. This is generally the note that the player reads from the written music. However, as described later, a repetitive waveform does not necessarily include a sinusoidal component at the repetition frequency, an effect referred to as the *missing fundamental*. Furthermore, depending on the strength of the various sinusoidal components present, there can often be an ambiguity in the pitch of a note perceived by the listener. The subjective pitch, when matched against

Amazon
EX1013 Page 550

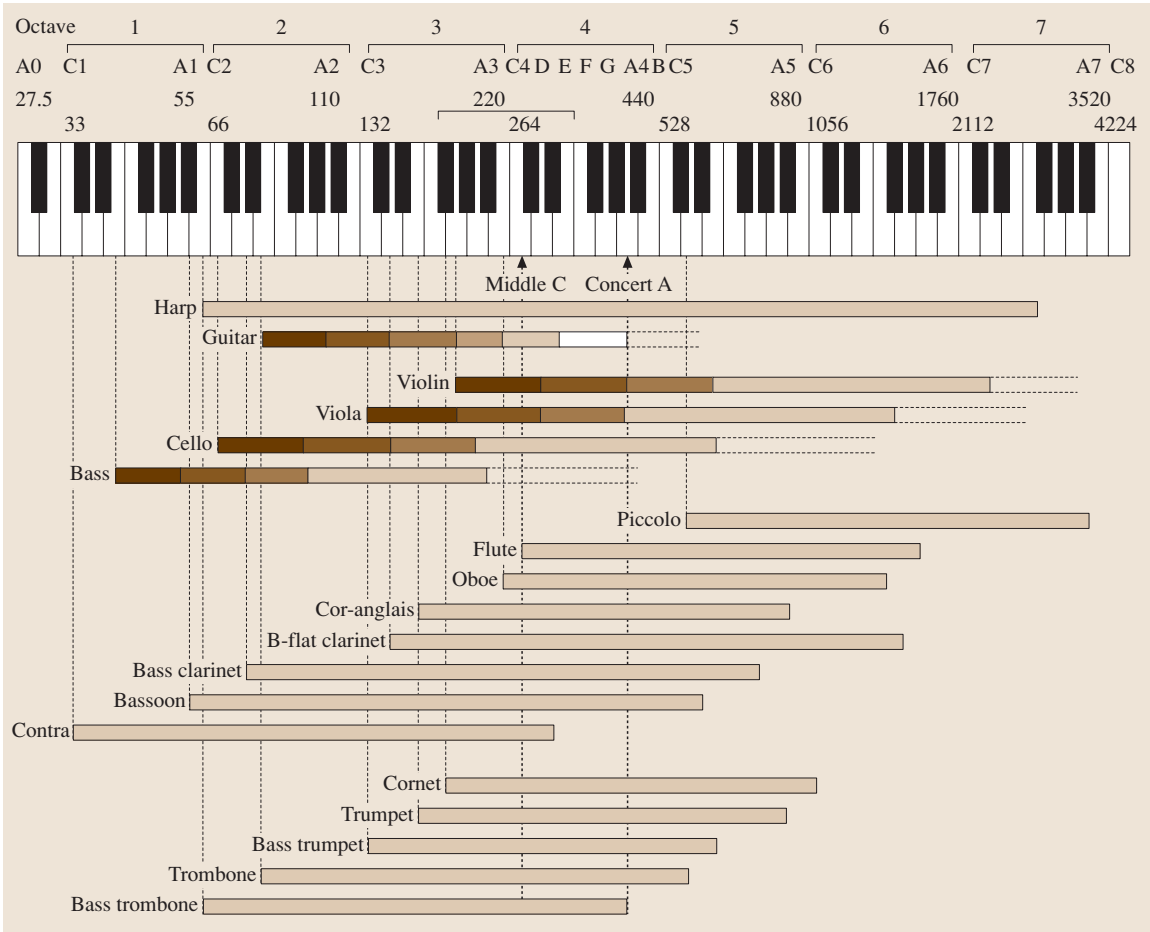


Fig. 15.4 Notation used for notes of the musical scale and the playing range of classical western musical instruments. Subdivisions for stringed instruments represent the tuning of the open strings

a pure sinusoidal wave, can often be an octave below or above the repetition frequency. The subjective pitch, as its name implies, can differ from person to person and within the musical context of the note being played.

Musical Intervals and Tuning

In western music the octave interval is divided into six tones (a whole-tone scale) and 12 semitones (the chromatic scale). Today, an equal temperament, logarithmic scale is used to tune a piano, with a fractional increase in frequency of $2^{1/12} = 1.059$ ($\approx 6\%$) between any two notes a semitone apart. The fractional increase between the frequencies of a given musical interval (a given number of semitones) is then always the same, whatever the starting note. Twelve successive semitones played in sequence therefore raises the frequency by an octave [$(2^{1/12})^{12} = 2$]. Any music played on the piano keyboard can therefore be transposed up or down by a given number of semitones, changing the pitch but leaving the relationship between the musical intervals unchanged. Such a scale was advocated as early as 1581, in a treatise by the lutenist Vincenzo Galileo (the father of Galileo Galilei). Although it is sometimes claimed that Bach exploited such a tuning in his 48 Preludes and Fugues, which uses all possible major and minor keys of the diatonic scale, historical research now suggests that Bach used a form of mean-tone tuning, which preserved some of the characteristic qualities of music written in particular keys [15.1].

Amazon
EX1013 Page 551

To provide a greater discrimination in the measurements of frequency, the semitone is divided into 100 further logarithmic increments called cents. The octave is therefore equivalent to 1200 cents and a quartertone to ≈ 50 cents, with the exact relationship between frequencies given by

$$\text{interval} = \frac{1200}{\ln 2} \ln(f_2/f_1) \text{ cents} \qquad (15.20)$$

corresponding to $\sim 1.73 \times 10^3 (\Delta f/f)$ cents for small fractional changes $\Delta f/f$.

Early musical scales were based on various variants of the natural harmonic series of frequencies, $f_n = n f_1$, where n is an integer (e.g. 200, 400, 600, ... 1600 Hz), illustrated by the audio CD-ROM. These notes correspond to the *harmonics* produced when lightly touching a bowed string at integral subdivisions of its length (1/2, 1/3, 1/4, etc.) CD-ROM. These simple divisions give successive musical intervals of the octave, perfect fifth, perfect fourth, major third and minor third, with frequency ratios 2/1, 3/2, 4/3, 5/4 and 6/5, respectively. The seventh member of the harmonic sequence has no counterpart in traditional western classical music, although it is sometimes used by modern composers for special effect [15.12].

Just temperament corresponds to musical scales based on these integer fraction intervals. The Pythagorean scale is based on the particularly consonant intervals of the octave (2/1) and perfect fifth (3/2), which can be used to generate individual intervals of the form $3^p/2^q$ or $2^p/3^q$, where p and q are positive integers. Although the Pythagorean and just-tempered scales coincide for the octave, perfect fifth and fourth, there are musically significant differences in the tuning for all other defined intervals, and all intervals other than the octave differ slightly from those of the equally tempered scale. A comparison between the musical intervals of just and equal temperament tuning is shown in Table 15.1, with the fractional mistuning indicated in cents. Because of the differences in tunings of the musical intervals, music transposed from one key to another will generally sound badly out of tune (particularly for commonly used intervals like the major and minor third) – unlike those played on a modern equal-tempered keyboard. Prior to the now almost universal practice of tuning keyboard instruments to a equal-tempered scale, many tuning schemes were devised which partially overcame the problems of tuning when playing in a succession of different keys (see *Fletcher* and *Rossing* [15.5], Sect. 17.6, and *Barbour* [15.13] for further discussion). Singers, stringed and wind instrument players can adjust the pitch of the notes they produce to optimise the tuning with other performers and for musical effect.

Table 15.1 Principal intervals and differences between just- and equal-temperament intervals

Interval	Just	Equal	$\Delta f/f$ Just-equal cents
Octave	2/1	2.00	0
Perfect fifth	3/2	$2^{7/12} = 1.498$	+2
Perfect fourth	4/3	$2^{5/12} = 1.334$	−2
Major third	5/4	$2^{4/12} = 1.260$	−13
Minor third	6/5	$2^{3/12} = 1.189$	+15
Tone	9/8	$2^{2/12} = 1.122$	−4
Semitone	16/15	$2^{1/12} = 1.066$	+1

Figure 15.5 and audio CD-ROM illustrate the difference in the sounds of a major triad formed from the just intervals (1, 5/4, 3/2) and the equivalent equal-tempered scale intervals (1 : 1.260 : 1.498). The rational Pythagorean intervals give a repetitive waveform of constant amplitude, while the less-consonant, inharmonic, equal-tempered intervals have a non-repetitive waveform with an easily discernable periodic beat in amplitude resulting from the departures in harmonicity of its component frequencies, as illustrated in Fig. 15.5. Interestingly, the pitch of the equally tempered intervals also sounds slightly higher, though both share the same fundamental.

A sequence of upward fifths (frequency ratio 3/2) and downward octaves (ratio 1/2) can be used to fill in all the semitones of an octave scale on the piano keyboard. However, the resulting octave formed from a succession of 12 upward fifths and six downward



Fig. 15.5 Wave envelope of a major triad chord based on the Pythagorean scale followed by the same chord on the equal-tempered scale, with pronounced beats in the amplitude arising from the departures from harmonicity in the frequencies of the major third and perfect fifth

Amazon
EX1013 Page 552

octaves gives a frequency ratio of $(3/2)^{12}/2^6 = 2.027$, which is significantly *sharper* (higher in frequency) than a true octave. In practice, a skilled piano tuner listens to the beats produced when playing the above intervals and tunes the upward fifth slightly flat, so that the sequence returns to the exact octave. However, there are striking psychoacoustic effects, in addition to physical shifts in the frequencies of upper partials arising from the finite rigidity of the strings, which result in pianos being tuned on a slightly stretched tuning scale with the octaves purposely tuned sharp at higher frequencies and flat at lower frequencies (*Fletcher* and *Rossing* [15.5], Sect. 12.8).

Repetitive Waveforms

Before considering the more complex waveforms of musical instruments, we consider the simple square, triangular, sawtooth and triangular repetitive waveforms (audio CD-ROM) and the corresponding Fourier spectra illustrated in Fig. 15.6.

Fourier Theorem

Fourier showed that any repetitive waveform, $f(t+nT) = f(t)$, can be described as a linear combination of sinusoidal components with frequencies that are exact multiples of the inverse repetition period T or physical pitch of the wave. This is formally expressed by the Fourier theorem,

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{\mathrm{i}n\omega_1 t} , \tag{15.21}$$

where $\omega_1 = 2\pi/T$ and

$$c_n = \frac{1}{T}\int_0^T f(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t , \tag{15.22}$$

where n takes on all positive and negative integer values. The Fourier coefficients c_n can be evaluated by performing the integral over any single period of the waveform. In general, the Fourier coefficients will have both real and imaginary components describing both the amplitude and phase.

For simple waveforms, such as the square, sawtooth and triangular waveforms, the origin of time can be chosen to make the waveforms symmetric or antisymmetric in time. The Fourier expansion can then be expressed in terms of the even cosine or odd sine functions,

$$f(t) = \sum_{n=0}^{\infty} \left\{ \begin{array}{l} a_n \sin(n\omega_1 t) \\ \text{or} \\ b_n \cos(n\omega_1 t) \end{array} \right\} , \tag{15.23}$$

with corresponding coefficients given by

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = \frac{2}{T}\int_0^T f(t) \begin{pmatrix} \sin(n\omega_1 t) \\ \cos(n\omega_1 t) \end{pmatrix} \mathrm{d}t , \tag{15.24}$$

where n is now restricted to positive integer values. The factor 2 is replaced by unity for the zero frequency average component b_0.

The first few terms of the square, sawtooth and triangular waveforms are listed in Table 15.2. The origin of time has been selected to make the waves odd-functions of time, as illustrated in Fig. 15.6, with the Fourier series only including sine terms. The Fourier components at integral multiples of the fundamental repetition frequency are referred to as partials, harmonics, or overtones. The nth partial has a frequency $f_n = n f_1$. This differs from the terminology used by musicians, who refer to f_2 as the first harmonic or overtone. Interestingly, a waveform depends critically on the sign (phase) of the individual Fourier components. In contrast, the ear is largely insensitive to the phase

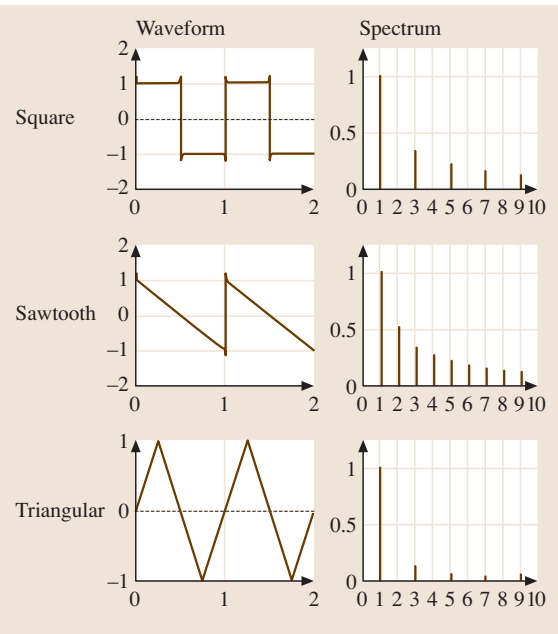


Fig. 15.6 Typical repetitive waveforms (synthesised from the first 100 components of a Fourier series) and the amplitudes of the first few partials normalised to the amplitude of the fundamental

Amazon
EX1013 Page 553

of the individual partials, with little change in the perceived sound when the sign or phase of a component partial is changed, though the waveforms will be very different.

For an arbitrarily chosen origin of time, the Fourier expansion will include both sine and cosine terms. The energy or intensity is proportional to the resultant amplitudes squared, $a_n^2 + b_n^2$, which is independent of the origin of time. The phase ϕ_n is given by $\tan^{-1}(b_n/a_n)$. The spectrum of a waveform is often plotted in terms of the modulus of the amplitude as a function of frequency, without reference to phase, as in Fig. 15.6. However, measurements of both amplitude and phase are important in any detailed comparison with theoretical models and in analytic measurements, such as modal analysis.

The square and sawtooth waveforms are closely related to the waveforms excited on bowed and plucked strings and loudly played notes on wind and brass instruments. The discontinuities in waveform generate a very rich harmonic spectrum with Fourier components or partials that decrease relatively slowly (as $1/n$) with increasing n. The strong higher partials give a much harsher and more penetrating sound than simple sinusoids, which is why the oboe, which has a sound that is very rich in higher partials, is used to sound concert A when an orchestra tunes up. In contrast, the partials of the triangular wave, with discontinuities in slope instead of amplitude, decrease more rapidly as $1/n^2$, with a resultant sound little different from that of a simple sinusoidal wave.

Note the large difference between the sound of a sawtooth waveform, which is closely related to the sound of an oboe in having a complete set of harmonic components, and the "hollow" sound of the square waveform, which is more like the sound of the lowest notes on a clarinet, with rather weak even-integer harmonics or overtones on its lowest notes (Fig. 15.7).

Table 15.2 Fourier expansions of the square, sawtooth and triangular waveforms

Square	$\frac{4}{\pi}\left(\sin\omega_1 t + \frac{1}{3}\sin 3\omega_1 t + \frac{1}{5}\sin 5\omega_1 t \ldots\right)$
Sawtooth	$\frac{2}{\pi}\left(\sin\omega_1 t - \frac{1}{2}\sin 2\omega_1 t + \frac{1}{3}\sin 3\omega_1 t - \frac{1}{4}\sin 4\omega_1 t \ldots\right)$
Triangular	$\frac{2}{\pi}\left(\sin\omega_1 t - \frac{1}{3^2}\sin 3\omega_1 t + \frac{1}{5^2}\sin 5\omega_1 t \ldots\right)$

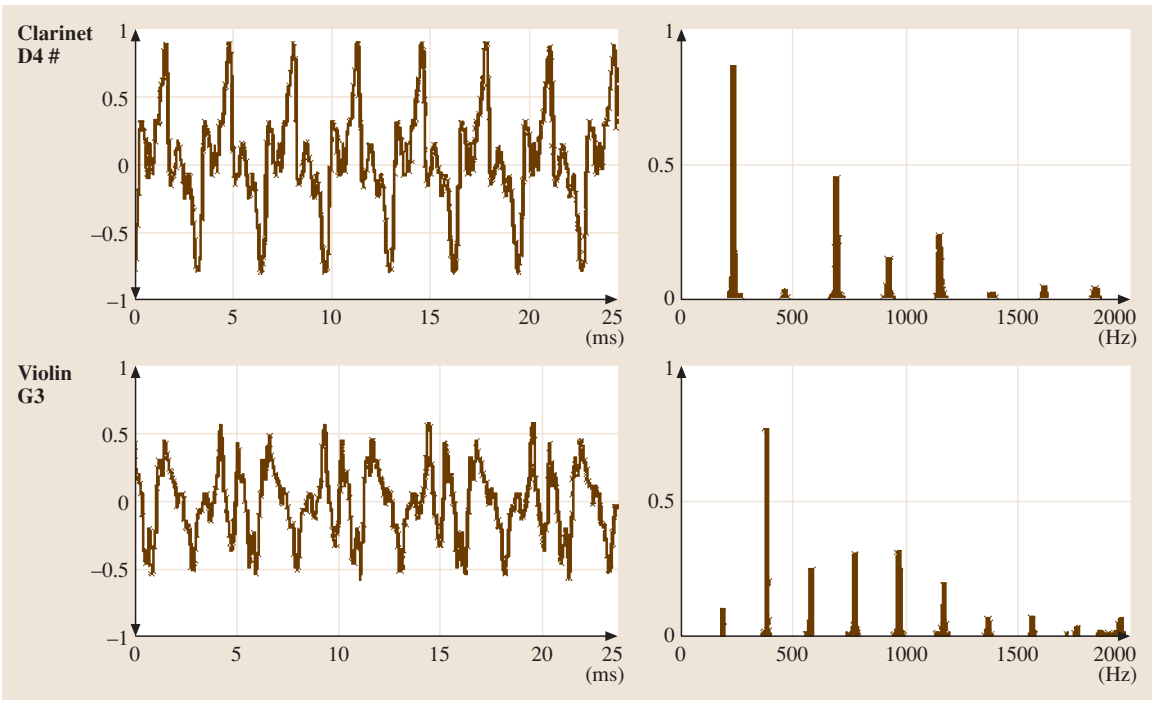


Fig. 15.7 Short-period samples of clarinet (D#4) and bowed violin (G3) tones and the corresponding Fourier spectra. The *vertical scales* are linear

Amazon
EX1013 Page 554

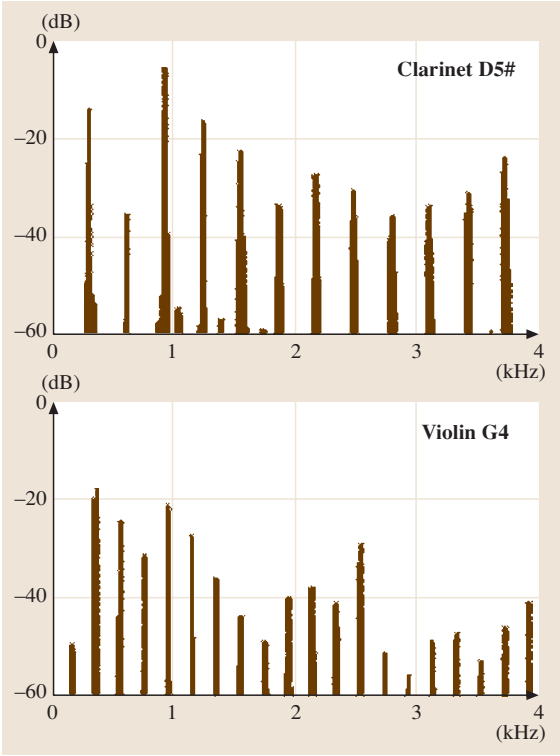


Fig. 15.8 Typical spectra for a clarinet and violin note plotted on a dB scale illustrating the large number of harmonics or overtones excited by bowed and blown instruments

Musical Waveforms and Spectra

The waveforms produced by musical instruments are generally far more complicated than the above simple examples, as already illustrated for an oboe note in Fig. 15.3. Waveforms and associated spectra of the clarinet note D#4 (309 Hz) and the violin bowed open-G-string G3 (195 Hz) are illustrated in Fig. 15.7. These are simply representative waveforms. Unlike the impression given in some elementary textbooks, there is no such thing as a defining violin or clarinet waveform or spectrum. Both the waveforms and the spectra change significantly from one note to the next – and even within a note when played with vibrato, particularly on stringed instruments. Despite the complexity of the waveforms, any repetitive waveform can be described as a linear superposition of sine waves, with frequencies that are integer multiples of the fundamental, as illustrated by the spectra.

Plotting the amplitudes of the Fourier coefficients on a linear scale often highlights the physical processes involved in the production of the sound. For example, the relatively small amplitudes of the second harmonic or partial in the sound of the clarinet reflects the absence of even-n modes of a cylindrical tube closed at one end, which approximates to that of the clarinet. Similarly, the small amplitude of the first partial in the sound of a violin reflects the absence of efficient radiating modes at low frequencies. However, because of the very wide dynamic range of hearing (a factor of $\approx 10^{10}$–10^{12} in intensity), it is often more appropriate to plot the Fourier coefficients in decibels on a logarithmic scale. Figure 15.8 shows the spectrum for clarinet and violin notes re-plotted on a dB scale, which illustrates the strength of all the partials over a very wide dynamical range. For bowed instruments such as the cello, well over 40 harmonic partials can be identified below 8 kHz. The sound of an instrument will be determined by all such components and not simply by the fundamental, which may make a relatively small contribution to the perceived sound. This is illustrated for a scale played on the violin with each note first played as recorded and then with the fundamental component removed by a digital filter (audio |?› CD-ROM). The lowest notes of the scale, for which the fundamental component is already very weak, are little affected by the removal of the fundamental component; however, the sound gets progressively "thinner" in the second half of the scale for notes for which the fundamental partial makes an increasingly significant contribution to the "richness" or "warmth" of the sound.

Transient and Non-Repetitive Tones

No musical note lasts for ever, so that musical sounds are all, to some extent, transient. Moreover, the sound of many percussion instruments is composed of many strongly inharmonic partials, with no regime in which the waveforms can be considered even quasi-repetitive. Nevertheless, one can still use the Fourier theorem to extract the spectrum of such a note, by considering each transient signal as one of a sequence of such transients repeated, say, every second, minute or even year. The spectrum of such a repeated waveform will therefore involve frequency components at integer multiples of the inverse repetition period, which we can make as long as we choose. In the limit of infinitely long repetition times, the Fourier series of a non-repeating waveform can therefore be replaced by a continuous spectral distribution over all possible frequencies, known as the Fourier transform $F(\omega)$,

$$f(t) = \int_{-\infty}^{\infty} F(\omega) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t \,, \qquad (15.25)$$

Amazon
EX1013 Page 555

where

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, dt \, . \qquad (15.26)$$

The Fourier transform spectra of important non-repetitive waveforms are shown in Fig. 15.9. In all cases, the width Δf of the Fourier spectrum is inversely proportional to the length τ of the input waveform, with $\Delta f \tau \sim 1$. For a rectangular pulse, the spectrum extends over a rather wide frequency range with the first zero when $f\tau = 1$, but with many ripples of decreasing amplitude extending to higher frequencies. For an impulse of negligible width (the delta-function), the spectrum is flat out to very high frequencies. The spectrum of a Gaussian waveform varying as $\exp[-(t/\tau)^2]$ is also a Gaussian proportional to $\exp[-(\pi f \tau)^2]$, with a width $\Delta f = 1/\pi\tau$. Similarly, the spectrum of a sinusoidal waveform with a Gaussian envelope of width τ is broadened by $\Delta f = 1/\pi\tau$.

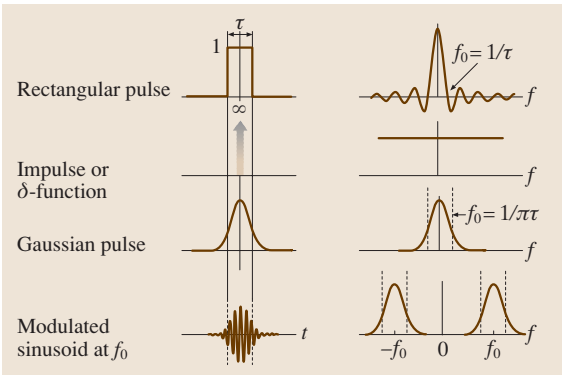


Fig. 15.9 Fourier transforms of transient waveforms

Any waveform that involves variations on a time scale τ will have Fourier components extending out to frequencies $\sim 1/\tau$. To reproduce such waveforms faithfully, the bandwidth of any recording or reproduction system must therefore extend to frequencies of at least $1/\tau$. Examples of non-repetitive waveforms and their associated spectra are illustrated in Fig. 15.10 for an orchestral rattle, a cymbal crash and timpani.

The ratchet sound consists of a sequence of short *clicks* illustrated by the selected short-section waveform. The spectrum is very broad with no individual frequencies particularly dominant. The crash of a cymbal generates a very large number of very closely spaced resonances, which appear as a fairly random set of peaks giving an overall broadband spectrum. The timpani spectrum shows a small number of large peaks corresponding to the prominent modes of vibration of the drum head, superimposed on a very wide-band spectrum largely

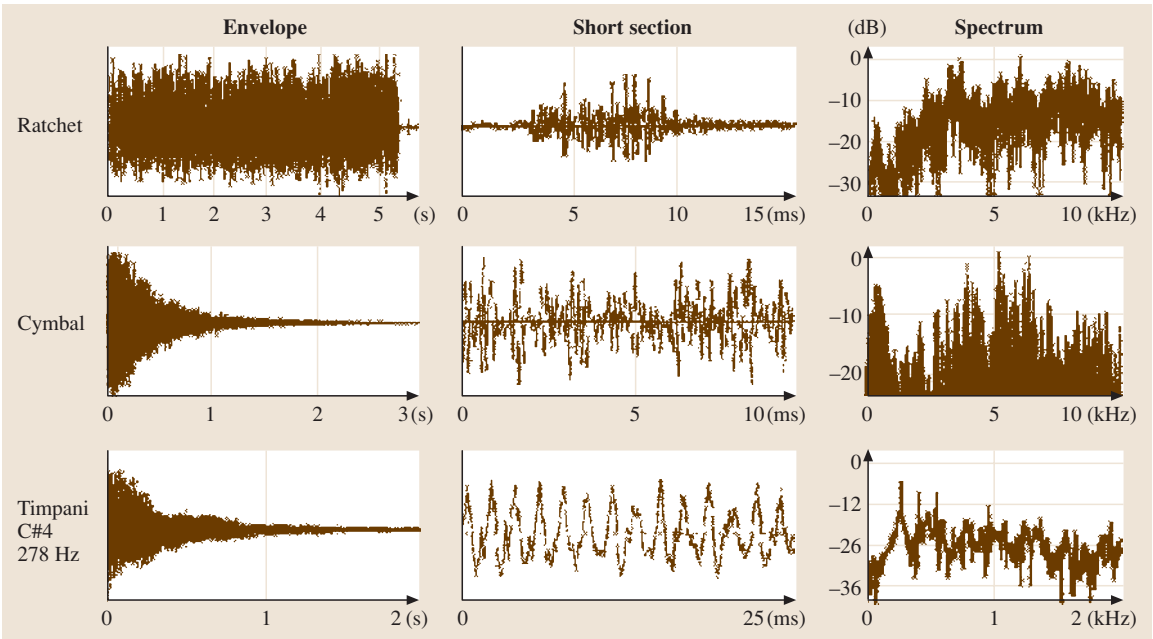


Fig. 15.10 Envelope, typical short-period waveform and spectrum for the sounds of a ratchet, cymbal and timpani

Amazon
EX1013 Page 556

Table 15.3 Dynamic range of an analogue-to-digital converter (ADC)

N-bit ADC	Dynamic range $\pm 2^{n-1}$	Dynamic range (dB)
8 bit	128	42
16 bit	32.8 k	90
24 bit	8.39 M	138

associated with the initial transients involving sound from all parts of drum.

Digital Recording

Nowadays almost all sound is recorded digitally using an analogue-to-digital converter (ADC). This converts the continuously varying analogue input signal into a stream of numbers, which can be recorded digitally on a computer or compact disc. Audio signals are typically recorded at a sampling rate of 44.1 kHz with 16-bit resolution corresponding to 1 part in 2^{16}. This allows signals to be recorded in 65 536 equally divided levels between the maximum positive and negative input signals (i. e. between ± 32.8 k levels).

For the highest-quality digital recordings, even faster recording rates with higher-bit resolution are used (typically 24-bit sampling at 96 kHz). This allows for over-sampling of the recorded signal, so that signals can be averaged, any errors detected and eliminated, and filtered more easily. As already noted, the dynamic range of human hearing can be as large as 100 dB. To exploit such a large range and to capture the details of both loud and soft sounds from a large orchestra accurately requires the recording system to have a large dynamic range. Table 15.3 shows the dynamic range in terms of the number of bits used to record the sound. Audio CD-ROM illustrates the greatly enhanced signal-to-noise ratio and hence increased dynamic range when sound is recorded at 16-bit rather than 8-bit resolution.

Aliasing

When sound is recorded digitally, ambiguities can arise when any of the input frequencies is larger than half the sampling rate f_D. This is known as the Nyquist limit $f_{Nyquist} = f_D/2$. For example, if a 2 kHz sine wave were to be sampled at 2 kHz, the digital signal would be recorded at exactly the same point of the waveform each cycle. The recorded digital signal would then be indistinguishable from a DC signal. It can easily be shown that sinusoidal inputs at $f_{Nyquist} + \Delta f$ give the same digital output as at $f_{Nyquist} - \Delta f$ and that the recorded signal is the same for all frequencies differing in frequency by

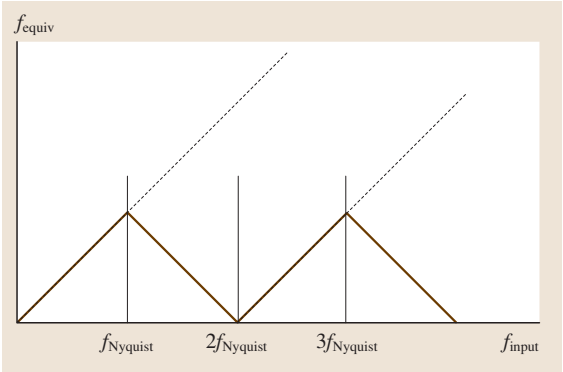


Fig. 15.11 Ambiguity of digital output for a steadily increasing frequency exceeding the Nyquist limit

the digitising frequency, $2f_{Nyquist}$. Thus for a steadily increasing input frequency the digital output is equivalent to that of a frequency which first increases up to $f_{Nyquist}$ then decreases to zero when $f_{in} = f_D = 2f_{Nyquist}$, with the process repeating for higher input frequencies, as illustrated in Fig. 15.11. In any replay system, an analogue output is generated that assumes a smooth curve between the sampled points. Hence, recorded frequencies above the Nyquist limit will be misinterpreted and will produce sounds below $f_{Nyquist}$ with no harmonic relevance to the original input.

This ambiguity is illustrated in audio CD-ROM, in which a sinusoidal input is swept in frequency from 200 Hz to 6 kHz. This is first recorded at 22.05 kHz, when aliasing is not a problem, and then at 6 kHz, when halfway through the increasing frequency signal, at 3 kHz, the replayed sound starts to descend to zero frequency at the end of the sweep, when the input signal has the same frequency as the sampling rate. The single-frequency sweep is then followed by an ascending major triad (with intervals in the ratios 1 : 5/4 : 3/2) recorded at 6 kHz, which illustrating the severe problems of aliasing in terms of musical harmonies, as soon as any of the higher-frequency components in a signal exceed the Nyquist limit, with the frequency of some partials ascending while others are descending.

To avoid such problems, a high-frequency cut-off input filter is generally used, with a cut-off frequency slightly below the Nyquist frequency (see Chap. 14 by *Hartmann* for further details).

Sound File Formats

A sound signal is frequently recorded and stored as an encoded WAVE file of the generic form `*.wav`, which includes additional information on data acqui-

Amazon
EX1013 Page 557

Table 15.4 Decoded WAVE file information

Data provided	Typical example
Mono (1), stereo (2)	2
Data acquisition rate	2.205×10^4
Resolution in bits	16
Data per second	4.41×10^4
First measurement from left channel	237
Simultaneous measurement from right channel	−1356
Second measurement from left channel	456
Simultaneous measurement from right channel	−1972
Repeated sequence until end of recorded sound	...

sition rate, stereo or mono format and bit resolution. The decoded structure of a WAVE file is shown in Table 15.4.

Recording audio signals as WAVE files is very expensive in memory, with 1 hour of stereo music recorded at 22 kHz requiring ≈ 300 MB. Music files on CDs are encoded so that the input is redistributed over time and therefore over the surface of the disc. This enables the original signal to be reproduced even in the presence of dust, scratches and other small imperfections on the disc surface, eliminating the clicks that were a familiar feature of older vinyl records. More sophisticated, adaptive, encoding schemes can be used to significantly reduce the amount of memory used, such as the now widely used mp3 format. An algorithm is used, based on physical principles and on the way the ear responds to musical sounds, to continuously analyse and process the incoming data. The input data can then be recorded using a much reduced number of bits, in much the same sort of way that digital pictures are encoded more efficiently in ZIP files and compact image formats. The information used to encode the digital signal is also recorded, so that the processed data can be unscrambled on playback with relatively little loss in perceived quality.

Discrete Fourier Transform

The digital form of the recorded data allows certain computational efficiencies in calculating the Fourier spectrum. Consider a recorded sample of N measurements, corresponding to a sample of length $T_s = N/f_D$. To calculate the spectrum, this data set is assumed to repeat indefinitely, to form a continuous waveform with a repetition frequency. From the Fourier theorem, the resulting spectrum is composed of Fourier components that are exact multiples of the repetition frequency, so that $f_n = n/T_s$. Hence, a 1 s set of data points will give a discrete Fourier spectrum with frequencies at 1, 2, ... n ... Hz. In practice, the number of Fourier components is limited to $N/2$, because each component has both an amplitude and a phase, which requires at least two independent measurements to be made per Fourier component.

Windowing Functions

Using the sampled waveform to form a continuously repeating waveform will, in general, introduce a repeating discontinuity Δ at the beginning and end of each repeated data set, since the start and end values will not usually be the same. Any such discontinuity will generate spurious contributions to the spectrum, with additional Fourier coefficients with amplitudes proportional to Δ/n. To circumvent this problem, a *windowing function* is used. This applies a smooth envelope to the data set, which reduces the values at the start and end to zero, thus eliminating the discontinuities. However, as described above, the application of such an envelope will give an extra width $\Delta f \sim 1/T_s$ to the spectral features.

A typical windowing function is the Hanning function $\sin^2(2\pi t/T_s)$. A number of other windowing functions are illustrated in Fig. 15.12, each of which has advantages for specific applications [15.14]. The Hanning windowing is widely used for general-purpose measurements, while the Hamming function is used to separate closely spaced sine waves. In general there is a trade-off between the accuracy that can be achieved in determining the frequency of individual spectral components and the width of the low-amplitude side lobes generated by application of a windowing function. Various forms of the Blackman–Harris windowing function

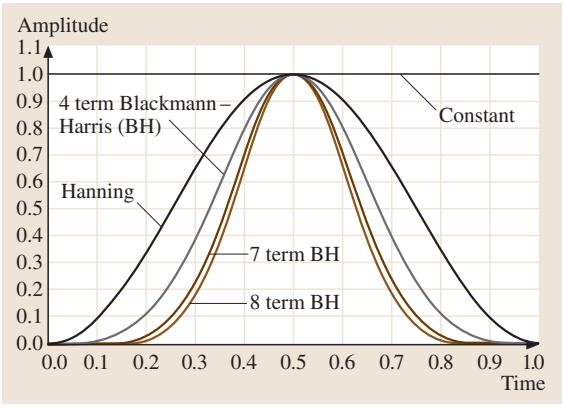


Fig. 15.12 Representative windowing functions (after [15.14])

Amazon
EX1013 Page 558

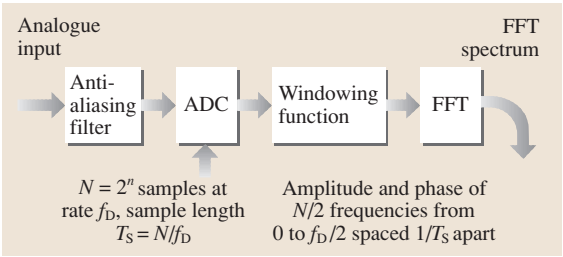


Fig. 15.13 A typical digital sampling and FFT analysis scheme

can be used to optimise the fast Fourier transform (FFT) for specific measurements. Windowing need not be used for the accurate measurements of widely separated sinusoidal waves with similar amplitudes, though one should be aware of the existence of the rather wide side lobes generated unless the sampling period is an exact integer multiple of the period of the waveform being measured.

Fast Fourier Transform (FFT)

To determining the amplitude and phase of the $N/2$ Fourier components from N input measurements requires the inversion of an $N \times N$ matrix, requiring a computation time proportional to N^2. However, if N is an integral power of 2 (e.g. $2^8 = 256$, $2^{16} = 65536$), the FFT computing algorithm can be used to reduce the computing time by many orders of magnitude (by a factor $\sim N/\log N$). The speed of modern computers is such that FFT spectra of the sound of musical instruments can be calculated and displayed with delays of only a few milliseconds, though any such delay will always be limited by the length and hence frequency resolution of the data set being analysed.

A typical implementation of the FFT method for spectral analysis is shown schematically in Fig. 15.13. An input anti-aliasing filter is first used to remove frequency components above the Nyquist limit $f_D/2$; an ADC then converts the incoming signal to a digital output to give a data set of $N = 2^n$ measurements over a time $T = N/f_D$. A windowing function removes problems from discontinuities at the start and end of the recorded set of data, and the computer evaluates the FFT giving the amplitudes and phases of the Fourier components at $N/2$ discrete frequencies spaced $1/T = f_D/N$ Hz apart.

A sequence of FFTs from data taken over successive short periods of time can be used to illustrate the decay of individual partials in transient and decaying waveforms, such as those of a plucked string, a piano note or struck bell, as illustrated for the sound of a plucked violin A-string in Fig. 15.14.

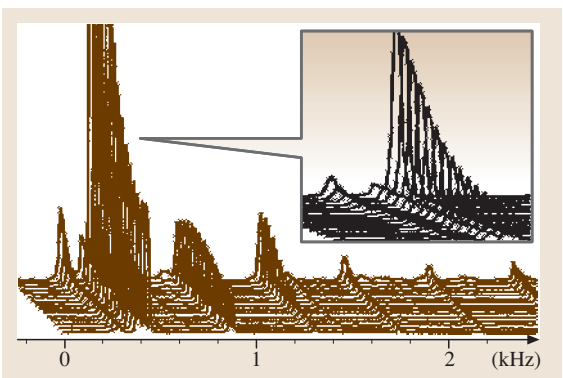


Fig. 15.14 Time sequence of delayed FFTs illustrating the decay of excited modes of a violin, when the A-string is plucked, with an expanded section of the frequency scale for the lowest resonances excited. The time between successive traces is 10 ms

Envelopes of Sound Waveforms

The time dependence or envelope of the amplitude of a sound signal is just as important a factor in the recognition of any musical instrument as the spectrum of the sound produced. In general, the envelope has a starting

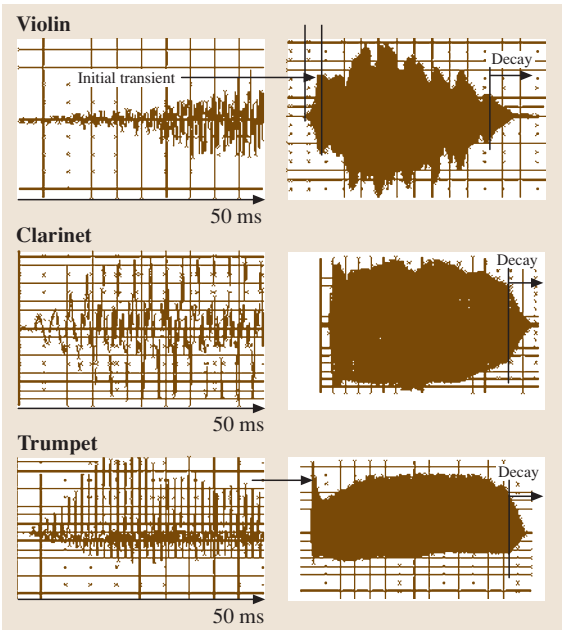


Fig. 15.15 Typical wave envelopes for a violin, clarinet and trumpet with a 50 ms expanded view of the initial starting transient

Amazon
EX1013 Page 559

transient, a period with a quasi-constant amplitude for a continuously bowed or blown instrument, and a period of free decay, when the instrument is no longer being excited. The sound produced by musical instruments is also significantly affected by the acoustic environment in which the instrument is played, but this will be ignored for the moment. Typical initial transients and overall envelopes of single notes played on a violin, clarinet and trumpet are shown in Fig. 15.15.

The starting transient provides an immediate clue to the ear enabling the listener to recognise the instrument being played quickly. However, the characteristic fluctuations in frequency and amplitude within the overall envelope and noise associated with the method of excitation (e.g. bowing and blowing) are just as important in the recognition of specific instruments. This can easily be shown by removing the starting transient from a musical sound altogether, as illustrated in CD-ROM. In this example comparisons are made between the sounds of a violin, flute, trombone and oboe played first with the initial 200 ms transient removed and then with the transient reinserted. In each case the instrument can immediately be identified even in the absence of the starting transient. The audio example ends with a constant amplitude sawtooth waveform having an unvarying sound quite unlike the sound of any real musical instrument.

Nevertheless, the starting transient and subsequent decay of sound are extremely important in the identification of the sounds of plucked or hammered strings and all percussion instruments, where the waveform and spectral content changes very rapidly with time after the start of the note. This is illustrated by the dramatic difference in the unrecognisable sound of a piano when played backwards and then replayed in the normal direction (CD-ROM).

Noise

There are several potential sources of fluctuations in the envelope of musical instruments, which help to characterise their characteristic sounds, such as the breathiness induced by the noise of turbulent air passing over the sound hole in a recorder, flute or organ pipe (*Verge* and *Hirschberg* [15.15]) and irregularities in the sound of any bowed instrument due to inherent noise in the slip–stick bowing mechanism (*McIntyre* et al. [15.16]).

Amplitude and Frequency Modulation

Another important source of fluctuations is vibrato, which involves periodic changes in the amplitude, frequency, or spectral content of a note and often all three (*Meyer* [15.17], *Gough* [15.18]). Vibrato is produced on a stringed instrument by periodically changing the length of the bowed string by rocking the finger stopping the string backwards and forwards. In singing (*Prame* [15.19]) and wind instruments (*Gilbert* et al. [15.20]) vibrato is produced by periodic modulations of the pressure exerted by the lungs or mouth on the exciting reed or air passage.

Amplitude modulation of a sinusoidal frequency component can be expressed as

$$y(t) = (1 + a_\mathrm{m} \cos \Omega t) \sin \omega t$$
$$= \sin \omega t + \frac{a_\mathrm{m}}{2} [\sin(\omega + \Omega)t + \sin(\omega - \Omega)t] \,, \tag{15.27}$$

where Ω is the modulation frequency and a the modulation parameter. Amplitude modulation introduces two "side-bands" with amplitude $a_\mathrm{m}/2$ at frequencies Ω above and below that of the principal central component. The side-bands have a net resultant that remains in phase with the central component giving a fractional change in amplitude $[1 + a_\mathrm{m} \cos \Omega t]$.

Frequency modulation should more strictly be referred to as phase modulation, with the phase varying as

$$\phi(t) = \omega t + a_\mathrm{f} \cos \Omega t \,. \tag{15.28}$$

where a_f is frequency-modulation index. The time-varying frequency can then be defined by the rate of change of phase, such that

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega - a_\mathrm{f} \Omega \sin \Omega t \,, \tag{15.29}$$

with a fractional shift in frequency varying as

$$\frac{\Delta\omega(t)}{\omega} = -a_\mathrm{f} \frac{\Omega}{\omega} \sin \Omega t \,. \tag{15.30}$$

For small modulation index, a phase-modulated wave can be written as

$$y(t) = \sin \omega t + \frac{a_\mathrm{f}}{2} [\cos(\omega + \Omega)t + \cos(\omega - \Omega)t] \,, \tag{15.31}$$

which again results in equally spaced side-bands about the central frequency with amplitude $a_\mathrm{f}/2$, but with a resultant now in phase-quadrature with that of the central frequency giving the above phase modulation.

Because of the multi-resonant frequency response of all musical instruments, changes in driving frequency also induce significant fluctuations in amplitude. Such fluctuations are particularly important for the strongly peaked multi-resonant instruments of the violin family, as illustrated in Fig. 15.15.

Amazon
EX1013 Page 560

15.1.4 Perception and Psychoacoustics

In this section, we briefly highlight a number of psychoacoustic aspects of particularly importance in any discussion of musical acoustics. See also Chap. 13 on Psychoacoustics by *Brian Moore*.

Sensitivity of Hearing

We have already commented that the brain interprets both frequency and intensity on a logarithmic scale. The recognition of familiar intervals such as the octave, perfect fifth, irrespective of the absolute frequencies, provide an immediate example, as is the use of the dB scale in the measurement of sound levels.

In the 1930s, *Fletcher* and *Munson* [15.22] undertook a survey of a large population of subjects to investigate how the sensitivity of the ear varies with frequency and intensity (and age). These measurements were later refined by *Robinson* and *Dadson* [15.21]. Their published values for normal equal-loudness level contours, shown in Fig. 15.16, were adopted by the International Standards Organization, as the original ISO 226 standard for audio sensitivity, with data recently refined to define the new ISO 226:2003 standard.

The plotted curves show population-averaged equal loudness contours for sinusoidal sound waves measured in phons on a dB scale, which equate to sound pressure level (SPL) measurements in dB at 1 kHz. The SPL dB scale is based on a reference root-mean-square pressure of 20 μPa (equivalent to $2\times10^{-5}\,\mathrm{Nm}^{-2}$), which is very close to an qintensity of $10^{-12}\,\mathrm{Wm}^{-2}$. The threshold contour is the population-averaged minimum sound pressure that can be just detected under the quietest environmental conditions. Above sound pressures of ≈ 120 dB, the ear experiences pain and potential permanent damage.

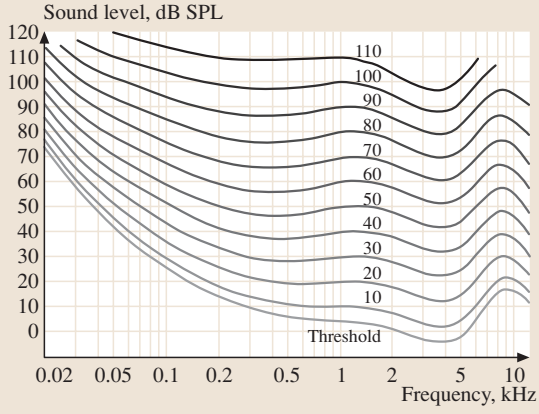


Fig. 15.16 *Robinson–Dadson* [15.21] curves with contours of perceived equal loudness measured in phons on a dB scale

The equal subjective sound level contours reflect the dynamics of the ear's detection system. There is a rapid fall-off in sensitivity at low frequencies, where the efficiency of the outer ear drum considered as a piston detector falls off as ω^4. The fall-off at high frequencies is due to the increasing inertial impedance of the ear drum and bones in the inner ear. However, the fall-off is partially compensated by peaks in sensitivity from the resonances of the outer air channel between the ears and ear-drum. Older people experience a considerable loss in sensitivity at high frequencies, which is strongly correlated with age. Fortunately, the losses are at relatively high frequencies and are generally not too important for the appreciation of music.

The sensitivity of the ear is particularly strong in the frequency range 2–6 kHz, which is important for recognising the consonants in speech. One would therefore expect such frequencies to be equally important in the identification and assessment of sound quality of musical instruments. Below around 200–400 Hz there is an increasingly rapid fall-off in sensitivity, which will differentially affect the subjective loudness of the lower partials of any complex musical sound at these and lower frequencies. At low frequencies the contours of equal amplitude are more closely spaced, so that the effect of increasing the SLP by 20 dB increases the subjective loudness by considerably more. Turning up the volume on any reproduction system changes the perceived quality from a rather thin sound to a more exciting sound with a much stronger bass and a somewhat stronger high-frequency response.

From a musical acoustics viewpoint, it is often sensible to invert the equal contour plot, as the inverted plot essentially acts as a subjective, mid-frequency band filter, de-emphasising the perceived intensities of the lowest and highest-frequency partials in a complex waveform.

Loudness levels will clearly vary with distance from any source. Sounds levels exceed 120 dB close to an aeroplane on take off or close to a loudspeaker in a noisy rock concert, resulting in potential permanent damage to the ear. Sound levels close to a heavily used motorway are typically around 90 dB, around 70 dB inside a car, about 50 dB in an office, 30 dB inside a quiet house at night, 20 dB in a very quiet recording studio and 0 dB inside an anechoic chamber. At the quietest levels, one begins to hear the beating of the heart and

Amazon
EX1013 Page 561

workings of other internal organs, which can be a somewhat disquieting experience. There would clearly be no evolutionary reason to have developed a more sensitive hearing system.

Audio CD-ROM is a short orchestral excerpt played at successively decreasing 6 dB steps (half the amplitude or a quarter the intensity). Musicians indicate the loudness with which music should be played using the dynamic markings *pp*, *p*, *mp*, *mf*, *f* and *ff*, which roughly correspond to a subjective doubling of intensity between each level. Such levels are clearly only relative, since absolute values will vary strongly with the distance of the listener from the source with a 12 dB decrease in intensity on doubling the distance in free space. Although the dynamic range of an individual note on a musical instrument rarely exceeds 20 dB, with only about six distinguishable dynamic levels within this range, the total dynamic range of an instrument is more like 45 dB (*Patterson* [15.23]). However, there is a much larger range of sounds produced by different instruments (e.g. the trombone and violin. The carrying power, penetration and prominence of musical sounds is not simply a matter of absolute intensity, but also depends on the harmonic content and transient structures of the complex tones produced. This helps to explain how a solo violinist can still be heard over the massive sound of a large orchestra.

Subjective Assessment of Pitch

We have already noted that the perceived pitch of a note is determined by the inverse period of a waveform and does not necessarily require the presence of a Fourier component at that frequency. This is illustrated in audio CD-ROM, in which simple sinusoidal tones at 300 Hz and 200 Hz are first played in succession and then played together to produce a repeating waveform sounding an octave lower at 100 Hz, which is then followed by a pure 100 Hz tone of the same amplitude but sounding very much quieter. The final sinusoidal tone at 100 Hz may well not be heard on a typical PC laptop or notebook sound production systems, which radiate little sound below around 200 Hz. The absence of a Fourier component at the pitch of a complex tone is often referred to as the *missing fundamental* phenomenon. It is important in many stringed instruments, which produce very little sound at the actual frequency of their lowest open strings.

The missing fundamental phenomenon is a psychoacoustic rather than a nonlinear effect produced by a distortion of the waveform in the ear. It reflects the way that the ear processes sounds in the time domain at low frequencies (*Moore* [15.24] and Chap. 13).

The ability of the ear to recognise the pitch at which an instrument is playing, even though the lower partials of the sound of individual instruments may be missing is very important in sound reproduction systems. It enables the listener to recognise the distinctive sounds of all the instruments in an orchestra, even when the recording or reproduction system may have a very poor low-frequency response – as in early gramophones and the loudspeakers used in cheap radios and typical PC laptops and notebooks.

Combining tones to produce a lower tone is exploited on the *quint* combination stop on the organ to produce low-pitched sounds (e.g. a 16 ft pipe and a $16 \times 2/3 = 10.66$ ft pipe sounding a fifth above, when sounded together, reproduce the sound of a 32 ft pipe, as illustrated for the combination of 200 and 300 Hz sine waves in CD-ROM above). Interesting, the effect is nothing like so strong in playing two bowed strings a fifth apart (e.g. open A and open E on a violin), presumably because of the very rich spectrum of higher partials and independent fluctuations of the two sounds. However, such sounds can often be heard when two flutes play well-tuned intervals together. The early 18th century virtuoso violinist Tartini recognised the existence of such mysterious tones, whenever pairs of notes were played together in exact intonation [e.g. integer ratios such as 3/2 (perfect fifth), 5/4 (major third), 6/5 (major third)], and reputedly attributed them to the devil. The effect is small, but is still used by violinists when practising playing such intervals exactly in tune.

In general, complex tones are composed of a number of spectral components which have no particular harmonic relationship to each other. One then has to consider what determines the subjective pitch of the perceived sound. This involves the way the brain processes the signal and the relative emphasis given to the spectral components present, which will depend on their frequencies and intensities. It is important to recognise that the perceived pitch is not necessarily that of the lowest-frequency component present. This is illustrated in CD-ROM in which the fundamental and first octave are fixed in frequency, while an intermediate partial is swept upwards from the lower to the higher note. First the fundamental is sounded by itself and then with the octave added producing a note at the same pitch but with a different timbre. An intermediate partial is then added and swept upwards in steps from the lower to the upper note, giving the sense of a note of continuously rising pitch, though the fundamental and its octave remain fixed. Although the fundamental and octave remain fixed, the rising partial gives the sense of a note of in-

Amazon
EX1013 Page 562

creasing pitch. In this particularly simple example, it is relatively easy to identify and follow the pitch of each harmonic component separately. However, for a musical instrument like a gong or bell, with no preconceived knowledge of the likely pitch of the individual partials, this is far more difficult. The perceived pitch of the strike note of a bell and many other percussion instruments, with an inharmonic combination of excited modes, depends in a rather complex way on the relative weighting of the partials present and the musical context.

In assessing the subjective pitch of a note there can often be an ambiguity of an octave in the apparent pitch. This is illustrated by the famous example of the apparent, ever-rising pitch of a note generated by a continuously rising comb of logarithmically spaced frequencies passing through a fixed hearing band of frequencies, CD-ROM [15.25], which appears to be increasing in pitch at all times though clearly repeating itself. This illustrates the *circularity* of pitch perception and is the audio equivalent of the visual illusion of continuously rising steps which return to the same point in space in an Escher drawing.

Precedence Effect

Another important time-domain phenomenon in the perception of musical sounds is the Haas precedence effect, which enables a listener to locate the source of a distant sound from the small difference in time that sound arriving at an angle to the head takes to reach the two ears. The brain gives precedence to the sound arriving first, even though later sounds from other directions may be significantly stronger. Any sound arriving within the first 20–40 ms (depending somewhat on frequency and intensity) of the first sound to arrive simply adds to the perceived intensity of the first sound. This is very important in musical performance, with reflections from close reflecting surfaces adding strongly to the intensity and definition of the music.

The precedence effect is illustrated in CD-ROM. This is a stereo recording of identical clicks recorded on the left and right channels with a delay of 20 ms between them, which is then reversed. Although the clicks are too close together for the ear to distinguish them separately, when replayed through a pair of stereo loudspeakers (not earphones), the sound will appear to come from the speaker providing the earlier click.

The precedence effect is one of the ways in which one can locate the origin of a particular sound within an orchestra or the sound of a particular voice in a crowded room. Once located, the brain is able to focus on the subsequent source even against a highly confusing background of other sources. It is likely that fluctuations within the characteristic sound of an individual person or musical instrument enable the brain to focus continuously on a particular source. In musical acoustics one must always recognise the formidable power of the brain's auditory processing capabilities, which is far beyond what can be achieved using present-day computers. Consequently, even very small effects on a physical measurement scale can have a very significant effect on the listener's subjective response to the sound of a particular instrument.

15.2 Stringed Instruments

In this section we describe the production of sound by the great variety of musical instruments based on the plucking, bowing and striking of stretched strings. This will include an introduction to the different modes of string vibrations excited by the player, the transfer of energy from the vibrating string to the acoustically radiating structural vibrations of the body of the instrument via the bridge, and the modification of such sound by the environment in which the instrument is played. Although the production of sound is based on the vibrations of relatively simple structures, such as strings and plates, it is the interactions between these, extending the physics well beyond introductory text-book treatments, which results in the characteristic sounds of individual stringed instrument, as summarised in this section.

The Physics of Musical Instruments by *Fletcher* and *Rossing* ([15.5], Chaps. 9–11) provides an authoritative account of the acoustics of a wide range of string instruments, and a comprehensive set of references to the research literature prior to 1998. The four volumes of research papers on violin acoustics, collated and edited by *Hutchins* [15.26] and *Hutchins* and *Benade* [15.27], also includes excellent introductions to almost every aspect of the acoustics of instruments of the violin family, much of which is just as relevant to other stringed instruments. Carleen Hutchins has been an inspirational figure in the field of violin acoustics. The Catgut Acoustical Society, which she cofounded, published a journal and an earlier newsletter [15.28] containing many important papers on violin research of interest to both professional acousticians and violin makers. Her inspiration has en-

Amazon
EX1013 Page 563

couraged a world-wide school of violin makers, who use scientific measurements and plate tuning in particular as an aide to making high-quality instruments. The comprehensive monograph on the *Physics of the Violin* by *Cremer* [15.29] provides an invaluable theoretical and experimental survey of research on instruments of the violin family, with particular emphasis on the bowed string, the action of the bridge, the vibrations of the body and the radiation of sound.

The production of sound by any stringed instrument is based on the same acoustic principles. The player excites the vibrations of a stretched string by bowing, plucking or striking. Energy from the vibrating string is then transferred via the supporting bridge to the acoustically radiating structural vibrations of the instrument. The radiated sound is then conditioned by the performing environment.

There are many different types of stringed instruments formally classified as chordophones. Harp-like lyres appear in Sumerian art from around 2800 BC. However, more primitive instruments, like a plucked string stretched over a bent stick and resonated across the mouth, probably date back to soon after the emergence of man the hunter [15.30, 31].

Stokes [15.32] was the first to recognise that the vibrating string was essentially a linear dipole, which radiated a negligible amount of sound at low frequencies (see also *Rayleigh* [15.3] Vol. 2, Sect. 341). To produce sound, the vibrating string has to excite the vibrations of a much larger area radiating surface. For bowed and plucked instruments, such as members of violin, lute and guitar families, almost all the sound is radiated by the shell of the instrument, with the acoustic output at low frequencies usually boosted by the Helmholtz resonance of the air inside the instrument vibrating in and out of the f- or rose-holes cut into the front plate. On larger instruments, such as the piano and harp, the sound is radiated by a large soundboard.

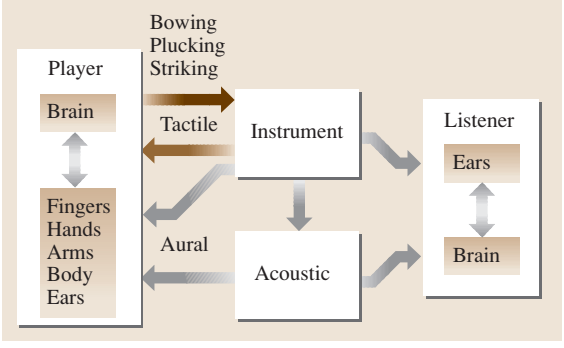


Fig. 15.17 A schematic representation of the complex feedback and sound radiation systems involved in the generation of sound by a bowed string instrument

For any continuously bowed (or blown) instrument, the sound is conditioned by a complex feedback loop involving the instrument, player and surrounding acoustic, illustrated schematically for the violin in Fig. 15.17. The expert string player controls the intonation and quality of the sound produced using slight adjustments of the position of the left-hand fingers stopping the string, and the pressure, velocity and position of the bow on the string, in response to the sound heard from both the instrument and the surrounding acoustic. In addition, there is direct tactile feedback through the fingers of both the left hand controlling the pitch of the note and the right hand controlling the bow. A similar overall feedback system is also involved in playing woodwind or brass instrument. The perception of the sound by both player and listener is also strongly influenced by the performing acoustic and the way the brain processes the sound received by the sensory organs in the ears, as illustrated schematically in Fig. 15.17. All such factors are involved in determining the perceived quality of the sound produced by a musical instrument. However, for simplicity and physical insight into the various mechanisms involved, it is convenient to consider the acoustics of musical instruments in terms of their component parts, like the vibrating string, the supporting bridge and shell of the instrument. Nevertheless, it is important not to lose sight of the fact that the sound produced by any instrument will involve the interactions of all such subsystems and, even more importantly, the skill of the player in exciting and controlling the vibrations ultimately responsible for the sound produced.

15.2.1 String Vibrations

The transverse vibrations $\xi(x, t)$ of a perfectly flexible stretched string, of mass μ per unit length and tension T, satisfy the one-dimensional wave equation (d'Alembert, 1747)

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{1}{c_T^2} \frac{\partial^2 \xi}{\partial t^2} , \qquad (15.32)$$

where the velocity of transverse waves $c_T = \sqrt{T/\mu}$. The tension $T = ES\Delta L/L$, where E is Young's modulus, S is the cross-sectional area of the string and $\Delta L/L$ is the fractional stretching of the string over its length L. For the relatively small transverse displacements of bowed and plucked strings on musical instruments, changes in tension can be ignored. However, at larger ampli-

Amazon
EX1013 Page 564

tudes, a number of interesting nonlinear effects can be observed, which will be described in Sect. 15.2.2.

A string can also support longitudinal and torsional modes, with velocities $c_L = \sqrt{E/\rho}$ and $c_\theta = \sqrt{E/\rho(1+\nu)}$, where ρ is the density and ν is the Poisson ratio (≈ 0.35 for most materials). The Poisson ratio ν is the ratio of transverse to longitudinal strain when the material is stretched along a given direction. For strings on musical instruments, the longitudinal and torsional wave velocities are typically an order of magnitude larger than the transverse velocity, with $c_L/c_T \sim \sqrt{L/\Delta L}$.

Although both longitudinal and torsional waves play important roles in the detailed physics of the bowed, plucked and struck string, the musically important modes of string vibration are the transverse modes – apart from unwanted squeaks from longitudinal modes, which are often excited by the beginner on the violin. Unless otherwise stated, we only consider transverse waves and drop the defining subscript, unless a distinction needs to be made.

Waves on an ideal string are dispersionless (independent of frequency), so that any wave initially excited on the string will travel along the string without change in amplitude or shape. D'Alembert obtained a general solution of the wave equation of the form

$$f(x,t) = f_1(x+ct) + f_2(x-ct)\,, \tag{15.33}$$

corresponding to two waves of unchanging shape travelling with wave velocity c in opposite directions along the string.

If the string is supported rigidly at its ends, the propagated waves are reflected with a change in sign giving zero displacement at the nodal end-points. Each propagating wave will continue to be reflected with change of sign on reflection at each end. For a string of length L, the string displacement will therefore return to its initial state in multiples of the transit time $2L/c$. The same is also true for the velocity and acceleration waveforms, since, if f satisfies the wave equation, then so must all its temporal and spatial derivatives, $\partial^n f/\partial x^n$ and $\partial^n f/\partial t^n$. It follows that the repetition frequency of any freely propagating wave on a given length of a stretched string will always be the same, however the string is excited (e.g. sinusoidally or by plucking, bowing or striking).

Excitation of Vibrations

First consider a string subject to a localised force F applied suddenly at a point along its length. This causes the string to move with velocity v at the point of contact exciting transverse waves travelling outwards in both directions with velocity c, as illustrated schematically in Fig. 15.18.

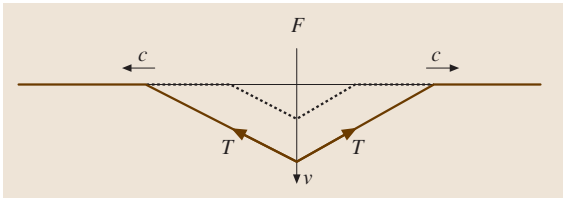


Fig. 15.18 Transverse motion of string induced by a localised force, with the *dotted lines* indicating the displacement at an earlier time

In a short time δt, the transverse waves travel a distance $c\delta t$ along the string while the string at the point of contact is displaced by a transverse distance $v\delta t$. For $v \ll c$, one can make the usual small-angle approximations, so that equating the applied force to the transverse force from the deformed string, we obtain

$$v = \left(\frac{c}{2T}\right) F = \frac{1}{2R_0} F\,, \tag{15.34}$$

where $R_0 = \mu c$ is the characteristic impedance (force/induced velocity) of the string, which for an ideal string ignoring intrinsic losses is purely resistive. The factor of two in the above equation arises because the force acts on the two semi-infinite lengths of string in parallel. In practice, any discontinuity in slope will be rounded by the finite flexibility of real strings, as discussed later.

Force on End-Supports

The characteristic resistance R_0 of the string is an important parameter, because it determines the transfer of energy from the vibrating string to the acoustically radiating modes of the instrument via the supporting bridge at the end of the string. The transverse force exerted by the string on an end-support at the origin can be written as $F_B = T(\partial\xi/\partial x)_0$. This induces a transverse velocity at the point of string support given by

$$v_B = \frac{1}{Z_B} F_B = A_B F_B\,, \tag{15.35}$$

where Z_B and A_B are the frequency-dependent characteristic impedance and admittance at the end-support. In general, the induced velocity at the point of string support on the supporting bridge will differ in phase from that of the driving force, so that $Z(\omega)$ and $A(\omega)$ will be complex quantities.

The bridge on a musical instrument is never a perfect node otherwise no energy could be transferred to the radiating surfaces of the instrument. Waves on the string

Amazon
EX1013 Page 565

are reflected at the bridge with a frequency-dependent reflection coefficient r and a fractional loss of energy ε given by

$$r = \frac{R_0 - Z_B}{R_0 + Z_B} \quad \text{and} \quad \varepsilon = \frac{2R_0\left(Z_B + Z_B^*\right)}{|R_0 + Z_B|^2}\,, \qquad (15.36)$$

where Z_B^* is the complex conjugate of the complex impedance at the terminating bridge.

For strings on musical instruments, $R_0 \ll |Z_B|$, so that to a first approximation we can consider the bridge as a node. If this were not so, the vibrational frequencies of strings would by strongly perturbed from their harmonic values. Nevertheless, first-order corrections are important, as they determine the energy transfer from the strings to the body of the instrument and hence the intensity of the radiated sound. The coupling via the bridge also affects the string vibrations themselves, with the resistive losses at the bridge causing damping and the reactive component of the admittance perturbing their vibrational frequencies, as described in Sect. 15.2.3. Such perturbations can sometimes be so large that it is no longer possible to sustain a stable bowed note, resulting in what is known as a *wolf-note* (for an illustration of a bad wolf-note on the cello listen to CD-ROM).

Before considering the interaction of real strings with the supporting structure, we first consider the simplest cases of sinusoidal and simple Helmholtz modes of vibration on an ideal string with perfectly rigid end-supports.

Sine-Wave Modes

An ideally flexible string stretched between rigid end-supports a distance L apart can support standing waves, or eigenmodes, with transverse string displacements given by

$$\xi_n(x, t) = a_n \sin\left(\frac{n\pi x}{L}\right)\cos\left(\omega_n t + \phi_n\right)\,, \qquad (15.37)$$

where $\omega_n = 2\pi f_n$ and a_n is the amplitude of the nth mode with frequency $f_n = nc/2L$ and phase ϕ_n. Such modes can be considered as the sum of two sine waves of the d'Alembert form (15.33) travelling in opposite directions. For an ideal string, these solutions form a complete orthogonal set of *eigenmodes* with a *harmonic set* of *eigenfrequencies*, which are integer multiples of the fundamental frequency $c/2L$.

The resonant response of individual modes of a metal or metal-covered string can be investigated, for example, with a photosensitive device to detect the transverse string motion induced by a sinusoidal current passing through the string placed in a magnetic field to give a transverse Lorentz force (*Gough* [15.33]).

Because the wave equation is linear, any waveform, however excited, can be described as a Fourier sum of harmonic modes, such that

$$\xi_n(x, t) = \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right) \left[A_n \cos\left(\omega_n t\right) + B_n \sin\left(\omega_n t\right)\right]\,, \qquad (15.38)$$

where the Fourier coefficients A_n and B_n are determined by the initial transverse displacement and velocity along the length of the string, so that

$$A_n = \frac{2}{L}\int_0^L \xi(x, 0)\sin\left(\frac{n\pi x}{L}\right)\mathrm{d}x\,, \qquad (15.39)$$

and

$$B_n = \frac{2}{L\omega_n}\int_0^L \frac{\mathrm{d}\xi(x, 0)}{\mathrm{d}t}\sin\left(\frac{n\pi x}{L}\right)\mathrm{d}x\,. \qquad (15.40)$$

The transverse force on the end-support at $x = L$ is given by

$$\begin{aligned} F_{\text{end}} &= -T\left(\frac{\partial \xi}{\partial x}\right)_L \\ &= -T\sum_n \left(\frac{n\pi}{L}\right)(-1)^n \\ &\quad \left[A_n \cos\left(\omega_n t_n\right) + B_n \sin\left(\omega_n t\right)\right]\,. \end{aligned} \qquad (15.41)$$

Helmholtz Modes

Although many physicists and most musicians intuitively associate waves on strings with the sinusoidal waves of textbook physics, in practice, the vibrations of a bowed, plucked or struck string are very different. Nevertheless, because such waves are repetitive, it follows from the Fourier theorem that all such solutions can be described as a sum of sinusoidal wave components. However, the motions of plucked, bowed and struck strings are much more easily described by what are known as Helmholtz solutions to the wave equation [15.34]. These are illustrated for the plucked and bowed string in Fig. 15.19a,b.

The Helmholtz solutions are made up of straight-line sections of string. There is no net force acting on any small segment within any such section, because the transverse tension forces acting on its ends are equal and opposite. By Newton's laws, any such segment must therefore be either at rest or moving with constant velocity. Only where there is a *kink* or discontinuity in the

Amazon
EX1013 Page 566

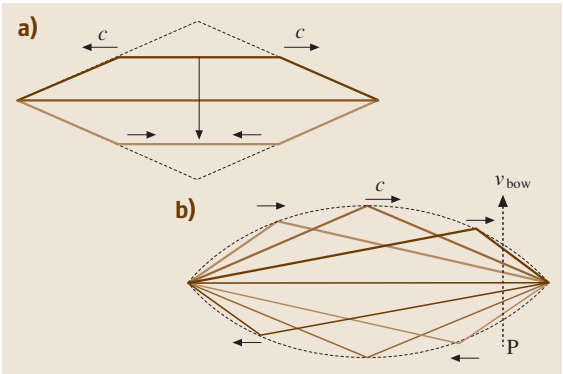


Fig. 15.19a,b Helmholtz waveforms for **(a)** a centrally plucked and **(b)** a bowed string. The *horizontal arrows* indicate the directions that the kinks are travelling in and the *vertical arrows* the directions of the moving string sections. The *different colours* represent string displacements at different times. P indicates a typical bowing position along the string

slope between adjacent straight-line sections (equivalent to a δ function in the spatial double derivative) can there be any acceleration. From our earlier discussion, any such kink must travel backwards and forwards along the string at the transverse string velocity c, reversing its sign on reflection at the ends. As the kink moves past a specific position along the string, the difference in the transverse components of the tension on either side of the kink results in a localised impulse, which changes the local velocity of the string from one moving or stationary straight-line section to the next. In general, there can be any number of Helmholtz kinks travelling along the string in either direction, each kink marking the boundary between straight-line sections either at rest or moving with constant velocity. Similar solutions also exist for torsional and longitudinal waves.

We now consider the Helmholtz wave solutions for the plucked, bowed and hammered string in a little more detail.

Plucked String

Consider an ideal string initially at rest with an initial transverse displacement a at its mid-point, as illustrated in Fig. 15.19a. On release, kinks will propagate away from the central point in both directions with velocity c, but points on the string beyond the kinks will remain at rest. When the kink arrives at a particular point along the string, the associated impulse will accelerate the string from rest to the uniform velocity of the central section of the string. After a time t, the solution therefore comprises

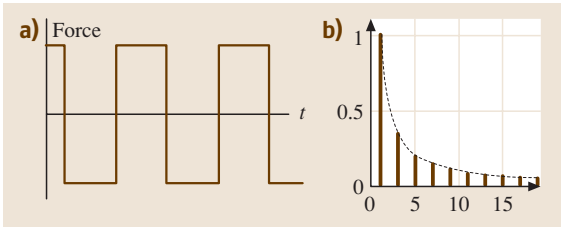


Fig. 15.20 **(a)** Square-wave time dependence of transverse force acting on the bridge from a string plucked at its centre and **(b)** the corresponding amplitudes of the odd Fourier components n varying as $1/n$ (*dotted curve*)

a straight central section of the string of width $2ct$ moving downward with constant velocity c ($2a/L$), with the outer regions remaining at rest until a kink arrives. After a time $L/2c$, the kinks separating the straight-line sections reach the ends and are reflected with change of sign. After half a single period L/c, the initial displacement will therefore be reversed and will return to the original displacement after one full period $2L/c$. In the absence of damping, the process would repeat indefinitely.

Now consider the transverse force acting on the end-support responsible for exciting sound through the induced motion of the supporting bridge and vibrational modes of the instrument. The initial transverse force on the bridge is $2Ta/L$, where we assume $a \ll L$. This force is unchanged until the first kink arrives. On reflection, the direction of the force is reversed and is reversed again when the second kink returns after reflection from the other end of the string. The two circulating kinks therefore cause a reversal in sign of the force on the end-supports every half-cycle, resulting in a square-wave waveform, as illustrated in Fig. 15.20. The spectrum of a square wave has Fourier components at odd multiples n of the fundamental frequency with amplitudes proportional to $1/n$, Fig. 15.20b.

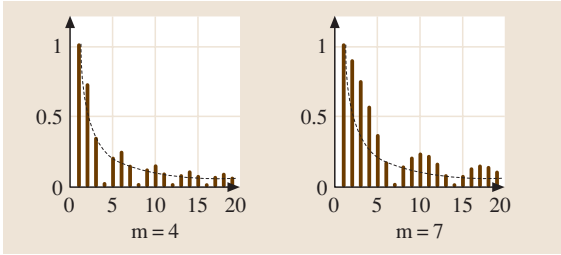


Fig. 15.21 Normalised Fourier amplitudes for the force on the bridge for a string plucked 1/4 and 1/7 of the string length from the bridge. The *dashed curves* show the $1/n$ envelope of the partials of a sawtooth waveform

Amazon
EX1013 Page 567

Note that plucking a string at the mid-point excites only the odd-n modes. This is a consequence of the initial force being applied at a node of all the even-n modes. If a string is plucked at a fractional position $1/m$ along its length, any partial that is an integer multiples of m will be missing. This is illustrated in Fig. 15.21, showing the spectra of the force on the supporting bridge for a string plucked at points 1/4 and 1/7-th along its length. By selecting the plucking position along the string, the guitar or lute player can change the harmonic content of the sound produced. When plucked near the bridge, the sound of the plucked guitar string is rather bright, with nearly all the prominent partials almost equally strongly excited (audio CD-ROM).

In practice, the finite width of the plucking point, the finite rigidity of the string and the loss of energy at the bridge perturb the Helmholtz wave, removing the unphysical discontinuities of the idealised model. This results in a more rapid decrease in the intensities of the higher partials excited.

Bowed String

The motion of the bowed string can be described rather accurately by a simple Helmholtz wave with a single kink circulating backwards and forwards along the string. The kink now separates two straight sections moving with constant angular velocity about the nodal end-points, as illustrated in Fig. 15.19b. This is again a solution that satisfies Newton's laws of motion, with the only acceleration occurring as the kink arrives at a particular point along the string. Such a wave is just as valid a solution to the wave equation as a sine wave and once excited would continue indefinitely, if there were no damping or energy losses on reflection at the bow or supported ends.

The energy required to excite and maintain such a wave is provided by frictional forces between the moving bow hair and the string, involving what is known as the *slip–stick* excitation mechanism. For a typical bowing position, marked by the line at P in Fig. 15.19b, the friction between the bow and string forces the string to remain in contact with the bow hair moving with constant bow velocity. This is referred to as the *sticking regime* and occurs all the time the kink is travelling to the left of the bowing position. However, when the kink is between the bow and supporting bridge, the string moves in the opposite direction to the bow. This is the *slipping regime*. Such motion is possible because the sliding friction between the bow and string can be much smaller than the sticking friction, when the bow and string are in contact. In this highly idealised model, the frictional force is assumed to be infinite in the sticking regime and zero in the slipping regime.

A more detailed discussion of the slip–stick bowing mechanism will be given later (Sect. 15.2.3), taking into account more-realistic models for the frictional forces between the bow and string and the transfer of energy from the string to the vibrational modes of the structure via the bridge. However, the idealised Helmholtz motion provides a surprisingly good description of the vibrations of real strings, as confirmed in early measurements by *Raman* [15.35] and many more-recent publications to be cited later.

The amplitude of the Helmholtz bowed waveform is determined by the velocity of the bow v_{bow} and its distance L_{B} from the bridge. The transverse displacement of the kink maps out a parabolic path as it traverses the string (Fig. 15.19b). At the mid-point, the string displacement executes a triangular-wave motion with time, moving with velocity $\pm 2ca/L$ in alternate half-periods, where the maximum displacement $a = (L^2/4L_{\text{B}})(v_{\text{B}}/c)$, for $a \ll L$. At the bowing position, the transverse string velocity alternates between v_{bow} in the sticking regime and $-v_{\text{bow}}\,(L - L_{\text{B}})\,/L_{\text{B}}$ in the slipping regime, as illustrated in Fig. 15.22.

To increase the sound, the player can therefore either use a faster bow speed or play with the bow nearer the bridge. *Schelling* [15.36] has shown that more-realistic frictional models limit the playing range, as discussed later (Fig. 15.31).

The transverse force on the bridge produced by an idealised Helmholtz bowed wave has a sawtooth-

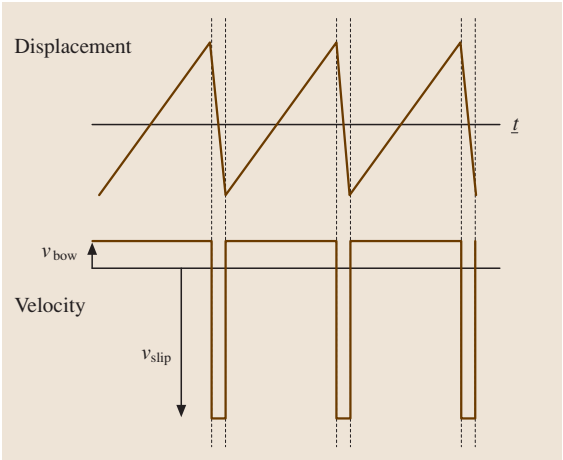


Fig. 15.22 Displacement and velocity of string at the bowing point. The mark-to-space ratio in the velocity is the same as the division of the string by the bow

Amazon
EX1013 Page 568

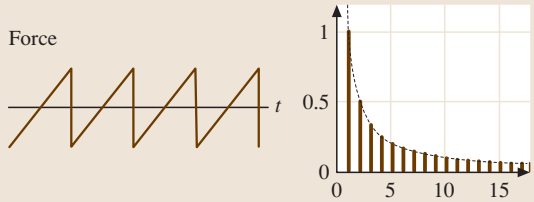


Fig. 15.23 Sawtooth time-dependence of force on the end supports from Helmholtz bowed waveform and corresponding amplitudes of the normalised Fourier spectrum with partials varying as $1/n$

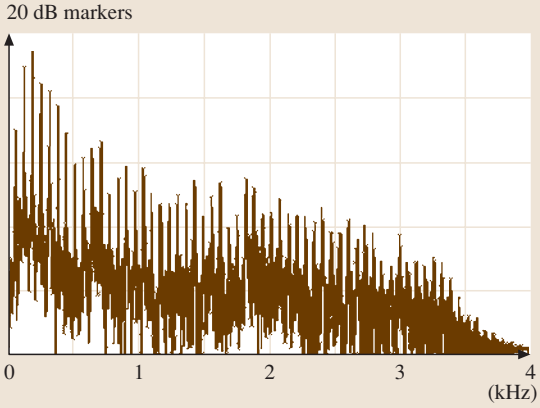


Fig. 15.24 The spectrum of the intensity of the lowest bowed note on a cello, illustrating the very large number of partials contributing to the sound of the instrument

waveform time dependence, as shown in Fig. 15.23. Each time the kink is reflected at the bridge, the transverse force acting on the bridge reverses in sign. It then increases monotonically with time until the process repeats again. The sense of the sawtooth motion reverses with bow direction. The spectrum of the force acting on the bridge includes both even and odd partials, with amplitudes varying as $1/n$.

The spectrum of the sound produced by the lowest plucked and bowed notes on stringed instruments can typically involve 40 or more significant harmonic partials, as illustrated in Fig. 15.24 by the spectrum of the sound produced by a bowed cello open C-string (C2 at ~64 Hz, audio CD-ROM). The FFT spectrum is plotted on a dB scale to illustrate the large range of amplitudes of the partials (Fourier components) excited. The amplitudes of the individual partials depend not only on the force at the bridge exerted by the plucked or bowed strings, but also on the frequency dependent response and radiative properties of the supporting structure, as discussed later.

Struck String

Many musical instruments are played by striking the string with a hammer. The hammer can be quite light and hard, as used for playing the dulcimer, Japanese koto and many other related Asian instruments, or relatively heavy and soft, like the felted hammers on a piano. Some time after the initial impact, the striking hammer bounces away from the string, leaving the string in a free state of vibration. There are a few instruments, such as the clavichord (*Thwaites* and *Fletcher* [15.37]), where the string is struck with a metal bar (the tangent), which remains in contact with the string, defining its vibrating length and hence the note produced.

Consider first a point mass m moving with velocity v striking an ideal stretched string of infinite extent. In any small increment of time, the moving mass will generate a wave moving outwards from the point of impact. This will result in a decelerating force on the mass equal to $2Tv/c = mcv$, as illustrated in Fig. 15.18. The displacement of the mass will then be described by the following equation of motion

$$m\frac{\mathrm{d}^2\xi_\mathrm{m}}{\mathrm{d}t^2} = -\frac{2T}{c_\mathrm{T}}\frac{\mathrm{d}\xi_\mathrm{m}}{\mathrm{d}t}\,. \qquad (15.42)$$

The transverse velocity of the impacting mass therefore decays exponentially with time as

$$\frac{\mathrm{d}\xi_\mathrm{m}}{\mathrm{d}t} = v_\mathrm{m}\exp\left(-t/\tau\right)\,, \qquad (15.43)$$

with $\tau = mc/2T$. The is identical to the dynamics of a trapeze artist dropping onto a stretched wire, with waves of displacement and velocity travelling outwards in both directions away from the point of impact, as illustrated in Fig. 15.25.

In general, the string will be struck at a distance a from one of its end-supports. Hence, in a time $(a/2L)T_0$, a reflected wave will return to the mass and exert an additional force, which will tend to throw the mass back

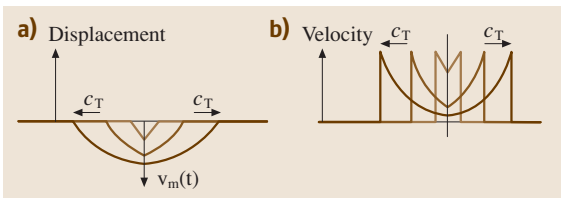


Fig. 15.25a,b Time sequences of **(a)** string displacement and **(b)** string velocity for a mass striking a string

Amazon
EX1013 Page 569

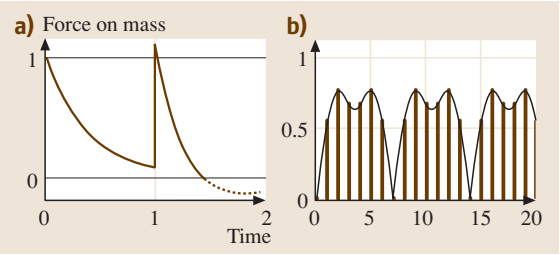


Fig. 15.26 **(a)** Time dependence of the upward force acting on a light hammer impacting a string in arbitrary units of time, and **(b)** amplitudes of Fourier coefficients of force acting on end supports for a hammer after hitting the string 1/7 of the length from an end-support. The *continuous line* shows the continuous spectrum for a string of semi-infinite length

off the string. However, because the mass cannot change its velocity instantaneously, any returning wave will be partially reflected, so that the mass acts as a source of secondary reflected waves travelling outwards in both directions. The total force acting on the hammer is then given by any residual force from the first impact plus the subsequent forces created by the succession of reflections from the end-supports. This problem was first correctly solved by Hall, in the first of four seminal papers on the string–piano hammer interaction [15.38–41].

Hall showed that the first reflected wave exerts an additional decelerating force $g(t') \sim (1 - t'/\tau)\mathrm{e}^{-t'/\tau}$ on the mass, where t' is the time after arrival of the first reflection. This is illustrated in Fig. 15.26 for a relatively light mass impacting the string at a position 1/7-th of the string length from an end. Provided the mass is sufficiently small, the force from the initial impact will have decayed significantly by the time the first reflection returns, so that the force acting on the mass will become negative (the dotted section in Fig. 15.26a), and the mass will detach itself from the string. The string will then move away from the mass and will vibrate freely, provided the hammer is prevented from falling back onto the string. An elaborate mechanism is used on the piano to prevent this from happening (see *Rossing*, *Fletcher* [15.5], Sect. 12.2), while the zither or dulcimer player quickly lifts the hammer well away from the string after the initial impact using much the same striking action as a percussionist playing a drum, where the same considerations apply.

The heavier the mass, the longer it will remain in contact with the string. Hall showed that it may then take several reflections from both ends of the string and sometimes several periods of attachment and detachment before the mass is finally thrown away from the string. A sufficiently heavy mass will never bounce back off the string.

In general, the waveforms excited on the string will therefore be rather complicated functions of the properties of the string, hammer and striking position. However, for a very light mass ($\ll$ mass of the string), which is thrown off the string by the first reflected wave, the Fourier coefficients of the induced velocity waveform, and hence the force on the end-supports, are given by $v_n \approx \left(1 + \mathrm{e}^{-1+\mathrm{i}n\pi\alpha}\right) \sin(n\pi\alpha)$, where $\alpha = a/L$, illustrated in Fig. 15.26b for an impact 1/7-th of the way along the string. Note that the seventh harmonic is missing, as again expected from general arguments, since no work can be transferred to a particular mode of string vibration for a force applied at a nodal position.

In practice, the spectrum is affected by the finite size of the hammer, multiple reflections occurring before the hammer is thrown from the string, and the elastic and often hysteretic properties of the hammer material [15.38].

Striking Tangent

On the clavichord (*Fletcher*, *Rossing* [15.5], Sect. 11.6), a string is struck by a rising end-support, or tangent, which remains in contact with the string, exciting transverse vibrations of the string on both sides of the tangent. If we assume a simplified model in which the rising tangent moves with constant velocity until its final displacement a is reached, there is again a simple Helmholtz wave solution. In practice, the length of string on one side of the tangent is damped, so that free vibrations are only excited on one side of the striking point. We therefore need only consider the length of string between the tangent and the end connected to the soundboard. The discontinuities $\pm v$ in the tangent velocity, occurring on initial impact and on reaching its final displacement after a time Δt, generate propagating kinks and discontinuities of velocity of opposite sign separated in time by Δt. The striking therefore excites waves with kinks, velocities and displacements along the string shown in Fig. 15.27a. The solutions are again Helmholtz waves, but now with two kinks of opposite signs travelling around the string in the same direction.

The Fourier coefficients of the velocity waveform shown in Fig. 15.27b can be written as

$$c_n \sim \frac{1}{n}\frac{a}{L'}\left(1 - \mathrm{e}^{\mathrm{i}n2\pi\beta}\right) , \qquad (15.44)$$

where $\beta = \Delta t/T_1$ is the fraction of the period $T_1 = L'/2c$ of the freely vibrating length of string during which the

Amazon
EX1013 Page 570

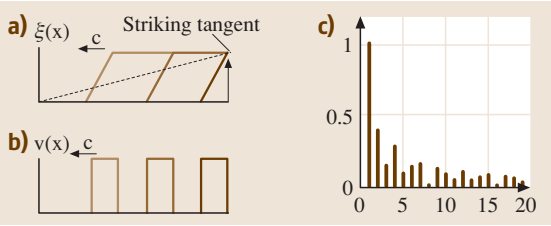


Fig. 15.27a–c Waveforms for a tangent hitting and sticking to string: **(a)** displacement and **(b)** velocity profiles along the string, at a succession of times (different colours) after the tangent hits the string, and **(c)** the spectrum of the resultant force on the end-supports for $\beta = 3/8$ (see text)

striking tangent moves from its initial to final position. Figure 15.27 also includes the spectrum of the force acting on the end-supports for $\beta = 3/8$.

Very similar modes to the above will be excited during the time a heavy hammer is initially in contact with the string on instruments like the dulcimer, zither and piano. Such modes therefore contribute to the initial transient sounds of such instruments. Another related example is the use of *col legno* on stringed instruments, when the strings are struck by the wooden part of the bow. By hitting the string at specific positions along the string, pitched initial transients can be produced, creating special sound effects, *nageln*, sometimes used in avant-garde contemporary music, audio CD-ROM). The above simplistic model for striking a clavichord string will, in practice, be modified by the way the player depresses the key, which is directly coupled to the rising tangent, both during and after the initial impact. The player can therefore influence the initial transient and the after-sound, including the use of a small amount of vibrato on the after-note, resulting in a particularly responsive and intimate but quiet instrument, which was particularly popular in the baroque period.

Real Strings

We now consider a number of departures from the above idealised models for real strings including:

1. the finite size of the plucking or striking point
2. the finite flexibility of the string
3. nonlinear effects

In a subsequent section, we consider the even larger perturbations resulting from coupling to the acoustically radiating modes of the body of the instrument via the bridge.

Finite Spatial Variation

Idealised models for the string, with infinitely sharp kinks produced by plucking, bowing or striking, involve waveforms with discontinuities in amplitude and slope and an infinite number of Fourier components are clearly unphysical. In practice, physics and geometrical limitations, like the finite size of the player's finger or plectrum, will always limit the maximum curvature of the string at the point of excitation. The kinks will therefore no longer be δ-functions (infinitely narrow) but will have a finite size. For illustration, travelling kinks can be modelled as Gaussian waveforms, $\xi_{\pm}(x,t) \sim \exp[-(x \pm ct)^2/2\,(\Delta x)^2]$, which approximate to δ-functions when $\Delta x \to 0$, where Δx characterises the width of the kink. The Fourier transform of such a function has a Gaussian distribution of Fourier coefficients varying as $c(k) \sim \exp[-(k/\Delta k)^2/2]$, where $\Delta k \Delta x = 1$. This is analogous to the uncertainty principle in position and momentum in quantum wave mechanics. For long bending lengths, the amplitudes of the higher-frequency Fourier components will be strongly attenuated.

The sound of a guitar string played with a sharp plectrum is therefore much brighter, with many more contributing higher partials, than when played with the fleshy part of a finger, which limits the bending radius to a few mm. This is illustrated by the sound of an acoustic guitar plucked first with a plectrum and then with the thumb, both at a distance of $\approx 10\,\mathrm{cm}$ from the bridge (audio CD-ROM).

Finite Rigidity

Even for an infinitely narrow plectrum, the bending at the plucking point will be limited by the finite flexibility of the string. The wave equation is then modified by an additional fourth-order bending stiffness term (*Morse* and *Ingard* [15.42, (5.1.25)]),

$$\rho S \frac{\partial^2 \xi}{\partial t^2} = T \frac{\partial^2 \xi}{\partial x^2} - ES\kappa^2 \frac{\partial^4 \xi}{\partial x^4}\,, \qquad (15.45)$$

where E is Young's Modulus, S is the cross-sectional area of the string (assumed homogeneous) and κ its radius of gyration. For a uniform circular wire of radius a, $S\kappa^2 = \pi a^4/4$. Using dimensional arguments, any changes in slope of the string will take place over a characteristic length $\delta \sim (ES\kappa^2/T)^{1/2} = (a^2 L/2\Delta L)^{1/2}$, where ΔL is the extension of the string required to bring it to tension. This provides an intrinsic limit to the sharpness with which the string is bent and hence to the wavelength and frequency of the highest partials con-

Amazon
EX1013 Page 571

tributing significantly to the sound of a plucked, bowed or struck string.

The additional stiffness energy required to bend the string will also affect wave propagation on the string and the frequencies of the excited modes. Assuming sinusoidal wave solutions varying as $e^{i(\omega t \pm kx)}$, the modified wave equation (15.45) gives modes with resonant frequencies

$$\omega_n^2 = c^2 k_n^2 \left(1 + \delta^2 k_n^2\right) . \tag{15.46}$$

Waves on a real string are therefore no longer dispersionless, but travel with a phase and group velocity that depends on their frequency and wavelength. Any Helmholtz kink travelling around a real string will therefore decrease in amplitude and will broaden with time. To maintain the Helmholtz slip–stick bowed waveform, with a well-defined single kink circulating around the string, the bow has to transfer energy to the string to compensate for such broadening each time the kink moves past the bow (*Cremer* [15.29], Chapt. 7 and Sect. 15.2.2.

If a rigidly supported string is free to flex at its ends (known as a hinged boundary condition), solutions of the form $\sin(n\pi x/L)\sin(\omega t)$. However, the mode frequencies remain are no longer harmonic;

$$\frac{\omega_n^*}{\omega_n} = \left(1 + Bn^2\right)^{1/2} \sim 1 + \frac{1}{2} Bn^2 , \tag{15.47}$$

with $B = (\pi/L)^2 \delta^2$, where the expansion assumes $Bn^2 \ll 1$.

When a string is clamped (e.g. by a circular collet), it is forced to remain straight at its ends. *Fletcher* [15.43] showed that this raises all the modal frequencies by an additional factor $\sim [1 + 2/\pi B^{1/2} + (2/\pi)^2 B]$. For a real string supported on a bridge, connected to another length of tensioned string behind the bridge, the boundary conditions will be intermediate between hinged and clamped.

Kent [15.44] has demonstrated that finite-flexibility corrections raise the frequency of the fourth partial of the relatively short C5 (an octave above middle-C) string on an upright piano by 18 cents relative to the fundamental. The inharmonicity would be even larger for the very short, almost bar-like, strings at the very top of the piano. However, the higher partials of the highest notes on a piano rapidly exceed the limits of hearing, so that the resulting inharmonicity becomes somewhat less of a problem. The inharmonicity of the harmonics of a plucked or struck string results in dissonances and beats between partials, providing an *edge* to the sound, which helps the sound of an instrument to penetrate more easily. This is particularly true for instruments like the harpsichord and the guitar when strung with metal strings.

Finite-rigidity effects are particularly pronounced for solid metal strings with a high Young's modulus. To circumvent this problem, modern strings for musical instruments are usually composite structures using a strong but relatively thin and flexible inner core, which is over-wound with one or more flexible layers of thin metal tape or wire to achieve the required mass (*Pickering* [15.45, 46]). The difference in sound of an acoustic guitar strung with metal strings and the same instrument strung with more flexible gut or over-wound strings is illustrated in CD-ROM.

15.2.2 Nonlinear String Vibrations

Large-amplitude transverse string vibrations can result in significant stretching of the string giving a time-varying component in the tension proportional to the square of the periodically varying string displacement. This leads to a number of nonlinear effects of considerable scientific interest, though rarely of musical importance.

Morse and *Ingard* [15.42] and (*Fletcher* and *Rossing* [15.5], Chap. 5) provide theoretical introductions to the physics of nonlinear resonant systems and to nonlinear string vibrations in particular. *Vallette* [15.47] has recently reviewed the nonlinear physics of both driven and freely vibrating strings.

The Nonlinear Wave Equation

Transverse displacements of a string result in a fractional increase of its length L by an amount $1/L_0 \int_0^L 1/2(\partial\xi/\partial x)^2 \, dx$ and hence to a similar fractional increase in tension and related frequency of excited modes. For a spatially varying sinusoidal wave, the induced strain and hence tension will vary with both position and time along the string. Any spatially localised changes in the tension will propagate along the string with the speed of longitudinal waves. As this is typically an order of magnitude larger than for transverse waves, $c_L/c_T \sim \sqrt{L/\Delta L}$, where ΔL is the amount that the string is stretched to bring it to tension, such perturbations will propagate backwards and forwards along the string many times during a single cycle of the transverse waves. Hence, as pointed out by *Morse* and *Ingard* [15.42], to a rather good approximation, transverse wave propagation is determined by the spatially averaged perturbation of the tension.

Amazon
EX1013 Page 572

Consider a stretched string vibrating with large amplitude in its fundamental mode with transverse displacement $u = a\sin(\pi x/L)\cos\omega t$. The spatially averaged increase in tension is given by

$$\left(1+\frac{\pi^2}{4}\frac{a^2}{\Delta LL}\cos^2\omega_1 t\right) = \left[1+\beta\left(1-\cos 2\omega_1 t\right)a^2\right], \quad (15.48)$$

where $\beta = \frac{\pi^2}{8}\frac{1}{L\Delta L}$. Inserting this change in tension into the equation of motion for transverse string vibrations coplanar with a localised external driving force $f(t)$, we can write

$$\frac{\partial^2 u}{\partial t^2}+\frac{\omega_1}{Q}\frac{\partial u}{\partial t}+\omega_1^2\left[1+\beta\left(1-\cos 2\omega_1 t\right)a^2\right]u = \frac{2}{m}f(t)\,. \quad (15.49)$$

Mode Conversion

Nonlinearity results in an increase in the static tension by the factor $(1+\beta a^2)$ and hence an increase in frequency of all the string modes. In addition, the term varying as $\cos 2\omega_1 t$, at double the frequency of the principal mode excited, will interact with any other modes present to excite additional with frequencies $f_n \pm 2f_1$. Of special note is the effect of this term on the principal mode of vibration itself, exciting a new mode at three times the fundamental frequency $3f_1$ and an additional parametric term (acting on itself) from the $f_1 - 2f_1 = -f_1$ contribution. The parametric term causes an additional increase in frequency of the principal mode excited, so that in total

$$\omega_{1*}^2 = \omega_1^2\left(1+\frac{3}{2}\beta a^2\right), \quad (15.50)$$

where ω_1 is the small-amplitude resonant frequency.

Nonlinear effects depend on the square of the amplitude of the strongly excited mode and inversely on the amount by which the string has been stretched to bring it to tension. To investigate nonlinear effects, it is therefore advantageous to use weakly stretched strings at low initial tension. Conversely, because the tension of strings on musical instruments tends to be rather high, nonlinear effects are not in general important within a musical context.

Figure 15.28 shows measurements by *Legge* and *Fletcher* [15.48], which illustrate the nonlinear excitation and subsequent decay of the third partial of a guitar string plucked one third of the way along its length, so that the third partial was initially absent.

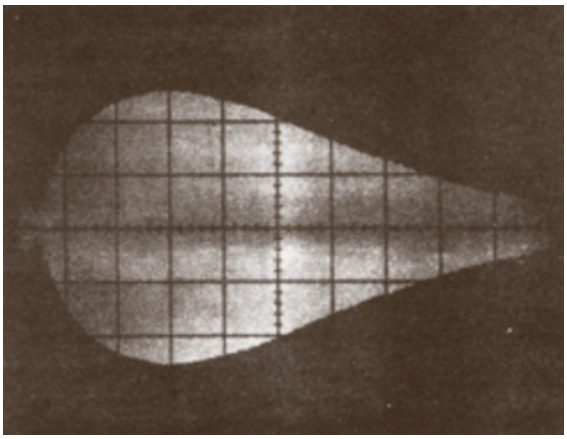


Fig. 15.28 Nonlinear excitation of the third partial of a stretched string plucked 1/3 of the way along its length; the graticule divisions are 50 ms apart (after *Legge* and *Fletcher* [15.48])

In general, bowed, plucked and struck waveforms have many Fourier components, each of which will contribute a term proportional to a_n^2 to the nonlinear increase in tension. However, in most cases, the fundamental will be the most strongly excited mode and will therefore dominate the nonlinearity.

The inharmonicity and changes in frequency associated with nonlinearity at large amplitudes can give a strongly plucked string an initial rather *twangy* sound. Nonlinear effects can also raise the frequency of a very strongly bowed open C-string of a cello by almost a semitone. However, under normal playing conditions, nonlinearity is rarely musically significant, at least in comparison with other more important perturbations of string vibrations, such as their interaction with the acoustically important structural resonances of an instrument, to be considered later.

Nonlinear Resonances

The nonlinear increase in frequency of modes with increasing amplitude leads to string resonances, which become increasingly skewed towards higher frequencies at large amplitudes, as illustrated in Fig. 15.29. For sufficiently large amplitudes and small damping, the resonance curves develop an overhang. On sweeping through resonance from the low-frequency side, the amplitude rises causing the resonance frequency to shift to higher frequencies, as indicated by the dashed line in Fig. 15.29a. Damping eventually leads to a sudden collapse, with the amplitude

Amazon
EX1013 Page 573

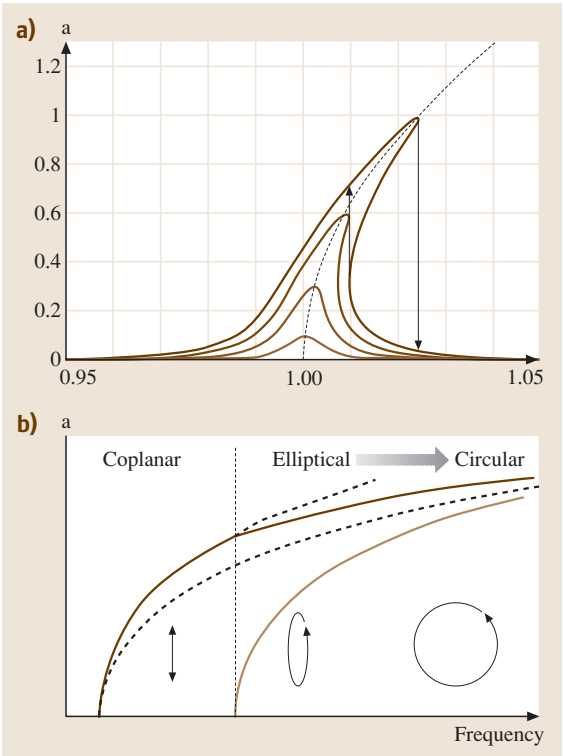


Fig. 15.29 **(a)** The effect of nonlinearity on the resonance curves of a stretched string with a Q of 100, for increasing drive excitation plotted against normalised resonant frequency. The *dashed curve* represents the nonlinear amplitude of the frequency for free decay. Note the hysteretic transitions at large amplitude; **(b)** the transition at large amplitudes from linearly polarised vibrations coplanar with the driving force to elliptical and finally circular orbital motion of the string at very large amplitudes. The two *continuous curves* represent the induced amplitudes in the directions parallel and perpendicular to the driving force

dropping to a much lower high-frequency value, illustrated by the downward arrow. On decreasing the frequency, the response initially remains on the low-amplitude curve before making a sudden hysteretic transition back to the large amplitude, strongly nonlinear, regime.

This behaviour is characteristic of any nonlinear oscillator with a restoring force that increases in strength on increasing amplitude. For a spring constant that softens with increases displacement, as we will discuss later in relation to Chinese gongs, the resonance curves are skewed in the opposite direction.

Orbital Motion

Nonlinearity results in another surprising effect on the driven resonant response. At sufficiently large amplitudes of vibration, a sinusoidally driven string suddenly develops motion in a direction orthogonal to and in phase-quadrature with the driving force, illustrated schematically in Fig. 15.29b. The transverse displacements then execute elliptical orbits about the central axis approaching circular motion at very large amplitudes (*Miles* [15.50]). In this limit, the string is under constant increased tension, producing an amplitude-dependent inward force balancing the centrifugal force of the orbiting string, resulting in an amplitude-dependent orbital frequency $\omega_{1*}^2 = \omega_1^2(1+2\beta a^2)$. For circular motion, the extension of the string and hence the increase in tension and resonant frequency are determined by the orbital radius of the whirling string, so there is now no variation in tension with time. The sense of clockwise or anticlockwise rotation is determined by chance or in practice by slight geometrical or material anisotropies of the string or supporting structure.

Such transitions have been investigated by *Hanson* and coworkers [15.51, 52] using a brass harpsichord string stretched to playing tension. The transition from linear to elliptically polarised motion was observed in addition to chaotic behaviour at very large amplitude. However, their measurements were complicated by the very long time constants predicted to reach equilibrium behaviour close to the transitional region and to rather strong and not well-understood splitting of the degeneracy of the transverse modes, even at low amplitudes when nonlinearity is unimportant.

A related effect occurs when a string is plucked so that it is given some orbital motion, as is invariably the case when plucking a string on a stringed instrument such as the guitar. Nonlinearity introduces

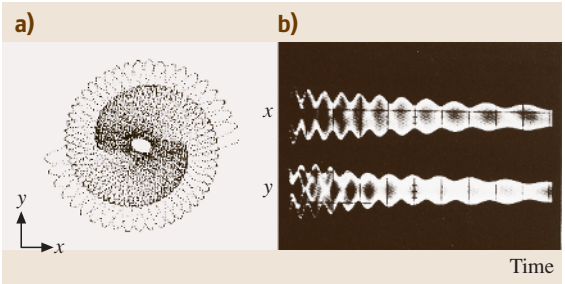


Fig. 15.30 **(a)** The computed precession of the damped elliptical orbits of a strongly plucked string, and **(b)** measurements of the orthogonal transverse components of such motion for a string plucked close to its mid-point (after [15.49])

Amazon
EX1013 Page 574

coupling between motions in orthogonal transverse directions, causing the orbits to precess (*Elliot* [15.53], *Gough* [15.49], *Villagier* [15.47]), as illustrated by computational simulations and measurements in Fig. 15.30. The precessional frequency Ω is given by $\frac{\Omega}{\omega} = \frac{ab}{L\Delta L}$, where a and b are the major and minor semi-axes of the orbital motion and ΔL is the amount by which the string is stretched to bring it to tension [15.49].

Such precession can lead to the rattling of the string against the fingerboard on a strongly plucked instrument, as the major axis of the orbiting string precesses towards the fingerboard. The nonlinear origin of such effects can easily by distinguished from other linear effects causing degeneracy of the string modes and hence beats in the measured waveform, by the very strong dependence of the precession rate on amplitude, as illustrated in Fig. 15.30b.

15.2.3 The Bowed String

Realistic Models

Although the main features of the bowed string can be described by a simple Helmholtz wave, it is important to consider how such waves are excited and maintained by the frictional forces between the bow and string. The simple Helmholtz solution is clearly incomplete for a number of reasons including:

1. the unphysical nature of infinitely sharp kinks,
2. the insensitivity of the Helmholtz bowed waveform to the position and pressure of the bow on the string. In particular, the simple Helmholtz waveform involves partials with amplitudes proportional to $1/n$, whereas such partials must be absent if the string is bowed at any integer multiple of the fraction $1/n$ along its length, since energy cannot be transferred from the bow to the string at a nodal position of a partial,
3. the neglect of frictional forces in the slipping regime,
4. the neglect of losses and reaction from mechanical coupling to structural modes at the supporting bridge,
5. the excitation of the string via its surface, which must involve the excitation of additional torsional modes.

Understanding the detailed mechanics of the strongly nonlinear coupling between the bow and string has been a very active area of research over the last few decades, with major advances in our understanding made possible by the advent of the computer and the ability to simulate the problem using fast computational methods. *Cremer* ([15.29], Sects. 3–8) provides a detailed account of many of the important ideas and techniques used to investigate the dynamics of the bowed string. In addition, *Hutchins* and *Benade* ([15.27], Vol. 1), includes a useful introduction to both historical and recent research prefacing 20 reprinted research papers on the bowed string. *Woodhouse* and *Galluzzo* [15.54] have recently reviewed present understanding of the bowed string.

Pressure, Speed and Position Dependence

In the early part of the 20th century, *Raman* [15.35], later to be awarded the Nobel prize for his research on optoacoustic spectroscopy, confirmed and extended many of Helmholtz's earlier measurements and theoretical models of the bowed string. Raman used an automated bowing machine to investigate systematically the effect of bow speed, position and pressure on bowed string waveforms. He also considered the attenuation of waves on the string and dissipation at the bridge. From both measurements and theoretical models, he showed that a minimum downward force was required to maintain the Helmholtz bowed waveforms on the string, which was proportional to bow speed and the square of bow distance from the bridge. He also measured and was able to explain the wolf-note phenomenon, which occurs when the pitch of a bowed note coincides with an over-strongly coupled mechanical resonance of the supporting structure. At such a coincidence, it is almost impossible for the player to maintain a steady bowed note, which tends to *stutter* and jump in a quasi-periodic way to the note an octave above, illustrated previously for a cello with a bad wolf note, audio CD-ROM.

Saunders [15.55], well known for his work in atomic spectroscopy (Russel–Saunders spin-orbit coupling) was a keen violinist and a cofounder of the Catgut Acoustical Society. He showed that, for any given distance of the bow from the bridge, there was both a minimum and a maximum bow pressure required for the Helmholtz kink to trigger a clean transition from the sticking to slipping regimes and vice versa. Subsequently, *Schelling* [15.56] derived explicit formulae for these pressures in terms of the downward bow force F as a function of bow speed v_B, assuming a simple model for friction between bow hair and string in the slipping region of $\mu_d F$ and a maximum sticking force of $\mu_s F$,

$$F_{min} = \frac{R_0^2 v_B}{2R\beta^2 (\mu_s - \mu_d)} \quad \text{and}$$
$$F_{max} = \frac{2R_0 v_B}{\beta (\mu_s - \mu_d)} = 4\beta \frac{R}{R_0} F_{min} , \qquad (15.51)$$

where R_0 is the characteristic impedance of the string terminated by a purely resistive load R at the bridge, and

Amazon
EX1013 Page 575

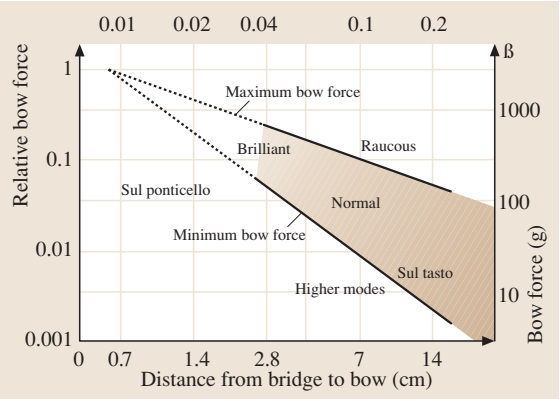


Fig. 15.31 The playing range for a bowed string as a function of bow force and distance from bridge, with the bottom and right-hand axis giving values for a cello open A-string with a constant bow velocity of 20 cms^{-1} (after *Schelling* [15.56])

β is the fractional bowing point along the string. If the downwards force is larger than $F_{\max}$ the string remains stuck to the string instead of springing free into the slipping regime, while for downward forces less than $F_{\min}$ an additional slip occurs leading to a double-slipping motion.

Figure 15.31 is taken from the article by *Schelling* on the bowed string in the Scientific American special issue on the *Physics of Musical Instruments* [15.57]. It shows how the sound produced by a bowed cello string changes with bow position and downward bow pressure for a typical bow speed of 20 cm/s. Note the logarithmic scales on both axes. In practice, a string can be bowed over a quite a large range of distances from the bridge, bow speeds and pressures with relatively little change in the frequency dependence of the spectrum and quality of the sound of an instrument, apart from regions very close and very distant from the bridge. Nevertheless, the ability to adjust the bow pressure, speed and distance from the bridge, to produce a good-quality steady tone, is one of the major factors that distinguish an experienced performer from the beginner.

Slip–Stick Friction

An important advance was the use of a more realistic frictional force, dependent on the relative velocity between bow and string, shown schematically for three downward bow pressures in Fig. 15.32. Such a dependence was subsequently observed by *Schumacher* [15.58] in measurements of steady-state sliding friction between a string and a uniformly moving bow.

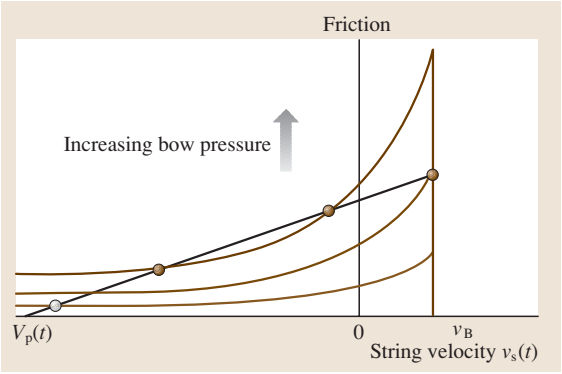


Fig. 15.32 Schematic representation of the dependence of the frictional force between bow and string on their relative velocity and downward pressure of the bow on the string. The *straight line* with slope $2R_0$ passes through the velocity v_p of the string determined by its past history and the intersection with the friction curves determines its current velocity. The *open circle* represents the single intersection in the slipping regime at low bow pressures, while the *closed circles* illustrate three intersections at higher pressures

The frictional force is proportional to the downward bow pressure.

Friedlander [15.59] showed that a simple graphical construction could be used to compute the instantaneous velocity v at the bowing point from the velocity $v_p(t)$ at the bowing point induced by the previous action of the bow. The new velocity is given by the intersection of a straight line with slope $2R_0$ drawn through v_p with the friction curve, where R_0 is the characteristic string impedance. This follows because the localised force between the bow and string generate secondary waves with velocity $F/2Z_0$ at the bowing point as previously described (15.34). In the slipping region well away from capture, there will be just a single point of intersection, so the problem is well defined. However, close to capture, as illustrated by the intersections marked by the black dots with the upper frictional curve, the straight line can intersect in three points (two in the slipping regime and one in the sticking regime) as first noted by Friedlander.

Computational Models

This model has been used in a number of detailed computational investigations of both the transient and steady-state dynamics of the bowed string, notably by the Cambridge group lead by *McIntyre* and *Woodhouse* [15.60–62], their close collaborator *Schumacher* [15.58, 63] from Carnegie-Mellon, and

Amazon
EX1013 Page 576

Guettler [15.64], who is also a leading international double-bass virtuoso. Readers are directed to the original publications for details of the various computational schemes used, which are also discussed in some detail by *Cremer* ([15.29], Sect. 8.2).

Whenever a string is bowed at an integer interval along its length, the secondary waves excited by the frictional forces between bow and string can give rise to coherent reflections between the bow and bridge, giving rise to pronounced *Schelling ripples* on the Helmholtz waveform and hence significant changes in the spectrum of the radiated sound. However, because the bowing force tends to be distributed across the ≈1 cm width of the bow hairs, such effects tend to be smeared out and are not generally of significant musical importance. *McIntyre* et al. [15.62] have also shown that uncertainties in the sticking point from the finite-width strand of bow hairs leads to a certain amount of jitter or aperiodicity in the pitch of the bowed string amounting of a few cents, which is again of little musical significance, though the noise generated may be significant in contributing to the characteristic sound of bowed string instruments.

It is instructive to consider the kind of computational methods developed by Woodhouse and his collaborators to investigate both the initial transient and the steady-state dynamics of the bowed string. This is illustrated schematically in Fig. 15.33, where u and u' represent the velocity under the bow from waves travelling towards the bow from the bridge and from the stopped end of the string respectively, and v and v' are the velocities at the bowing point of the waves travelling away from the bow. In the absence of any bowing force $v = u'$ and $v' = u$. However, in the presence of a frictional force between the bow and string, the outgoing waves will acquire an additional velocity $f/2R_0$, where the frictional force is determined by the velocity from the incoming waves $u + u'$ excited by previous events. The outgoing wave travelling towards the stopped end or nut of the string will simply be reflected, while the outgoing wave reaching the bridge will not only be reflected, but will also excite continuing vibrations at the bridge from the excitation of the coupled structural modes.

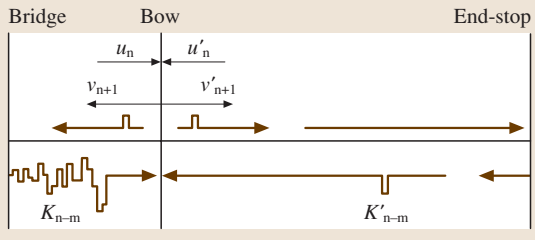


Fig. 15.33 Schematic representation of the model used by *McIntyre* and *Woodhouse* to compute bowed string dynamics. The velocities u and v represent incoming and outgoing waves from the two ends, with reflections of impulse functions from the bridge and end-stop represented in their digitised form

Such problems can be solved using a Green's function approach *Cremer* ([15.29], Sect. 8.4), in which the outgoing waves can be considered in terms of the response to forces represented as a succession of short impulses. The problem is then reduced to understanding the response of the system for the reflection of a sequence of short impulses or δ functions. At the end-stop, an impulse will simply be reflected with reversed sign, but reduced amplitude in the case of a soft finger stopping the string. The incoming wave u' generated by the reflected impulse will therefore be an impulse function delayed in time by the transit time from the bow to the end-stop and back. Similarly, the impulse returning from the bridge will be an impulse delayed by the transit time between bow and bridge and back followed by a wave generated by the induced motions of the bridge on reflection. The time-delayed impulse responses from reflections at the bridge and end-stop can described by the functions $K(t)$ and $K'(t)$. The incoming waves $(u(t), u'(t))$ can then be described by the convolution of $K(t)$ and $K'(t)$ with the outgoing waves $(v(t), v'(t))$ considered as a succession of impulse functions at all previous times t', such that

$$u(t) = \int^{t} v(t')K(t-t')\,\mathrm{d}t' \quad \text{and}$$

$$u'(t) = \int^{t} v'(t')K'(t-t')\,\mathrm{d}t\,. \tag{15.52}$$

To compute the resulting dynamics of string motion digitally, one simply computes the above velocities at a succession of short time intervals, with the outgoing waves determined from the incoming waves plus the secondary waves induced by the resulting frictional force, such that

$$v'_{n+1} = u_n + f_n/2R_0 \quad \text{and}$$
$$v_{n+1} = u'_n + f_n/2R_0 \tag{15.53}$$

and

$$u_{n+1} = \sum^{n} v_m K_{n-m} \quad \text{and}$$

$$u'_{n+1} = \sum^{n} v'_m K'_{n-m}\,, \tag{15.54}$$

where K_j and K'_j are now the digital equivalents of the time-delayed impulse responses, illustrated schemat-

Amazon
EX1013 Page 577

ically in Fig. 15.33. The frictional force f_n entering (15.53) is evaluated from the pressure- and velocity-dependent frictional force using the Friedlander construction with the computed string velocity under the bow given by $u_n + u'_n$.

Pressure Broadening and Flattening

As an example, Fig. 15.34 illustrates the computed velocity of the string under the bow as a function of increasing bow pressure (*McIntyre* et al. [15.60]). In contrast to the rectangular waveform predicted by the simple Raman model, the waveform is considerably rounded, especially at low bow pressures. This results in a less strident, less intense sound, with the higher partials strongly attenuated. At higher pressures, but at the same position and with the same bow velocity, the rounding is less pronounced, so that higher partials become increasingly important. The increased intensity of the higher partials leads to an increased perceived intensity with bow pressure, in contrast to the Raman model, in which the waveform and hence intensity remains independent of bow pressure. This is referred to as the *pressure effect*. At even higher pressures, the ambiguity in intersections noted by Friedlander leads to a pronounced increase in the capture period and hence the pitch of the bowed note, known as the *flattening effect*. These features are discussed in considerable detail along with his own important research and that of his collaborators on such effects by *Cremer* ([15.29], Chaps. 7 and 8).

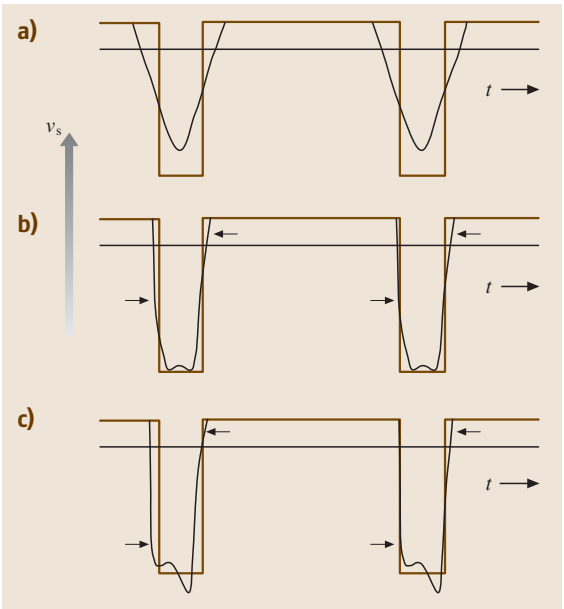


Fig. 15.34 Computed velocity of string at bowing point for increasing bow pressures in the ratios 0.4 : 3 : 5 (after *McIntyre* and *Woodhouse* [15.60]) illustrating both the broadened waveform and pitch dependence on bow pressure compared with the idealised rectangular Helmholtz bowed waveform

Initial Transients

Computational models can also describe the initial transients of the bowed string before the steady-state Helmholtz wave is established. Figure 15.35 compares the computed and measured initial transients of the string velocity under the bow for a string played with a sharp attack (a martelé stroke) (*McIntyre*, *Woodhouse* [15.60]). These computations also include the additional excitation of torsional waves, which are excited because the bowing force acts on the outer diameter of the wire, exerting a couple in addition to a transverse force. The excitation and loss of energy to the torsional waves appears to encourage the rapid stabilisation of the bowed Helmholtz waveform.

For low-pitched stringed instruments such as the double bass, it is very important that the Raman bowed waveform is established very quickly, otherwise there will be a significant delay in establishing the required pitch. Remarkably, *Guettler* [15.64] has shown that, by simultaneously controlling both bow speed and downward pressure, the player can establish a regular Raman waveform in a single period. The speed with which a steady-state bowed note can be established can be represented on a *Guettler diagram*, where the number of slips before a steady-state Helmholtz motion is established can be illustrated as a two-dimensional histogram as a function of bowing force and acceleration of the bow speed from zero.

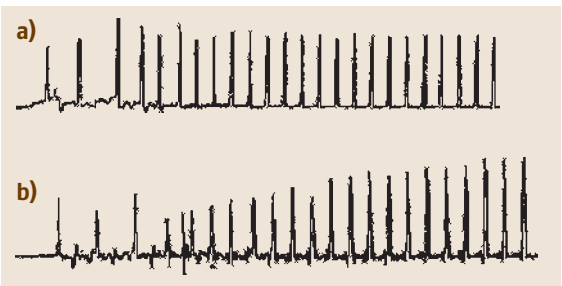


Fig. 15.35 **(a)** Computed transient string velocity at the bowing point for a strongly bowed string including coupling to both transverse and torsional modes and **(b)** the measured string velocity for a strongly played martelé bow stroke (after *McIntyre and Woodhouse* [15.60])

Amazon
EX1013 Page 578

To investigate such effects experimentally, *Galluzzo* and *Woodhouse* [15.54, 65] have recently developed a dynamically controlled bowing machine with active feedback, providing programmable control of both downward bow pressure and bow speed. This enables reliable and reproducible results to be made over a very wide range of possible playing parameters, extending Guettler's original measurements.

Viscoelastic Friction

Recent measurements have shown that the frictional model assumed in these investigations is over-simplistic. The force between the bow hairs and the string is maintained by a thin layer of rosin which coats them both. Rosin is a rather soft, sticky substance, with a glass-to-liquid transition not far above room temperature, resulting in viscoelastic properties, which are very sensitive to temperature (*Smith, Woodhouse* [15.66]). As the bow slides past the bow hair, the frictional forces will heat the rosin and hence reduce its viscoelasticity frictional properties. During the sticking regime, with no work being done at the bow–string interface, the rosin will cool down and the friction will increase. The frictional forces are therefore hysteretic and will be strongly dependent on past history within a given period of string vibration. *Woodhouse* et al. [15.68] and *Smith* [15.69] have investigated this hysteretic behaviour in some detail using rosin-coated glass rods. The hysteretic properties shown in Fig. 15.36 were deduced from measurements at the two supported ends of the string. *Woodhouse* [15.70] subsequently extended his computational models to incorporate the hysteretic frictional properties. Somewhat surprisingly, this more realistic model made little qualitative difference to the predicted behaviour. Such measurements contribute to our understanding of the physical processes underlying viscoelastic properties of various coatings and lubricants and have become an important tool in the field of tribology (studies of friction).

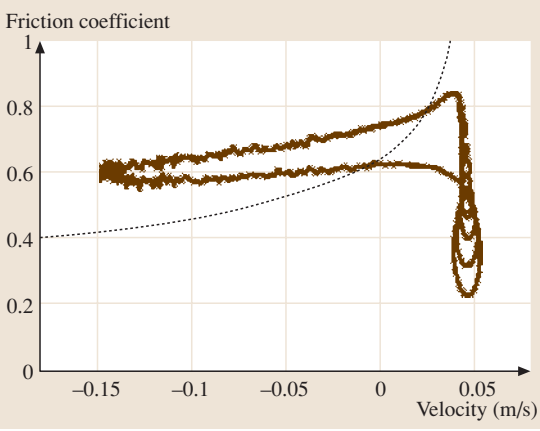


Fig. 15.36 Measured hysteretic frictional force between string and a glass bow coated with rosin, with the *dashed line* indicating previously assumed velocity dependence (after *Smith* and *Woodhouse* [15.66])

15.2.4 Bridge and Soundpost

We now consider the role of the bridge and soundpost in providing the coupling between the vibrating strings and the vibrational modes of the body of instruments of the violin family. We also consider the influence of such coupling on the modes of string vibration, which involves a discussion of the very important influence of damping on the normal modes of any coupled system.

Bridges

Many plucked and struck stringed instruments, such as the piano or guitar, use a rather low solid bridge to support the strings and transfer energy directly from the transverse string vibrations perpendicular to the supporting soundboard or front-plate of the instrument. The bridge needs only to be sufficiently high to prevent the strings from vibrating against the fingerboard or shell of the instrument. This is also true for the Chinese two-string violin, the *erhu*, which is held and played so that the bow excites string vibrations perpendicular rather than parallel to the stretched snake-skin membrane supporting the bridge and strings. The strings of a harp are attached to an angled sounding board, so that transverse string vibrations in the plane of the strings couple directly to the perpendicular vibrations of the supporting soundbox *Fletcher* and *Rossing* ([15.5], Sect. 11.2).

For such instruments, the bridge and other string terminations play a relatively insignificant acoustic role,

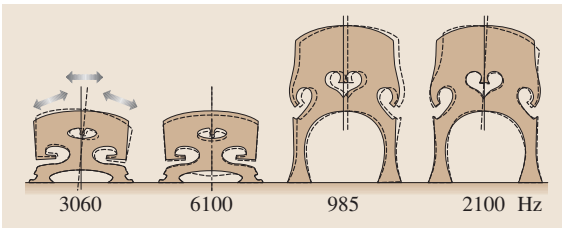


Fig. 15.37 The lowest in-plane resonant modes and frequencies of violin and cello bridges (after *Reinicke* [15.67]). The *arrows* represent the vibrational directions of the bowed outer and middle strings

Amazon
EX1013 Page 579

apart from adding a small inertial mass and additional stiffness to the soundboard or top plate, which only slightly perturbs the frequencies of the structural modes of vibration.

In contrast, the rather high bridges on instruments of the violin and viol families have a profound influence on the acoustical properties, particularly at frequencies comparable with and above any mechanical resonances of such structures. Figure 15.37 illustrates the shape of modern violin (and viola) and cello bridges and indicates their principal vibrational modes, as measured by *Reinicke* and *Cremer* [15.67, 71] using laser interference holography. Bridges are cut from maple and taper in thickness from the two feet to the top surface supporting the four strings, which are set in small v-shaped locating grooves.

Reinicke [15.67, 71] showed that the lowest violin bridge resonance at typically around 3 kHz involves a rotational motion of the top half of the bridge about its waist. The rotational motion induced by the vibrating strings supported on the top of the bridge results in a couple acting on the top plate via the two feet. The next most important in-plane resonance is at ≈ 6 kHz and involves the top of the bridge bouncing up and down on its feet, resulting in forces via the legs perpendicular to the supporting surface. The cello bridge has rather longer legs, resulting in two low-frequency twisting modes with resonances at around 1 and 2 kHz, both of which exert a couple on the top plate. Longitudinal forces from the vibrating strings can also induce bridge motion perpendicular to its plane (at double the frequency of the vibrating strings), but such motion is generally rather small and will be ignored for the purposes of this chapter.

Any transverse string force at the top of the bridge, from bowing, plucking or striking the string, will be transferred to the supporting body via the two feet. This will induce a linear motion of the centre of mass of the instrument, rotation about its centre of mass and the excitation of both flexural and longitudinal waves in the plates of the instrument. Because bowing involves a static force which reverses with bow direction, a bowed instrument has to be held fairly firmly by the player, which introduces an extra channel for energy loss through the supporting chin and fingers. The induced linear and rotational motions of an instrument are relatively unimportant at audio frequencies as they involve the whole mass M of the instrument with admittances varying $\sim 1/\mathrm{i}M\omega$.

If a tall bridge is placed centrally on a symmetric shell structure, like the body of an early renaissance viol, the plucked or bowed motion of the strings parallel to the supporting top plate would excite only asymmetrical modes of the supporting structure, whereas perpendicular string vibrations would excite only symmetrical modes. For instruments of the violin family, an offset soundpost is wedged between the front and back plates, which destroys the symmetry. The coupled modes will then involve a linear combination of symmetrical and asymmetric body modes, as discussed later (Sect. 15.2.6).

The arching of the top of the bridge allows each of the supported four strings to be bowed separately or together (double stopping), with the bow direction making an angle of around $\pm 15-20°$ relative to the top plate for the outer two strings and almost parallel for the middle two strings. Bowing on the outer two strings therefore involves significant perpendicular in addition to parallel forces, but only slightly different sounds from a single type of string when supported in different positions on the bridge. Audio CD-ROM compares the sound of a bowed covered-gut D-string mounted in the normal position and in the G-, A- and E-string positions on the same violin. On a guitar almost all the sound is produced by the vertical motion of the plucked string rather than by parallel vibrations, which primarily excite non-radiating longitudinal modes of the top plate.

Simplified Bridge Model

Cremer ([15.29], Chap. 9) gives a detailed historical and scientific introduction to research on violin and cello bridges and their coupling to the body of the instrument. Relatively complicated mechanical models are described composed of several masses and springs to account for the various possible vibrational modes of the bridge. However, the principal resonances of the violin bridge shown in Fig. 15.37 can be modelled very simply by a two-degree-of-freedom mechanical model, with effective masses representing the linear and rotational energy of the top of the bridge coupled to the supporting surface through the two supporting feet via a rotational or vertical spring, illustrated schematically in Fig. 15.38. The relatively light mass and added rigidity of the lower

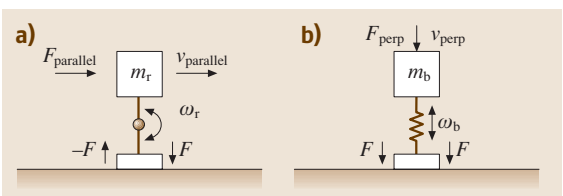


Fig. 15.38a,b Simplified mechanical models for the lowest **(a)** rotational and **(b)** bouncing motions of the bridge supported by its two feet on a rigid surface

Amazon
EX1013 Page 580

half of the bridge will only slightly perturb the resonant frequencies of the more massive supporting plates and can therefore be ignored as a first approximation. The effective masses and strength of the coupling springs can be chosen to reproduce the vibrational characteristics of the first two vibrational modes of the violin (or cello) bridge, which dominate the acoustical properties of the instrument.

At low frequencies, well below any resonant frequency, the bridge will vibrate as a rigid body, adding a small amount of additional mass, moment of inertia and rigidity to the top plate, which will again only slightly perturb the vibrational frequencies of the supporting shell structure. The additional relative height of the cello bridge compared with that of the violin bridge enables a rather larger couple to be exerted by the bowed string on the more massive top plate. There is a delicate balance between increasing the coupling to enhance the intensity at low frequencies without making it so strong that troublesome wolf-note problems arise, as referred to earlier.

Bridge-Hill (BH) Feature

Reinicke [15.67,71] and *Cremer* [15.29] highlighted the importance of the bridge resonance on both the sound of the violin and on admittance measurements, which are traditionally made by exciting the violin at the top of the bridge using an external force parallel to the top supporting plate. In recent year, this problem has attracted renewed interest, in an attempt to describe the rather broad peak and associated phase changes superimposed on the multi-resonant response of the instrument, which Jansson refers to as the *Bridge-Hill* (BH) feature [15.74,75].

Figure 15.39 shows recent measurements by *Woodhouse* [15.72] of the modulus of the admittance at the bridge for a particular instrument using a series of bridges with different masses but the same resonant frequency at ≈ 2 kHz. A strong but rather wide overall BH peak is observed in the vicinity of the bridge resonance. Note the marked decrease in admittance with increasing bridge mass above the bridge resonance. There is also an associated overall 90° change in the phase of the admittance on passing through the peak.

Evidence for the BH feature can also be seen in *Dünnwald*'s [15.73] superimposed measurements of the sound output of a large number of high-quality Italian, modern master and factory violins as a function of sinusoidal input force at top of the bridge, shown in

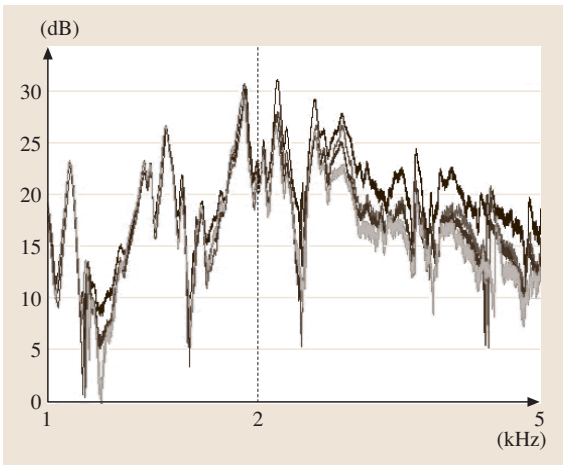


Fig. 15.39 The admittance at the top of the bridge on a single violin, plotted on the same but arbitrary dB scale, for a number of violin bridges having the same height and resonant frequency (≈ 2 kHz) but different masses. The upper curve corresponds to the lightest bridge that could be fabricated from a standard bridge blank and the lowest curve by the heaviest. Subtracting 45 dB from the results would give the approximate admittance in units of $ms^{-1}N^{-1}$ (data kindly provided by *Woodhouse* [15.72])

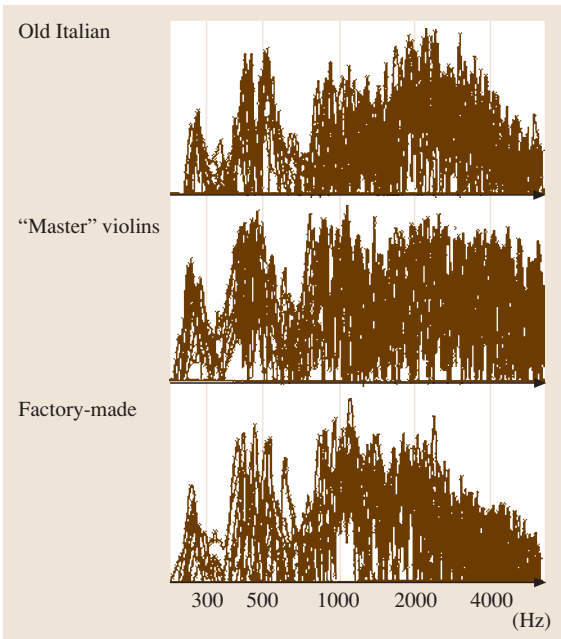


Fig. 15.40 Overlays of the sound output of 10 typical old Italian, modern master instruments, and 10 factory instruments for a constant sinusoidal force at the top of the bridge (after *Dünnwald* [15.73])

Amazon
EX1013 Page 581

Fig. 15.40. A surprising aspect of these measurements is the apparent lack of any such feature for modern master violins, possibly because of a wider variation in bridge resonances and effective masses of bridge and plate resonances in the chosen instruments. From measurements of the radiated sound of over 700 violins, Dünnwald proposed that the presence of a number of strong acoustic resonances in the broad frequency band from 1.5 to 4 kHz was one of the distinguishing features of a really fine instrument. The influence of the bridge in accounting for such a peak and the reduced response at higher frequencies is clearly important.

Woodhouse [15.76] has recently revisited the problem of the coupling between bridge and body of the instrument and the origin of the BH peak. A simple theoretical model shows that the peak depends on many factors, such as the effective masses, Q-values and resonant frequencies of the major vibrational modes of the bridge and the multi-resonant properties of the instrument. To demonstrate the overall effect of the bridge without having to consider the detailed vibrational response of a particular instrument, Woodhouse first considered coupling to a simplified model for the vibrational modes of the coupled instrument. This assumed a set of coupled vibrational modes each having the same effective mass M and Q-value, with a constant spacing of resonances $\omega_0 = 2\pi\Delta f$. Different values for these parameters would need to be used to model the independent rotational or bouncing modes, though Woodhouse concentrates on the influence of the lowest frequency "rocking" bridge mode. The merit of such a model is that the multi-resonant response of such a system varies monotonically with frequency. The features introduced by the resonant properties of the bridge can then be easily identified and the input admittance expressed relative to the admittance A_V for a completely rigid bridge of the same mass, where

$$A_V(\omega) = \frac{1}{M}\sum_n \frac{\mathrm{i}\omega}{(n\omega_0)^2 - \omega^2 + \mathrm{i}\omega n\omega_0/Q} \,. \tag{15.55}$$

The corresponding input admittance for the one-degree-of-freedom model bridge is then given by

$$A_{BB}(\omega) = \frac{A_V + \mathrm{i}\omega/m\omega_B^2}{1 - (\omega/\omega_B)^2 + \mathrm{i}\omega m A_V} \,, \tag{15.56}$$

where m is the effective mass of the bridge and ω_B its resonant frequency and internal damping of the bridge has been neglected.

We can also define a nonlocal admittance or mobility A_{VB} to describe the induced body motion per unit force

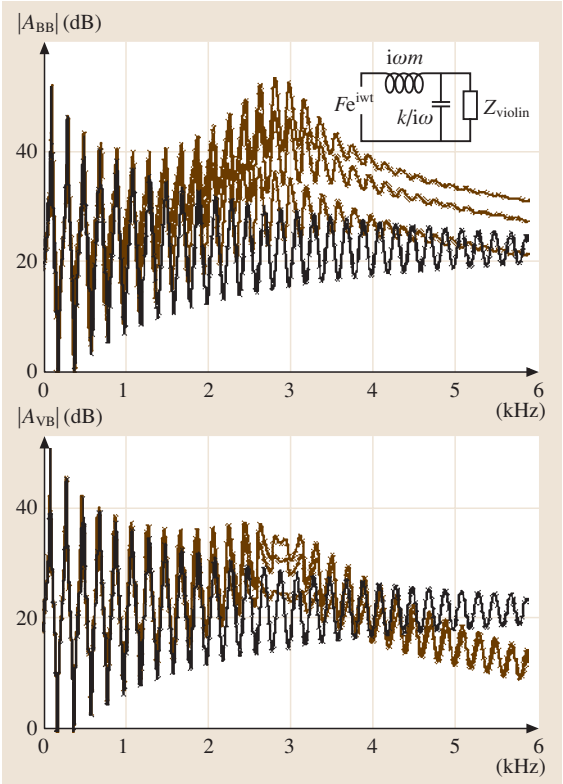


Fig. 15.41 Response curves for a one-degree-of-freedom bridge coupled to an artificial set of regularly spaced (200 Hz), constant effective mass (100 g) and constant Q (50) structural resonances. The *upper curves* illustrate the effect of bridge mass on the admittance A_{BB} measured at the point of excitation at the top of the bridge, while the *lower curves* illustrate the corresponding induced body mobility A_{VB}. The *coloured response curves* are for lossless bridges with effective masses 1, 1.5 and 3 g (highest to lowest response), having the same resonant frequency at 3 kH (after *Woodhouse* [15.76]) The *black curves* show the violin body response A_V that would be measured using a massless rigid bridge

at the foot of the bridge given by

$$A_{VB}(\omega) = \frac{A_V}{1 - (\omega/\omega_B)^2 + \mathrm{i}\omega m A_V} \,. \tag{15.57}$$

The simulations in Fig. 15.41 illustrate the major effect of the bridge resonance on both the input response and induced body motion and hence radiated sound at, around and above the resonant frequency of the bridge (3 kHz in the above example). For a real instrument, the spacing and Q-values of the individ-

Amazon
EX1013 Page 582

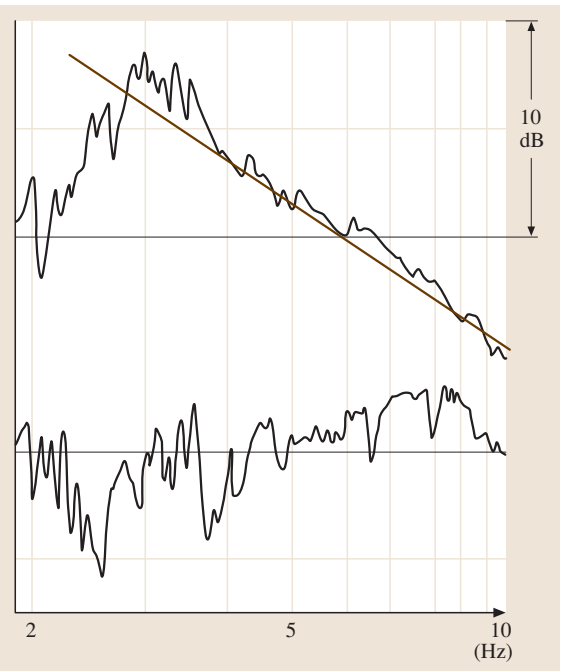


Fig. 15.42 Admittances of a violin measured at the top of the bridge (top trace) and at the left foot of the bridge (lower trace) illustrating a strong BH peak when measured at the top of the bridge but a relatively monotonic dependence of the body of the instrument (after *Cremer* [15.29], Fig. 12.9). The added *solid line* represents the $1/f$ reduction in the predicted BH response above the bridge resonance

ual modes will be very irregular and highly instrument dependent; nevertheless, the effect of the bridge resonance on the overall response will be very similar. In particular, the bridge resonance gives a broad peak in input admittance followed by a 6 dB/octave decrease in the admittance above resonance, where the response is largely dominated by the bridge dynamics rather than that of the instrument itself, with $A_{BB} \sim 1/im\omega$. Note that the height and width of the peak is largely determined by energy lost to the coupled structural vibrations (including, in practice, additional energy lost to all the supported strings) rather than from internal bridge losses, which have been neglected in this example.

The bridge resonance introduces a somewhat smaller peak in the induced body mobility and hence radiated sound. Well above the bridge resonance, the induced body velocity is given by $A_{VB}(\omega_B/\omega)^2$, with an intensity decreasing by 12 dB/octave. Unlike the input bridge admittance, the induced body motion and output sound retains the characteristic resonances of the instrument, though attenuated.

The predicted difference in admittance at the top of the bridge A_{BB} and top of the instrument A_V is illustrated in Fig. 15.42, in measurements by *Moral* and *Jansson* [15.77] reproduced by *Cremer* ([15.29], Fig. 15.9). Whereas the average admittance of the violin varies relatively little with frequency, the admittance at the bridge shows a pronounced BH peak with a relatively featureless and approximately $1/f$ (the added solid line) variation above the peak, as anticipated from the above model.

Woodhouse [15.76] has extended this idealised model to describe the coupling of the bridge to a more realistic, but still simplified, model for the vibrational modes of the violin with a soundpost. This changes the detailed response, but not the overall qualitative features. Because the response of a violin depends rather randomly at higher frequencies on the positions and Q-values of the structural modes, Woodhouse uses a logarithmic scale to average the peaks and troughs at the maxima and minima of the admittance (approximately proportional to Q and $1/Q$), to give a *skeleton curve* describing the global variation of the violin's complex admittance (more details are given in the later Sect. 15.2.3 on shell modes). This enables Woodhouse to illustrate the influence of various bridge parameters on the acoustical properties of the instrument, suggesting ways in which violin makers could vary bridge properties to optimise the sound quality of an instrument, though that will always be a matter of personal taste rather than being scientifically defined.

The important role of the bridge in controlling the sound of the violin or cello has often been overlooked, even by many skilled violin makers. Indeed one of the reasons why Cremonese violins generally produce such highly valued sounds is the experience and skill involved in adjusting the mass, size and fitting of the bridge (and the position of the soundpost) to optimize the sound quality, investigated experimentally by Hacklinger [15.78].

Added Mass and Muting

A familiar demonstration of the importance of the mass of the bridge on the sound of an instrument is to place a light mass or mute on the top of the bridge. This dramatically softens the tone of the instrument by decreasing the resonant frequency of the bridge and hence amplitude of the higher-frequency components in the spectrum of sound. The added mass Δm lowers the resonant frequency ω_B by a factor $[m/(m+\Delta m)]^{1/2}$.

Amazon
EX1013 Page 583

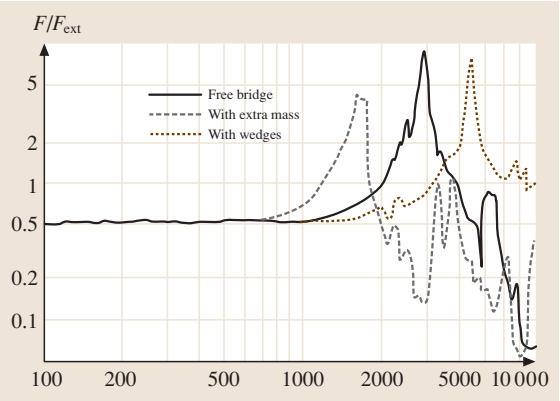


Fig. 15.43 Measurements of bridge resonances from measurements of the ratio of the force exerted by one bridge foot on a rigid surface to the applied force, for an added mass of 1.5 g and for wedges introduced between the *wings* of the bridge to increase its rotational stiffness (after *Reinicke* data reproduced in *Fletcher* and *Rossing* [15.5], Fig. 10.19)

Figure 15.43 illustrates changes in resonant frequency measured by Reinicke for a bridge mounted on a rigid support for an additional mass of 1.5 g and when wedges are inserted between the *wings* of the bridge to inhibit the rotational motion of the top of the bridge and hence the resonant frequency. Audio CD-ROM illustrates the changes in sound of a violin before and after first placing a commonly used 1.8 g mute and then a much heavier practice mute on top of the bridge, and after wedges were inserted in the bridge to inhibit the rocking motion.

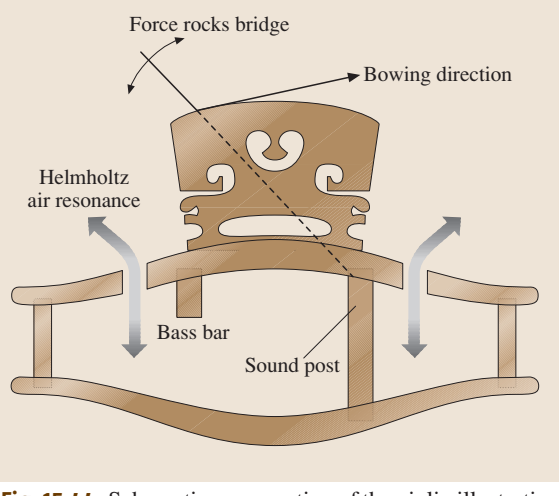


Fig. 15.44 Schematic cross section of the violin illustrating the position of the soundpost, bass-bar and f-hole openings

Soundpost and Bass Bar

In instruments of the violin family, a soundpost is wedged asymmetrically between the top and back plates, as illustrated schematically in Fig. 15.44. Additionally, a bass bar runs longitudinally along much of the length of the bass-side of the front plate. The soundpost and bass bar give added mechanical strength to the instrument, helping it to withstand the rather large downward force from the angled stretched strings passing over the bridge, which is typically ≈ 10 kg weight for the violin.

The influence of the soundpost on the quality of sound is so strong that the French refer to it as the *âme* (soul) of the instrument. Its acoustic function is to provide a rather direct coupling of the induced bridge vibrations to both the back and the front plates of the instrument and to provide an additional mechanical constraint, so that the bowed string vibrations excite normal modes, which are linear combinations of the asymmetric and symmetric modes of vibration of the front and back plates of the instrument.

Of modern stringed instruments, only the violin family makes use of a soundpost. However, soundposts were probably used in the medieval fiddle and other early instruments including the viol. The ancient Celtic *crwyth* effectively combined the functions of the bridge and soundpost by using a bridge with feet of unequal length, the first resting on the top plate and the second passing through a hole in the front face to rest on the back plate – a bridge design still used today in the folk-style Greek *rebec* (see *Gill* [15.79]).

15.2.5 String–Bridge–Body Coupling

We now consider the interaction of the strings with the vibrational modes of the body of the instrument via the bridge. Because we are dealing with the coupling of the vibrational modes of the strings, bridge and body of the instrument, the problem has to be considered in terms of the normal modes of the coupled system. An important aspect of this problem that is often not widely recognised, but is always important in dealing with musical instruments, is the profound influence of damping on the nature of the coupled modes. This is a generic phenomenon for any system of coupled oscillators. As we will see, the strength of the damping relative to the strength of the coupling determines whether a system can be considered as weakly or strongly coupled.

String–Body Mode Coupling

For simplicity, we only consider the perturbation of string resonances from the induced motion of the bridge

Amazon
EX1013 Page 584

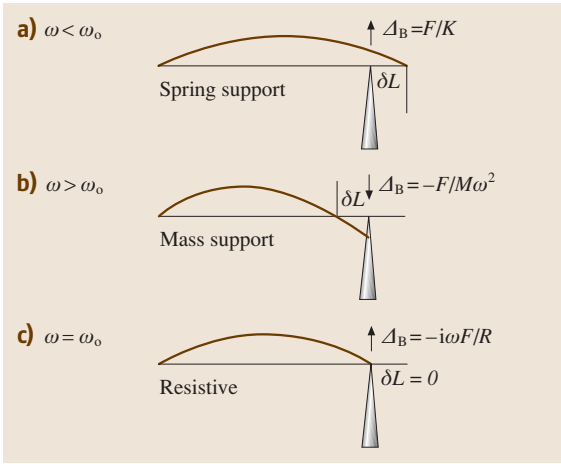


Fig. 15.45a–c Coupling of vibrating string to a weakly coupled normal mode via the bridge for the string resonance **(a)** below the resonant frequency of the coupled mode, **(b)** above the resonant frequency, and **(c)** at resonance

and ignore any damping introduced by, for example, a finger stopping the string at its opposite end. In general, as we have already seen, the admittance at the point of string support on the bridge will be a complicated, multi-resonant, function of frequency reflecting the normal modes of vibration of the coupled structure. The normal modes will include the combined motions of all parts of the violin body, including the body, bridge, neck, tailpiece, etc.

Each coupled normal mode will contribute a characteristic admittance, which will be spring-like below its resonant frequency, resistive at resonance and mass-like above resonance. The effect of such terminations on the vibrating string is therefore to shift its effective nodal position, as illustrated in Fig. 15.45a–c, for a spring-like string termination with spring constant K, an effective mass M and a lossy support with resistance R.

For a spring-like termination with spring constant K, the bridge will move in phase with the force acting on it. This will increase the effective length of the vibrating string between nodes by a distance $\Delta_B = T/K$, lowering the frequency of a string mode by a fraction T/KL. For a mass-like termination M, the end-support will move in anti-phase with the forces acting on it, so that the effective string length is shortened. The string frequencies are then increased by the fraction $T/M\omega_n^2 = (1/n\pi)^2 m/M$, where m is the mass of the string. It is less easy to visualise the effect of a resistive support because the induced displacement is in phase-quadrature with the driving force. Mathematically, however, a resistive termination can be considered as an imaginary mass $m^* = R/i\omega$ leading to an imaginary fractional increase in frequency $i\omega_n 1/(n\pi)^2 m/R$. This imaginary frequency is equivalent to an exponential decay $e^{-t/\tau}$ for all modes with $\tau = \pi^2 R/m\omega_1^2$, where ω_1 is the frequency of the fundamental string mode. This result can also be derived using somewhat more physical arguments, by equating the loss of stored vibrational energy to the energy dissipated at the end-support.

The terminating admittance at the bridge for a single coupled vibrational mode can be written in the form

$$A_n(\omega) = \frac{1}{M_n} \frac{i\omega}{\omega_n^2 - \omega^2 + i\omega\omega_n/Q_n}\,, \qquad (15.58)$$

with the real part of this function determining the decay time of the coupled string resonances and the imaginary part the perturbation in their resonant frequencies. The perturbations are proportional to the ratio of mass of the vibrating string to the effective mass of the coupled resonance at the point of string support on the bridge and vary with frequency with the familiar dispersion and dissipation curves of a simple harmonic oscillator. For a multi-resonant system like the body of any stringed instrument, the string perturbations from each of the coupled structural resonances are additive.

Normal Modes and Damping

Strictly speaking, whenever one considers the coupling between any two or more vibrating systems, one should always consider the normal modes or coupled vibrations rather than treat the systems separately, as we have done above. However, the inclusion of damping has a profound influence on the normal modes of any system of coupled oscillators (*Gough* [15.80]) and justifies the above *weak-coupling approximation*, provided that the coupling at the bridge is not over-strong. Although we consider the effect of damping in the specific context of a lightly damped string coupled to a more strongly damped structural resonance, the following discussion is completely general and is applicable to the normal modes of any coupled system of damped resonators.

Consider a string vibrating in its fundamental mode coupled via the bridge to a single damped structural resonance. The string has mass m, an unperturbed resonant frequency of ω_s a Q-value of Q_s and a displacement at its mid-point of v. The coupled structural resonance has an effective mass M at the point of string support, an unperturbed resonant frequency of ω_M a Q-value of Q_M and displacement of u.

Amazon
EX1013 Page 585

The vibrating string exerts a force on the coupled body mode, such that

$$M\left(\frac{\partial^2 v}{\partial t^2}+\frac{\omega}{Q_\mathrm{m}}\frac{\partial v}{\partial t}+\omega_\mathrm{M}^2 v\right)=T\left(\frac{\pi}{L}\right)u\,. \qquad (15.59)$$

Multiplying this expression through by $\partial v/\partial t$, one recovers the required result that the rate of increase in stored kinetic and potential energy of the coupled mode is simply the work done on it by the vibrating string less the energy lost from damping. Similar energy balance arguments enable us to write down an equivalent expression for the influence of the coupling on the string vibrations,

$$\frac{m}{2}\left(\frac{\partial^2 u}{\partial t^2}+\frac{\omega}{Q_\mathrm{m}}\frac{\partial u}{\partial t}+\omega_\mathrm{M}^2 u\right)=T\left(\frac{\pi}{L}\right)v\,, \qquad (15.60)$$

where the effective mass of the vibrating string is $m/2$ (i. e. its energy is $1/4\,m\omega^2u^2$). To determine the normal-mode frequencies, we look for solutions varying as $\mathrm{e}^{\mathrm{i}\omega t}$. Solving the resultant simultaneous equations we obtain

$$\left(\omega_\mathrm{M}^2-\omega^2\left(1-\mathrm{i}/Q_\mathrm{M}\right)\right)\left(\omega_\mathrm{m}^2-\omega^2\left(1-\mathrm{i}/Q_\mathrm{m}\right)\right)$$
$$=\left(T\frac{\pi}{l}\right)^2\frac{2}{mM}=\alpha^4\,, \qquad (15.61)$$

where α is a measure of the coupling strength.

Solving to first order in $1/Q$-values and α^2 we obtain the frequencies of the normal modes $\Omega_\pm$ of the coupled system,

$$\Omega_\pm^2=\omega_+^2\pm\left(\omega_-^4+\alpha^4\right)^{1/2}\,, \qquad (15.62)$$

where

$$\omega_\pm^2=\frac{1}{2}\left[\omega_\mathrm{M}^2\pm\omega_\mathrm{m}^2+\mathrm{i}\left(\frac{\omega_\mathrm{M}^2}{Q_\mathrm{M}}\pm\frac{\omega_\mathrm{m}^2}{Q_\mathrm{m}}\right)\right]\,. \qquad (15.63)$$

If the damping terms are ignored, we recover the standard perturbation result with a splitting in the frequencies of the normal modes at the crossover frequency (when the uncoupled resonant frequencies of the two systems coincide) such that $\Omega_\pm^2=\omega_\mathrm{M}^2\pm\alpha^2$.

In the absence of damping, the two normal modes at the crossover frequency are linear combinations of the coupled modes vibrating either in or out of phase with each other, with equal energy in each, so that $v/u=\pm\sqrt{m/2M}$. Well away from the crossover region, the mutual coupling only slightly perturbs the individual coupled modes, which therefore retain their separate identities. However, close to the crossover region, when $|\omega_\mathrm{M}-\omega_\mathrm{m}|\leqslant 2\alpha^2/\left(\omega_\mathrm{M}+\omega_\mathrm{m}\right)$, the coupled modes lose their separate identities, with the normal modes involving a significant admixture of both.

The inclusion of damping significantly changes the above result. If we focus on the crossover region, coupling between the modes will be significant when

$$\alpha^2\sim\omega_\mathrm{M}^2-\omega_\mathrm{m}^2\,. \qquad (15.64)$$

At the crossing point, when the uncoupled resonances coincide, the frequencies of the coupled normal are given by

$$\Omega_\pm^2=\omega_\mathrm{M}^2\left(1+\mathrm{i}/2Q_+\right)\pm\left(\alpha^4-\left(\frac{\omega_\mathrm{M}^2}{2Q_-}\right)^2\right)^{1/2}\,, \qquad (15.65)$$

where

$$\frac{1}{Q_\pm}=\frac{1}{Q_\mathrm{M}}\pm\frac{1}{Q_\mathrm{m}}\,. \qquad (15.66)$$

The sign of the terms under the square root clearly depends on the relative strengths of the coupling and damping terms. When the damping is large and the coupling is weak, such that $\left(\omega_\mathrm{M}^2/2Q_-\right)^2>\alpha^4$, one is in the *weak-coupling* regime, with no splitting in frequency of the modes in the crossover region. In contrast, when the coupling is strong and the damping is weak, such that $(\omega_\mathrm{M}^2/2Q_-)^2<\alpha^4$, the normal modes are split, but by a somewhat smaller amount than had there been no damping.

Figure 15.46 illustrates the very different character of the normal modes in the crossover region in the weak- and strong-coupling regimes. The examples shown are for an undamped string interacting with a structural resonance with a Q of 25, evaluated for coupling factors,

$$K=\frac{2Q_\mathrm{M}}{\omega_\mathrm{M}^2}\alpha=\frac{2Q_\mathrm{M}}{n\pi}\sqrt{\frac{2m}{M}}\,, \qquad (15.67)$$

of 0.75 and $\sqrt{5}$, in the weak- and strong-coupling regimes, respectively.

In the weak-coupling limit, the frequency of the vibrating string exhibits the characteristic perturbation described in the previous section, with a shift in frequency proportional to the imaginary component of the terminating admittance and an increased damping proportional to the real part. Note that the coupling also weakly perturbs the frequency and damping of the coupled structural resonance. However, there is no splitting of modes at the crossover point and the normal modes retain their predominantly string-like or body-like character throughout the transition region.

Amazon
EX1013 Page 586

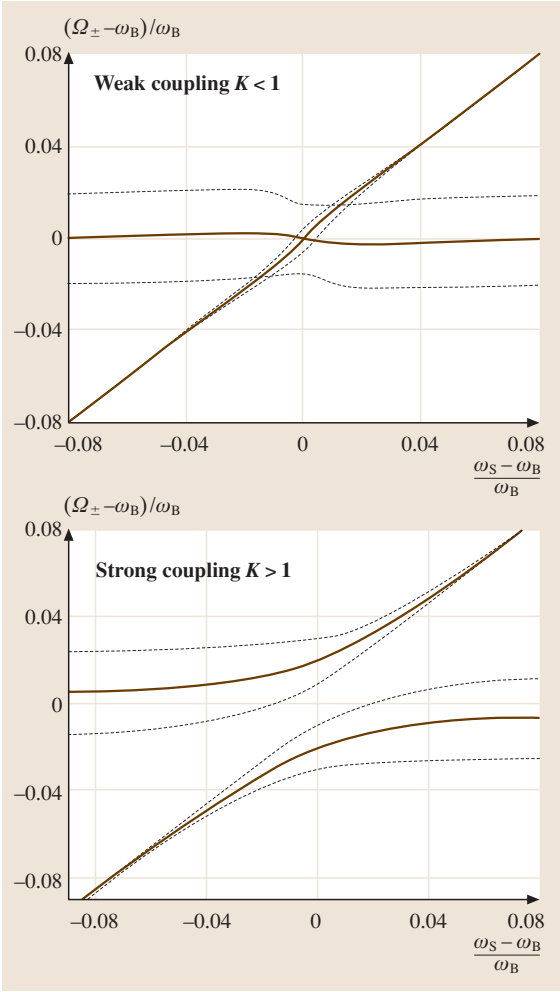


Fig. 15.46 Normal modes of coupled oscillators illustrating the profound effect of damping on the behaviour in the cross-over region illustrated for K-values of 0.75 and $\sqrt{5}$ for an undamped string resonance coupled to a body resonance with a typical $Q = 25$. The *solid line* shows the shifted frequencies of the normal modes as the string frequency is scanned through the body resonance, while the *dashed lines* show the 3 dB points on their damped resonant response (*Gough* [15.80])

In the strong-coupling limit, $K > 1$, the normal modes are split at the crossover point. The losses are also shared equally between the split modes. As the string frequency is varied across the body resonance, one mode changes smoothly from a normal mode with a predominantly string-like character, to a mixed mode at cross over, and to a body-like mode at higher frequencies, and vice versa for the other normal mode.

Our earlier discussion of the perturbation of string resonances by the terminating admittance is therefore justified in the weak-coupling regime ($K \ll 1$), which is the usual situation for most string resonances on musical instruments. However, if the fundamental mode of a string is over-strongly coupled at the bridge to a rather light, weakly damped body resonance, such that $K > 1$, the normal-mode resonant frequency of the vibrating string, when coincident in frequency with the coupled body mode, will be significantly shifted away from its position as the fundamental member of the harmonic set of partials. It is then impossible to maintain a steady Helmholtz bowed waveform on the string at the pitch of the now perturbed fundamental, which is the origin of the wolf-note problem frequently encountered on otherwise often very fine-stringed instruments, and cellos in particular.

To overcome such problems, it is sometimes possible to reduce K by using a lighter string, but more commonly the effective Q-value is reduced by extracting energy from the coupled system by fitting a resonating mass on one of the strings between the bridge and tailpiece. A lossy material can be placed between the added mass and the string to extract energy from the system, which might otherwise simply move the wolf note to a nearby frequency.

String Resonances

Figure 15.47 illustrates: (a) the frequency dependence of the in-phase and phase-quadrature resonant response of an A-string as its tension increased, so that its frequency passes through a relatively strongly coupled body resonance at ≈460 Hz; (b) the splitting in frequency of the normal modes of the second partial of the heavier G-string frequency tuned to coincide with the frequency of the coupled body resonance. Superimposed on these relatively broad resonances is a very sharp resonance arising from transverse string vibrations perpendicular to the strong coupling direction, to be explained in the next section. This very weakly perturbed string resonance provides a marker, which enables us to quantify the shifts and additional damping of string vibrations in the strong coupling direction.

When the frequency of the lighter A-string is tuned below that of the strongly coupled body resonance, the coupling lowers the frequency of the coupled string mode, as anticipated from our earlier discussion. In contrast, when tuned above the coupled resonance the frequency of the coupled string mode is increased, while

Amazon
EX1013 Page 587

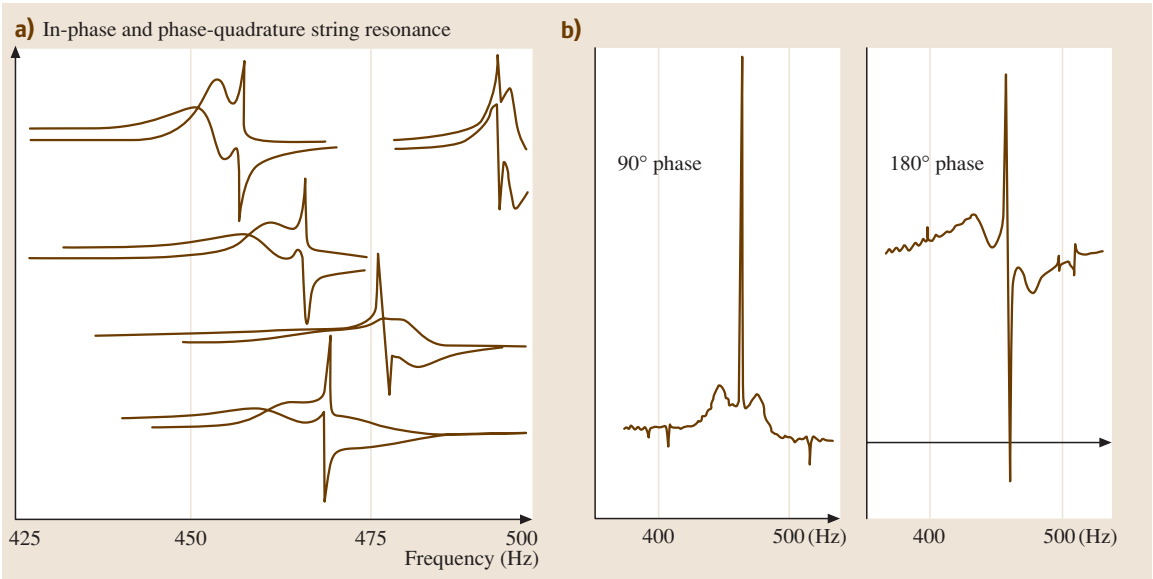


Fig. 15.47a,b Measurements of the in-phase and in-quadrature resonant response of violin strings coupled via the bridge to a strong body resonance (*Gough* [15.33]). The shift of the broader resonances relative to the unperturbed narrow resonance indicates the extent of the perturbative coupling. **(a)** tuning the A-string resonance through a coupled resonance at ≈ 460 Hz; **(b)** the splitting of the string–body normal modes for the more strongly coupled, second partial, of the heavier G-string

at coincidence there is a slight indication of split modes somewhat smaller than the widths. The splitting of modes is clearly seen for the second partial of the much heavier G-string (Fig. 15.47b), with symmetrically split broad string/body modes above and below the narrow *uncoupled mode*. Not surprisingly, this violin suffered from a pronounced wolf note when played at 460 Hz in a high position on the G-string, but not on the lighter D- or A-string. Such effects tend to be even more pronounced on cellos due to the very high bridge providing strong coupling between the vibrating strings and body of the instrument.

On plucked string instruments the inharmonicity of the partials of a plucked note induced by coupling at the bridge to prominent structural resonances causes beats in the sound of plucked string, which contribute to the characteristic sound of individual instruments. *Woodhouse* [15.81,82] has recently made a detailed theoretical, computational and experimental study of such effects for plucked notes on a guitar taking account of the effect of damping on the coupled string–corpus normal modes. This is sometimes not taken into proper account in finite-element software, in which the normal modes of an interacting system are first calculated ignoring damping, with the damping of the modes then added. As is clear from Fig. 15.46, such an approach will always break down whenever the width of resonances associated with damping becomes comparable with the splitting of the normal modes in the absence of damping, as is frequently the case in mechanical and acoustical systems.

Polarisation

We have already commented on the response of a bridge mounted centrally on a symmetrically constructed instrument, with string vibrations perpendicular to the front plate exciting only symmetric modes of the body of the instrument, while string vibrations parallel to the front plate induce a couple on the front plate exciting only asymmetric modes. The terminating admittance at the bridge end of the string will therefore be a strongly frequency dependent function of the polarisation direction of the transverse string modes. The angular dependence of the terminating admittance lifts the degeneracy of the string modes resulting in two independent orthogonal modes of transverse string vibration, with different perturbed frequencies and damping, polarised along the frequency-dependent principal directions of the admittance tensor. If a string is excited at an arbitrary angle, both modes will be excited, so that in free decay the directional polarisation

Amazon
EX1013 Page 588

will precess at the difference frequency. The resultant radiated sound from the excited body resonances will also exhibit beats, which unlike the nonlinear effects considered earlier will not vary with amplitude of string vibration.

In instruments of the violin family, the soundpost removes the symmetry of the instrument, with normal modes involving a mixture of symmetric and asymmetric modes. Measurements like those shown in Fig. 15.47 demonstrate that below ≈ 700 Hz, the effect of the soundpost is to cause the bridge to rock backwards and forwards about the treble foot closest to the soundpost, which acts as a rather rigid fulcrum. This accounts for the very narrow string resonances shown in Fig. 15.47, which correspond to string vibrations polarised parallel to the line between the point of string support and the rigidly constrained right-hand foot, as indicated in Fig. 15.44. In contrast, string vibrations polarised in the orthogonal direction result in a twisting couple acting on the bridge, with the left-hand foot strongly exciting the vibrational modes of the front plate giving the frequency-shifted and broadened string resonances of the strongly coupled string modes.

By varying the polarisation direction of an electromagnetically excited string, one can isolate the two modes and determine their polarisations (*Baker* et al. [15.84]). When such a string is bowed, it will in general be coupled to both orthogonal string modes. The unperturbed string mode may well help stabilise the repetitive Helmholtz bowed waveform.

String–String Coupling

A vibrating string on any multi-stringed instrument is coupled to all the other strings supported on a common supporting bridge. This is particularly important on the piano, where pairs and triplets of strings tuned to the same pitch are used to increase the intensity of the notes in the upper half of the keyboard. Such coupling is also important on instruments like the harp, where the strings and their partials are coupled via the soundboard. On many ancient bowed and plucked stringed instruments, a set of coupled sympathetic string were used to enhance the sonority and long-term decay of plucked and bowed notes. Even on modern instruments like the violin and cello, the coupling of the partials of a bowed or plucked string with those of the other freely vibrating open (unstopped) strings enhances the decaying after-sound of a bowed or plucked note. This may be one of the reasons why string players have a preference for playing in the bright key signatures of G, D and A major associated with the open strings, where both direct and sympathetic vibrations can easily be excited.

The musical importance of such coupling on the piano is easily demonstrated by first playing a single note and holding the key down so that the note remains undamped and then holding the sustaining pedal down, so that many other strings can also vibrate in sympathy and especially those with partials coincident with those of the played note. Composers, such as Debussy, exploit the additional sonorities produced by such coupling, as in La Cathédrale Engloutie CD-ROM.

The influence of coupling at the bridge of the normal modes of string vibration on the piano has been discussed by *Weinreich* [15.83] and for sympathetic string in general by the present author [15.80]. Consider first two identically tuned string terminated by a common bridge with string vibrations perpendicular to the soundboard and relative phases represented by arrows. The normal modes can therefore be described by the combination ↑↑ and ↓↑ with the strings vibrating in phase or in anti-phase. When the strings vibrate in anti-phase ↓↑, they exert no net force on the bridge, which therefore remains a perfect node inducing no perturbation in frequency or additional damping or transfer of energy

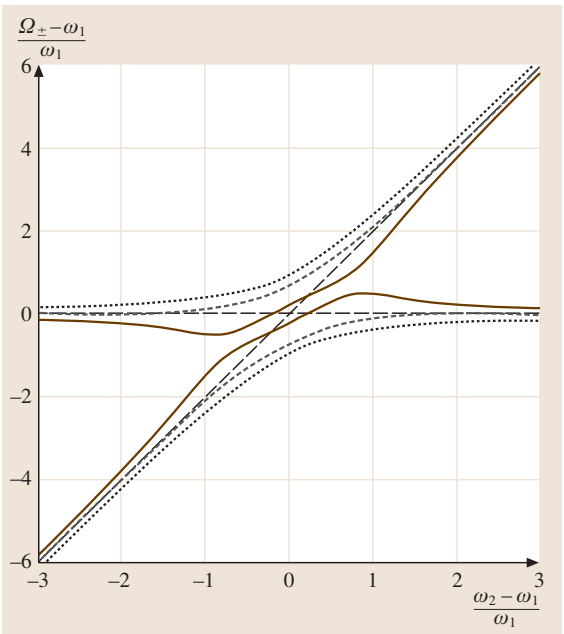


Fig. 15.48 Normal modes of a string doublet coupled at the bridge by a complex impedance. The *dashed* and *dotted curves* illustrate the effect of increasing the reactive component (after *Weinreich* [15.83])

Amazon
EX1013 Page 589

to the soundboard. In contrast, when the strings vibrate in the same phase ↑↑, the force on the bridge and resultant amplitude of sound produced will be doubled, as will the perturbation in frequency and damping of the normal modes, and the amplitude of the resultant sound, relative to that of a single string.

Reactive terms in the common bridge admittance tend to split the frequencies of the normal modes in the vicinity of the crossover frequency region, while resistive coupling at the bridge tends to draw the modal frequencies together over an appreciable frequency range. This is illustrated by Weinreich's predictions for two strings coupled at the bridge by a complex admittance shown in Fig. 15.48, which shows the veering together of the normal modes induced by resistive coupling and the increase in splitting as the reactive component of the coupling is increased.

If the two strings are only slightly mistuned, the amplitudes of the string vibrations involved in the normal modes can still be represented as ↑↑ and ↓↑, but the amplitudes of the two string vibrations will no longer be identical. Hence, when the two strings of a doublet are struck with equal amplitude by a hammer, the mode can be represented by a combination of vibrations ↑↑ with equal amplitude with a small component with opposite amplitudes the ↓↑ dependent on the mistuning. This leads to a double decay in the sound intensity, with the strongly excited ↑↑ mode decaying relatively quickly, leaving the smaller amplitude but weakly damped ↓↑ mode persisting at longer times. Figure 15.49, from *Weinreich* [15.83], shows the rapid decay of a single C4 string, when all other members of the parent string triplet are damped, followed by the much longer long-term decay of the normal mode excited when one other member of the triplet is also allowed to vibrate freely. More-complicated decay patterns with superimposed slow beats are observed for various degrees of mistuning, from interference with small-amplitude orthogonally polarised string modes excited when the strong-coupling direction is not exactly perpendicular to the soundboard. Weinreich suggests that skilled piano tuners deliberately mistune the individual strings of string doublets and triplets to maximise their long-term ringing sound. Any weakly decaying component of a decaying sound is acoustically important because of the logarithmic response of the ear to sound intensity.

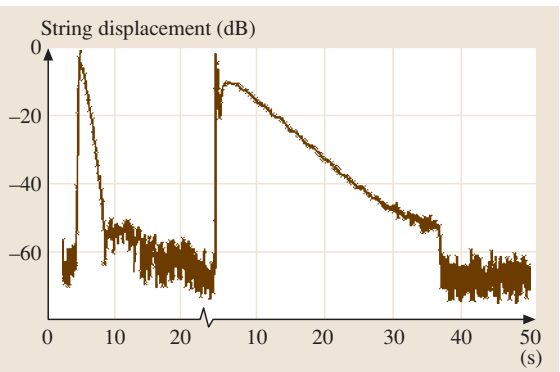


Fig. 15.49 Decay in string vibration of a struck C4 (262 Hz) piano string, first with other members of the string triplet damped and then with one other similarly tuned string allowed to vibrate also (after *Weinreich* [15.83])

15.2.6 Body Modes

Stringed instruments come in a great variety of shapes and sizes, from strings mounted on simple boxes or on skins stretched over hollow gourds to the more complex renaissance shapes of the viols, guitars and members of the violin family. The vibrational modes of all such instruments, which are ultimately responsible for the radiated sound, involve the collective modes of vibration of all their many component parts. For example, when a violin or guitar is played, all parts of the instrument vibrate – the individual strings, the bridge, the front and back plates and ribs that make up the body of the instrument, the air inside its hollow cavity, the neck, the fingerboard, the tailpiece and, for members of the violin family, the internal soundpost also.

Because of the complexity of the dynamical structures, it would be well nigh impossible to work out the modal shapes and frequencies of even the simplest stringed instruments from first principles. However, as we will show later in this section, with the advent of powerful computers and finite-element analysis software, it is possible to compute the modal vibrations and frequencies of typically the first 20 or more normal modes for the violin and guitar below around 1 kHz. Such calculations do indeed show a remarkable variety of vibrational modes, with every part of the instrument involved in the vibrations to some extent. Such modes can be observed by direct experiment using Chladni plate vibrations, laser holography and modal analysis techniques, as briefly described in this section.

The frequencies of the vibrational modes can be obtained even more simply from the admittance measured at the position of string support or other selected position on the body of an instrument, when the instrument is excited at the bridge by a sinusoidal electromagnetic

Amazon
EX1013 Page 590

force or a simple tap. However, unless a large number of measurements over the whole body of the instrument (normal-mode analysis) are made, such measurements provide very little direct information about the nature of the normal modes and the parts of the violin which contribute most strongly to the excited vibrations.

Although a particular structural mode can be very strongly excited, it may contribute very little to the radiated sound and hence the quality of sound of an instrument. Examples of such resonances on the violin or guitar include the strong resonances of the neck and fingerboard. However, even if such resonances produce very little sound, their coupling to the strings via the body and bridge of the instrument can lead to considerable inharmonicity and damping of the string resonances, as discussed in the previous section. Such effects can have a significant effect on the sound of the plucked string of a guitar and the ease with which a repetitive waveform can be established on the bowed string.

To produce an appreciable volume of sound, the normal modes of instruments like the violin and guitar have to involve a net change in volume of the shell structure forming the main body of the instrument. This then acts as a monopole source radiating sound uniformly in all directions. However, when the acoustic wavelength becomes comparable with the size of the instrument, dipole and higher-order contributions also become important.

For the guitar and instruments of the violin family, there are several low-frequency modes of vibration which involve the flexing, twisting and bending of the whole body of the instrument, contributing very little sound to the lowest notes of the instruments. To boost the sound at low frequencies, use is often made of a Helmholtz resonance involving the resonant vibrations of the air inside the body cavity passing in and out of f-holes or rose-hole cut into the front plate of the instrument. This is similar to the way in which the low-frequency sound of a loudspeaker can be boosted by mounting it in a bass-reflex cabinet. The use of a resonant air cavity to boost the low-frequency response has been a common feature of almost every stringed instrument from ancient times.

Although finite-element analysis and modal analysis measurement techniques provide a great wealth of detailed information about the vibrational states of an instrument, considerable physical insight and a degree of simplification is necessary to interpret such measurements. This was recognised by *Savart* [15.85] in the early part of the 19th century, when he embarked on a number of ingenious experiments on the physics of the violin in collaboration with the great French violin maker Vuillaume. To understand the essential physics involved in the production of sound by a violin, he replaced the beautiful, ergonomically designed, renaissance shape of the violin body by a simple trapezoidal shell structure fabricated from flat plates with two central straight slits replacing the elegant f-holes cut into the front. As Savart appears to have recognised, the detailed shape is relatively unimportant in defining the essential acoustics involved in the production of sound by a stringed instrument.

We will adopt a similar philosophy in this section and will consider a stringed instrument made up of its many vibrating components – the strings and bridge, which we have already considered, the supporting shell structure, the vibrations of the individual plates of such a structure, the soundpost which couples the front and back plates, the fingerboard, neck and tailpiece, which vibrate like bars, and the air inside the cavity. Although we have already emphasised that it is never possible to consider the vibrations of any individual component of an instrument in isolation, as we have already shown for the string coupled to a structural resonance at the bridge, it is only when the resonant frequencies of the coupled resonators are close together that their mutual interactions are so important that they change the character of the vibrational modes. Otherwise, the mutual interactions between the various subsystems simply provide a first-order correction to modal frequencies without any very significant change in their modal shapes.

Flexural Thin-Plate Modes

To radiate an appreciable intensity of sound, energy has to be transferred via the bridge from the vibrating strings to the much larger surfaces of a soundboard or body of an instrument. The soundboards of the harp and keyboard instruments and the shell structures of stringed instruments like the violin and guitar can formally be considered as thin plates. Transverse or flexural waves on their surface satisfy the fourth-order thin-plate equation (*Morse* and *Ingard* [15.42] Sect. 5.3), which for an isotropic material can be written as

$$\frac{\partial^2 z}{\partial t^2} + \frac{Eh^2}{12\rho(1-\nu^2)}\left(\frac{\partial^4 z}{\partial x^4} + 2\frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} + \frac{\partial^4 z}{\partial y^4}\right) = 0\,, \tag{15.68}$$

where z is the displacement perpendicular to the xy-plane, h is the plate thickness, E is the Young's modulus, ν is the Poisson ratio, and ρ the density.

It is instructive first to consider solutions for a narrow quasi-one-dimensional thin plate, like a wooden ruler or

Amazon
EX1013 Page 591

the fingerboard on a violin. One-dimensional solutions can be written in the general form

$$z = (a\cos kx + b\sin kx + c\cosh kx + d\sinh kx)\,\mathrm{e}^{\mathrm{i}\omega t}\,, \tag{15.69}$$

where

$$\omega = \left(\frac{E}{12\rho(1-\nu^2)}\right)^{1/2} hk^2\,. \tag{15.70}$$

The hyperbolic functions correspond to displacements that decay exponentially away from the ends of the bar as $\exp(\pm kx)$. Well away from the ends, the solutions are therefore very similar to transverse waves on a string, except that the frequency now depends on k^2 rather than k with a phase velocity $c = \omega/k$ proportional to $k \sim \omega^{1/2}$. Flexural waves on thin plates are therefore dispersive and unlike waves travelling on strings any disturbance will be attenuated and broadened in shape as it propagates across the surface.

The k values are determined by the boundary conditions at the two ends of the bar, which can involve the displacement, couple $M = -ES\kappa^2\partial^2 z/\partial x^2$ and shearing force $F = \partial M/\partial x = -ES\kappa^2\partial^3 z/\partial x^3$ at the ends of the bar, where k is the radius of gyration of the cross section (*Morse* and *Ingard* [15.42] Sect. 5.1).

For a flexible bar there are three important boundary conditions:

1. freely hinged, where the free hinge cannot exert a couple on the bar, so that

$$z = 0 \quad \text{and} \quad \frac{\partial^2 z}{\partial x^2} = 0\,, \tag{15.71}$$

2. clamped, where the geometrical constraints require

$$z = 0 \quad \text{and} \quad \frac{\partial z}{\partial x} = 0\,, \tag{15.72}$$

3. free, where both the couple and the shearing force at the ends are zero, so that

$$\frac{\partial^3 z}{\partial x^3} = \frac{\partial^2 z}{\partial x^2} = 0\,. \tag{15.73}$$

A bar of length L, freely hinged at both ends, supports simple sinusoidal spatial solutions with m half-wavelengths between the ends and modal frequencies

$$\omega_m = h\sqrt{\frac{E}{12\rho(1-\nu^2)}}\left(\frac{m\pi}{L}\right)^2\,. \tag{15.74}$$

For long bars with clamped or free ends, the nodes of the sinusoidal component are moved inwards by a quarter of a wavelength and an additional exponentially decaying solution has to be added to satisfy the boundary conditions, so that at the $x = 0$ end of the bar

$$z \sim A\left[\sin(k_\mathrm{m}x - \pi/4) \pm \frac{1}{\sqrt{2}}\,\mathrm{e}^{-k_m x}\right]\,, \tag{15.75}$$

where the plus sign corresponds to a clamped end and the minus to a free end, and $k_m = (m+1/2)\pi/L$. The modal frequencies are given by

$$\omega_m = h\sqrt{\frac{E}{12\rho(1-\nu^2)}}\left(\frac{(m+1/2)\,\pi}{L}\right)^2 \tag{15.76}$$

which, for the same m value, are raised slightly above those of a bar with hinged ends. Corrections to these formulae from the leakage of the exponentially decaying function from the other end of the bar are only significant for the $m = 1$ mode and are then still less than 1%.

The solutions close to the end of a bar for hinged, clamped and free boundary conditions are illustrated in Fig. 15.50, with the phase-shifted sinusoidal component for the latter two indicated by the dotted line. The exponential contribution is only significant out to distances $\sim \lambda/2$.

The above formulae can be applied to the bending waves of quasi-one-dimensional bars of any cross section, by replacing the radius of gyration $\kappa = h/\sqrt{12}$ of the thin rectangular bar with $a/2$ for a bar of circular cross section and radius a, and $\sqrt{a^2+b^2}/2$ for a hollow cylinder with inner and outer radii a and b (*Fletcher* and *Rossing* [15.5], Fig. 2.19).

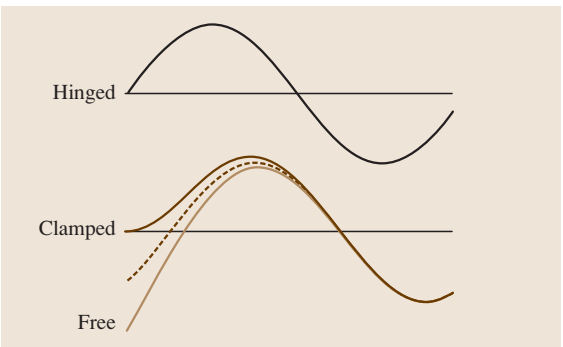


Fig. 15.50 Boundary conditions for flexural waves at the end of a one-dimensional bar. The *dashed line* represents the phase-shifted sinusoidal component, to which the exponentially decaying component has to be added to satisfy the boundary conditions

Amazon
EX1013 Page 592

Another case of practical importance in musical acoustics is a bar clamped at one end and free at the other. This would, for example, describe the bars of a tuning fork or could be used to model the vibrations of the neck or finger board on a stringed instrument. In this case, there is an addition $m = 0$ vibrational mode, with exponential decay length comparable with the length of the bar. The modal frequencies are then given [*Fletcher* and *Rossing* [15.5], (2.64)] by

$$\omega_m = \frac{h}{4}\left(\frac{\pi}{L}\right)^2 \sqrt{\frac{E}{12\rho(1-\nu^2)}} \left[1.194^2, 2.988^2, 5^2, \ldots, (2m+1)^2\right] . \tag{15.77}$$

In the above discussion, we have described the modes in terms of the number m of half-wavelengths of the sinusoidal component of the wave solutions within the length of the bar. A different nomenclature is frequently used in the musical acoustics literature, with the mode number classified by the number of nodal lines (or points in one dimension) m in a given direction *not including the boundaries* rather than the number of half-wavelengths m between the boundaries, as in Fig. 15.51.

Twisting or Torsional Modes

In addition to flexural or bending modes, bars can also support twisting (torsional) modes, as illustrated in Fig. 15.51 for the $z = xy$ (1,1) mode.

The frequencies of the twisting modes are determined by the cross section and shear modulus G, equal to $E/2(1+\nu)$ for most materials (*Fletcher* and *Rossing* [15.5], Sect. 2.20). The wave velocity of torsional waves is dispersionless (independent of frequency) with $\omega_n = nc_T k$, where

$$c_T = \frac{\omega}{k} = \sqrt{\frac{GK_T}{\rho I}} = \alpha\sqrt{\frac{E}{2\rho(1+\nu)}} , \tag{15.78}$$

where GK_T is the torsional stiffness given by the couple, $C = GK_T\partial\theta/\partial x$, required to maintain a twist of the bar through an angle θ and $I = \int \rho r^2 \,\mathrm{d}S$ is the moment of inertia per unit length along the bar. For a bar of circular cross section $\alpha = 1$, for square cross section $\alpha = 0.92$, and for a thin plate with width $w > 6h$, $\alpha = (2h/w)$. For a bar that is fixed at both ends, $f_n = nc_T/2L$, while for a bar that is fixed at one end and free at the other, $f_n = (2n+1)c_T/4L$, where n is an integer including zero.

Thin bars also support longitudinal vibrational modes, but since they do not involve any motion perpendicular to the surface they are generally of little acoustic importance, other than possibly for the lowest-frequency soundpost modes for the larger instruments of the violin family.

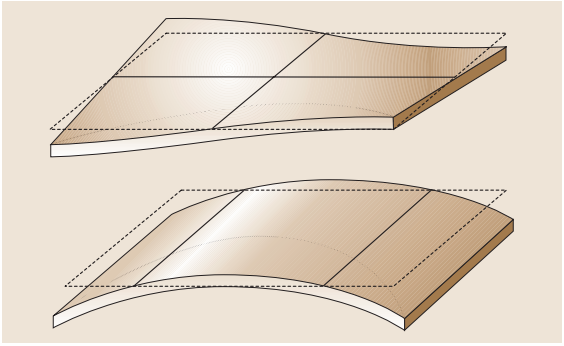


Fig. 15.51 Schematic illustration of the lowest-frequency twisting (1,1) and bending (2,0) modes of a thin bar with free ends

Two-Dimensional Bending Modes

Solutions of the thin-plate bending wave solutions in two dimensions are generally less straightforward, largely because of the more-complicated boundary conditions, which couple the bending in the x- and y-directions. For a free edge parallel to the y-axis, the boundary conditions are (*Rayleigh* [15.3] Vol. 1, Sect. 216)

$$\frac{\partial^2 z}{\partial x^2} + \nu\frac{\partial^2 z}{\partial y^2} = 0$$

and

$$\frac{\partial}{\partial x}\left[\frac{\partial^2 z}{\partial x^2} + (2-\nu)\frac{\partial^2 z}{\partial y^2}\right] = 0 . \tag{15.79}$$

Thus, when a rectangular plate is bent downwards along its length, it automatically bends upwards along its width and vice versa. This arises because downward bending causes the top surface of the plate to stretch and the bottom surface to contract along its length. But by Poisson coupling, this causes the top surface to contract and lower surface to stretch in the orthogonal direction, causing the plate to bend in the opposite direction across its width. This is referred to as anticlastic bending. The Poisson ratio ν can be determined from the ratio of the curvatures along the bending and perpendicular directions.

In addition, for orthotropic materials like wood, from which soundboards and the front plates of most stringed instruments are traditionally made, the elastic constants are very different parallel and perpendicular to the grain structure associated with the growth rings. *McIntyre*

Amazon
EX1013 Page 593

and *Woodhouse* [15.86] have published a detailed account of the theory and derivation of elastic constants from measurement of plate vibrations in both uniform and orthotropic thin plates, including the influence of damping.

For an isotropic rectangular thin plate, hinged along on all its edges, a simple two-dimensional (2-D) sine-wave solution satisfies both the wave equation (15.68) and the boundary conditions, with m and n half-wavelengths along the x- and y-directions, respectively, giving modal frequencies

$$\omega_{mn} = h\sqrt{\frac{E}{12\rho(1-\nu^2)}}\left[\left(\frac{m\pi}{L_x}\right)^2+\left(\frac{n\pi}{L_y}\right)^2\right]. \tag{15.80}$$

By analogy with our discussion of flexural waves in one-dimensional bars, we would expect the modal frequencies of plates with clamped or free edges to be raised, with the nodes of the sinusoidal components of the wave solution moved inwards from the edges by approximately quarter of a wavelength. For the higher-order modes, the modal frequencies would therefore be given to a good approximation by

$$\omega_{mn} = h\sqrt{\frac{E}{12\rho(1-\nu^2)}}\left[\left(\frac{(m+1/2)\,\pi}{L_x}\right)^2+\left(\frac{(n+1/2)\,\pi}{L_y}\right)^2\right]. \tag{15.81}$$

As recognised by *Rayleigh* ([15.3] Vol. 1, Sect. 223), it is difficult to evaluate the modal shapes and modal frequencies of plates with free edges. The method used by Rayleigh was to make an intelligent guess of the wave-functions which satisfied the boundary conditions and to determine the frequencies by equating the resulting potential and kinetic energies. *Leissa* [15.88] has reviewed various refinements of the original calculations. For a plate with free edges, the nodal lines are also no longer necessarily straight, as they were for plates with freely hinged edges.

Chladni Patterns

The modal shapes of vibrating plates can readily be visualised using Chladni patterns. These are obtained by supporting the plate at a node of a chosen mode excited electromagnetically, acoustically or with a rosined bow drawn across an edge. A light powder is sprinkled onto the surface. The plate vibrations cause the powder to bounce up and down and move towards the nodes

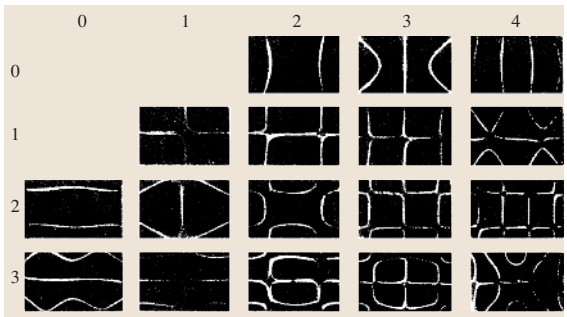


Fig. 15.52 Chladni pattern with *white lines* indicating the nodal lines of the first few modes of a rectangular plate (after *Waller* [15.87])

of the excited mode, allowing the nodal line patterns to be visualised. Figure 15.52 illustrates Chladni patterns measured by *Waller* [15.87] for a rectangular plate with dimensions $L_x/L_y = 1.5$, with the number of nodal lines between the boundary edges determining the nomenclature of the modes. Note the curvature of the nodal lines resulting from the boundary conditions at the free edges.

Figure 15.53 illustrates the nodal line shapes and relative frequencies of the first 10 modes of a square plate with free edges, where $f_{11} = hc_L/L^2\sqrt{1-\nu/2}$ (after *Fletcher* and *Rossing*, [1] Fig. 3.13).

Another important consequence of the anticlastic bending is the splitting in frequencies of combination modes that would otherwise be degenerate. This is illustrated in Fig. 15.54 by the combination $(2,0) \pm (0,2)$ normal modes of a square plate with free edges. The $(2,0) \pm (0,2)$ modes are referred to as the X- and ring-modes from their characteristic nodal line shapes. The $(2,0)-(0,2)$ X-mode exhibits anticlastic bending in the same sense as that induced naturally by the Poisson coupling. It therefore has a lower elastic energy and hence lower vibrational frequency than the $(0,2)+(2,0)$ ring-

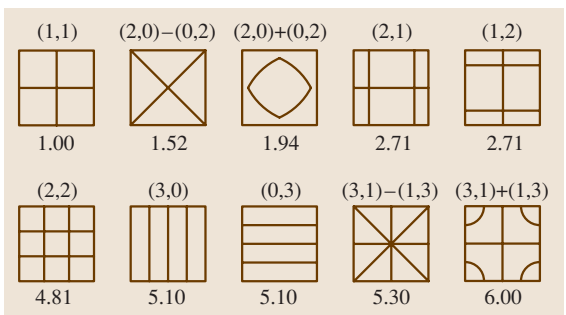


Fig. 15.53 Schematic representation of the lowest 10 vibrational modes of a square plate with free edge

Amazon
EX1013 Page 594

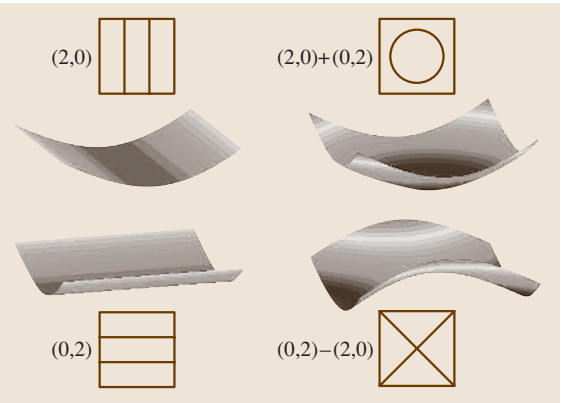


Fig. 15.54 Formation of the ring- and *X*-modes by the superposition of the (2, 0) and (0, 2) bending modes

mode, with curvatures in the same sense in both the *x*- and *y*-directions. The ring- and *X*-modes will therefore be split in frequency above and below the otherwise degenerate mode, as illustrated in *Fletcher* and *Rossing* ([15.5], Fig. 13.11).

Plate Tuning

The modal shapes and frequencies of the lowest-order (1,1) twisting mode and the *X*-and ring-modes are widely used for the *scientific tuning* of the front and back plates of violins following plate-tuning guidelines developed by *Hutchins* [15.89, 91]. These are referred to as violin plate modes 1, 2 and 5, as illustrated in Fig. 15.55 by Chladni patterns for a "well-tuned" back plate. The violin maker aims to adjust the thinning of the plates

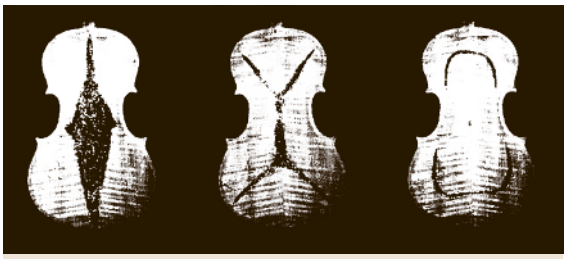


Fig. 15.55 Chladni patterns for the first twisting-(#1), X-(#2) and ring-(#5) modes of a viola back plate (after *Hutchins* [15.89])

across the area of the plate to achieve these symmetrical nodal line shapes at specified modal frequencies.

The use of such methods undoubtedly results in a high degree of quality control and reproducibility of the acoustic properties of the individual plates before assembly and, presumably, of the assembled instrument also, especially for the lower-frequency structural resonances. Unfortunately, they do not necessarily result in instruments comparable with the finest Italian instruments, which were made without recourse to such sophisticated scientific methods. Traditional violin makers instinctively assess the elastic properties of the plates by their feel as they are twisted and bent, and also by listening to the sound of the plates as they are tapped or even rubbed around their edges, rather like a bowed plate. From our earlier discussion, it is clear that the mass of the plates is also important in governing the acoustical properties.

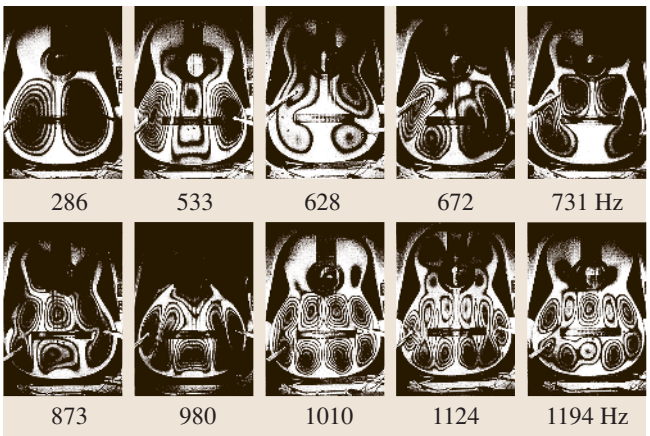


Fig. 15.56 Typical modal shapes for a number of low-frequency modes of a guitar top plate from time-averaged holographic measurements by *Richardson* and *Roberts* [15.90]

Geometrical Shape Dependence

The above examples demonstrate that the lower-frequency vibrational modes of quite complicated shaped plates can often be readily identified with those of simple rectangular plates, though the frequencies of such modes will clearly depend on the exact geometry involved. This is further illustrated in Fig. 15.56 by the modal shapes of a guitar front plate obtained from time-averaged holography measurements by *Richardson* and *Roberts* [15.90], where the contours indicate lines of constant displacement perpendicular to the surface. For the guitar, the edges of the top plate are rather good nodes, because of the rather heavy supporting ribs and general construction of the instrument. The boundary conditions along the edges of the plate are probably intermediate between hinged and clamped. The modes can be denoted by the number of half-wavelengths along the length and width of the instrument. Note that cir-

Amazon
EX1013 Page 595

cular rose-hole opening in the front face, which plays an important role in determining the frequency of the Helmholtz air resonance boosting the low-frequency response of the instrument, tends to concentrate most of the vibrational activity to the lower half of the front plate.

Mode Spacing

Although the frequencies of the modes of complicated shapes such as the violin, guitar and piano soundboard are rather irregularly spaced, at sufficiently high frequencies, one can use a statistical approach to estimate the spacing of the modal frequencies (*Cremer* [15.29], Sect. 11.2). For large m and n values, the modal frequencies of an isotropic rectangular plate are given by

$$\omega_{mn} \sim h\sqrt{\frac{E}{12\rho(1-\nu^2)}}\left({k_m}^2+k_n^2\right) , \qquad (15.82)$$

where $k_m = m\pi/L_x$ and $k_n = n\pi/L_y$. The modes can be represented as points (m, n) on a rectangular grid in k-space, with a grid spacing of π/L_x and π/L_y along the k_x and k_y directions. Each mode therefore occupies an area in k-space of π^2/L_xL_y. For large m and n, the number of modes ΔN between k and $k+\Delta k$ is therefore on average just the number of modes in the k-space area between k and $k+\Delta k$, so that

$$\Delta N = \frac{\pi}{2}k\Delta k\frac{L_xL_y}{\pi^2} = \frac{\pi}{2}\frac{\Delta\omega}{2\beta}\frac{L_xL_y}{\pi^2} , \qquad (15.83)$$

where we have made use of the dispersion relationship $\omega = \beta k^2$.

The density of modes per unit frequency is then constant and independent of frequency,

$$\frac{\mathrm{d}N}{\mathrm{d}f} = \frac{1}{2\beta}L_xL_y = \frac{\sqrt{3\left(1-\nu^2\right)}}{c_\mathrm{L}h}S \sim 1.5\frac{S}{c_\mathrm{L}h} , \qquad (15.84)$$

where S is the area of the plate. The spacing of modes Δf will therefore on average be $\sim c_\mathrm{L}h/1.5S$, proportional to the plate thickness and inversely proportional to plate area. For large k values, this result becomes independent of the shape of the plate. For an orthotropic plate like the front plate of the violin or guitar, the spacing is determined by the geometric mean $(c_xc_y)^{1/2}$ of the longitudinal velocities parallel and along the grain. For the violin, *Cremer* ([15.29], p. 292) estimates an asymptotic average mode spacing of 73 Hz for the top plate and 108 Hz for the back plate. Above around 1.5 kHz the width of the resonances on violin and guitar plates becomes comparable with their spacing, so that experimentally it becomes increasingly difficult to excite or distinguish individual modes.

On many musical instruments such as the violin and guitar, the presence of the f- and rose-hole openings introduce additional free-edge internal boundary conditions, which largely confine the lower-frequency modes to the remaining larger areas of the plate. The effective area determining the density of lower-frequency modes for both instruments will therefore be significantly less that that of the whole plate. Any reduction in plate dimensions, such as the island region on the front plate of the violin between the f-holes, will limit the spatial variation of flexural waves in that direction. Such a region will therefore not contribute significantly to the normal modes of vibration of the plate until $\lambda(\omega)/2$ is less than the limiting dimension.

Anisotropy of Wood

Wood is a highly anisotropic material with different elastic properties perpendicular and parallel to the grain. Furthermore, the wood used for soundboards and plates of stringed instruments are cut from nonuniform circular logs (slab or quarter cut), so that their properties can vary significantly across their area.

McIntyre and *Woodhouse* [15.86] have described how the anisotropic properties affect the vibrational modes of rectangular thin plates and have shown how the most important elastic constants including their loss factors can be determined from the vibrational frequencies and damping of selected vibrational modes. For a rectangular plate with hinged edges

$$\omega_{mn}^2 = \frac{h^2}{\rho}\left[D_1k_m^4 + D_3k_n^4 + (D_2+D_4)\,k_m^2k_n^2\right] , \qquad (15.85)$$

where D_1–D_4 are the four elastic constants required to describe the potential energy of a thin plate with orthotropic symmetry. These are related to the more familiar elastic constants by the following relationships

$$D_1 = E_x/12\mu\,,\; D_3 = E_y/12\mu\,,\; D_4 = G_{xy}/3\,,$$
$$D_2 = \nu_{xy}E_y/6\mu = \nu_{yx}E_x/6\mu\,, \qquad (15.86)$$

where $\mu = 1-\upsilon_{xy}\upsilon_{yx}$. G_{xy} gives the in-plane shear energy when a rectangular area on the surface is distorted into a parallelogram. This is the only energy term involved in a pure twisting mode (i. e. $z = xy$). For an isotropic plate, $D_4 = E/6(1+\nu)$.

Amazon
EX1013 Page 596

Table 15.5 Typical densities and elastic properties of wood used for stringed instrument modelling (after *Woodhouse* [15.76]). (The values with asterisks are intelligent guesses in the absence of experimental data)

Property	Symbol	Units	Spruce	Maple
Density	ρ	kg/m^3	420	650
	D_1	MPa	1100	860
	D_2	MPa	67	140*
	D_3	MPa	84	170
	D_4	MPa	230	230*
Relative scaling factors	$\sqrt[4]{D_1/D_3}$		1.9	1.4

For many materials, $(D_2 + D_4) \sim 2\sqrt{D_1 D_2}$, so that (15.82) can be rewritten as

$$\omega_{mn}^2 = \frac{h^2}{\rho}\sqrt{D_1 D_3}\left[\sqrt{\frac{D_1}{D_3}}\left(\frac{m\pi}{L_x}\right)^2 + \sqrt{\frac{D_3}{D_1}}\left(\frac{n\pi}{L_y}\right)^2\right]^2 . \tag{15.87}$$

The vibrational frequencies are therefore equivalent to those of a shape of the same area with averaged elastic constant $\sqrt{D_1 D_3}$ and length scales L_x multiplied by the factor $\sqrt[8]{D_1/D_3}$ and L_y by its inverse. The relative change in scaled dimensions is therefore $\sqrt[4]{D_1/D_3}$. These scaling factors account for the elongation of the equal contour shapes along the stiffer bending direction in the holographic measurements of mode shapes for the front plate of the guitar (Fig. 15.56), where the higher elastic modulus along the grains is further increased by strengthening bars glued to the underside of the top plate at a shallow angle to the length of the top plate.

Typical values for the elastic constants of spruce and maple traditionally used for modelling the violin are listed in Table 15.5 from *Woodhouse* [15.76]. The anisotropy of the elastic constants along and perpendicular to the grain of a spruce plate cut with the growth rings running perpendicular to the surface would give a relative scaling factor for a violin front plate of almost double the relative width, if one wanted to consider the flexural vibrations in terms of an equivalent isotropic thin plate. The anisotropy is therefore very important in determining the vibrational modes of such instruments.

Plate Arching

The front and back plates of instruments of the violin family have arched profiles, which give the instrument a greatly enhanced structural rigidity to support the downward component of the string tension (≈ 10 kg weight). The arching also significantly increases the frequencies of the lowest flexural plate modes. In the case of lutes, guitars and keyboard instruments with a flat sounding board or front plate, the additional rigidity is achieved by additional cross-struts glued to the back of the sounding board. The bass bar in members of the violin family serves a similar purpose in providing addition strengthening to that of the arching.

The influence of arching on flexural vibration frequencies is easily understood by considering the transverse vibrations of a thin circular disc. For a flat disc, the modal frequencies are determined by the flexural energy associated with the transverse vibrations. The longitudinal strains associated with the transverse vibrations are only second order in displacement and can therefore be neglected. However, if the disc is *belled out* to raise the centre to a height H, the transverse vibrations now involve additional first-order longitudinal strains stretching the disc away from its edges. The energy involved in such stretching, which is resisted by the rigidity of the circumferential regions of the disc, introduces an additional potential energy proportional to H^2. By equating the kinetic to the increased potential energy, is follows that the frequency of the lowest-order symmetrical mode will be increased by a factor $[1+\alpha(H/h)^2]^{1/2}$, where $\alpha \approx 1$ has to be determined by detailed calculation. *Reissner* [15.92] showed that, when the arching is larger than the plate thickness, $H \gg h$, the frequency of the fundamental mode is raised by a factor $\omega/\omega_0 = 0.68H/h$ for a circular disc with clamped edges, and $0.84H/h$ with free edges. For a shallow shell with $H/a < 0.25$, where a is the radius of the disc, the asymptotic frequency can conveniently be expressed as $\omega_n \sim 2(E/\rho)^{1/2}H/a^2$. The arching dependence of the modal frequencies is greatest for the lowest-frequency modes. At high frequencies, the radius of curvature is large compared to the wavelength, so arching is much less important.

The combined effect of the arching and the f- and rose-holes cut into the front face of many stringed instruments is to raise the frequency of the acoustically important lower-frequency modes to well above the asymptotic spacing of modal frequencies predicted by (15.82). For example, the lowest-frequency plate modes of the violin front and back plates are typically in the range 400–500 Hz compared with Cremer's predictions for an asymptotic spacing of modes ≈73 Hz for the top plate and 108 Hz for the back plate ([15.29], p. 292).

The more highly arched the plates the stiffer they will be and, for a given mass, the higher will be their

Amazon
EX1013 Page 597

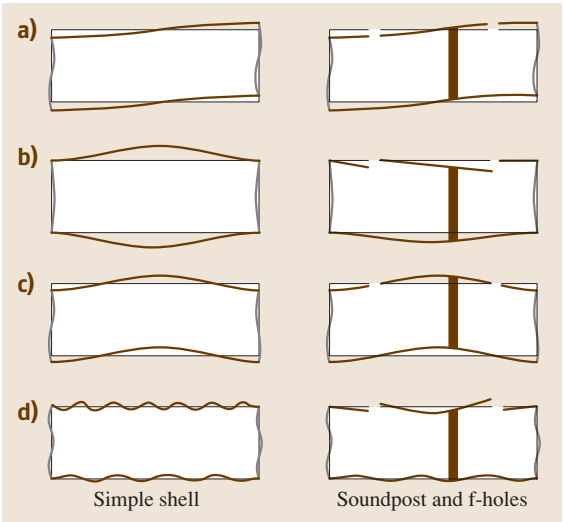


Fig. 15.57 Schematic cross-sectional representation of typical shell modes for a simple box and a "violin-type" structure with f-holes and a soundpost

associated vibrational frequencies. High arching may well contribute to the relatively soft and sweet sounds of the more highly arched early Amati and Stainer violins and the more powerful and brilliant sounds of the flatter later Stradivari and Guarneri models.

Shell Structures

Although it is interesting to investigate the free plates of violins and other instruments before they are assembled into the instrument as a whole, once assembled their vibrational properties will generally be very different, as they are subject to completely different boundary conditions at their supporting edges and by the soundpost constraint for instruments of the violin family. The supporting ribs tie the outer edges of the back and front plates together. The curvature of the outer edge shape gives the 1–2 mm-thick ribs of a violin considerable structural strength and rigidity, in much the same way as the bending of a serpentine brick wall. In many instruments there are extra strips and blocks attached to the ribs and plate edges to strengthen the joint, which still allow a certain amount of angular flexing, as indicated by the schematic normal-mode vibrations illustrated in Fig. 15.57. The supporting ribs add mass loading at the edges of the plates and impose a boundary condition for the flexing plates intermediate between freely hinged and clamped.

For a simple shell structure, Fig. 15.57a represents a low-frequency twisting mode in which the two ends of the instrument twist in opposite directions, just like the simple ($z = xy$) twisting mode of a rectangular plate. Figure 15.57b–d schematically represent normal modes involving flexural modes of the front and back plates. Mode (b) is the important *breathing mode*, which produces a strong monopole source of acoustic radiation at relatively low frequencies. In mode (c), the two plates vibrate in the same direction, resulting in a much weaker dipole radiation source along the vertical axis. The above examples assumed identical top and back plates, whereas in general they will have different thicknesses arching, and will be constructed from different types of wood with different anisotropic elastic properties: spruce for the front and maple for the back plate of the violin. Hence, for typical high-frequency normal modes (e.g. shown schematically in Fig. 15.57d), the wavelengths of flexural vibrations will be different in the top and back plates.

Note that, at low frequencies, several of the normal modes involve significant motion of the outer edges of the front and back plate, since the centre of mass of the freely supported structure cannot move. Hence, when an instrument is supported by the player, additional mode damping can occur by energy transfer to the chin, shoulder or fingers supporting the instrument at its edges, as indeed observed in modal analysis investigations on hand-held violins by *Marshall* [15.93] and *Bissinger* [15.94].

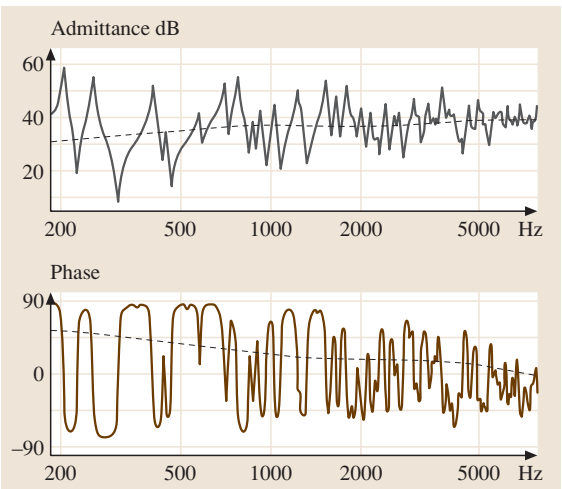


Fig. 15.58 The rotational admittance across the two feet of a bridge on an idealised rectangular violin structure with a soundpost under the treble foot but with no f-holes (after Woodhouse [15.76])

Amazon
EX1013 Page 598

Skeleton Curves

In this idealised model, the normal modes of the structure at high frequencies will be similar to those of the individual plates. Such modes will only be significantly perturbed when the resonances of the separate plates are close together, apart from a general background interaction from the average weak but cumulative interaction with all other distant modes. *Woodhouse* [15.76] has recently shown that the averaged amplitude and phase of the admittance can be described by *skeleton curves*, indicated by the dashed lines in Fig. 15.58, on which peaks and troughs of height Q and $1/Q$ and phase changes from individual resonances are superimposed. These curves were evaluated analytically for the rotational admittance across the two feet of a bridge mounted on the front plate of an idealised rectangular box-like violin without f-holes but with a soundpost close to the treble foot of the bridge.

At low frequencies the averaged input impedance across the two feet of the bridge is largely reactive, with a value close to the static *springiness*, which can be identified with the low-frequency limit of the normal-mode admittance $\sum_n \mathrm{i}\omega/m_n\omega_n^2$, where the effective mass of each mode will depend on where and how the instrument is excited. However, at high frequencies the admittance becomes largely resistive resulting from internal damping and energy loss to the closely overlapping modes. The use of skeleton curves enables Woodhouse to illustrate the effect of various different bridge designs on the overall frequency response of a violin, without having to consider the exact positions, spacing or damping of the individual resonances of the shell structure. Although the idealised model is clearly over-simplistic, the general trends predicted by such a model will clearly be relevant to any multi-resonant shell model.

Soundpost and f-Holes

The soundpost and f-holes cut into the front plate of the violin and related instruments have a profound effect on the frequencies and waveforms of the normal modes, illustrated schematically by the right-hand set of examples in Fig. 15.57. The f-holes create an island area on which the bridge sits, which separates the top and lower areas of the front plate. Like the rose-hole on a guitar illustrated in Fig. 15.56, the additional internal free edges introduced by the f-holes tend to localise the vibrations of the front plate to the regions above and below the *island area*. In addition, the soundpost acts as a rather rigid spring locking the vibrations of the top and back plates together at it its ends. At low frequencies, the soundpost introduces an approximate node of vibration on both the top and back plates, unless the frequencies of the uncoupled front and back plates modes are close together.

For some low-frequency modes, the soundpost and f-hole have a relatively small effect on the modes of the shell structure, such as the twisting mode (a) and mode (c), when the plates vibrate in the same direction. However, the breathing mode (c) will be strongly affected by the soundpost forcing the front and back plates to move together across its ends.

As indicated earlier, any string motion parallel to the plates will exert a couple on the top of the bridge. In the absence of the soundpost, only asymmetric modes of the top plate could then be excited. However, to satisfy the boundary conditions at the soundpost position, the rocking action now induces a combination of symmetric and antisymmetric plate modes (illustrated schematically in Fig. 15.57b), approximately doubling the number of modes that can contribute to the sound of an instrument including the very important lower-frequency symmetric *breathing modes*. Because of the f-holes, the central island can vibrate in the opposite direction to the wings on the outer edges of the instrument. The mixing of symmetric and antisymmetric modes is strongly dependent on the position of the soundpost relative to the nodes of the coupled waveforms. As a result, the sound of a violin instrument is very sensitive to the exact placing of the soundpost. The difference in the sound of a violin with the soundpost first in place and then removed is illustrated in CD-ROM.

To a good approximation, in the audible frequency range, the violin soundpost can be considered as a rigid body, as its first longitudinal resonance is ≈100 kHz, though lower-frequency bending modes can also be excited, particularly if the upper and lower faces of the soundpost fail to make a flat contact with the top and back plates (*Fang* and *Rogers* [15.95]). At high frequencies, there is relatively little induced motion of the outer edges of top and back plates, so that the impedance $Z(\omega)$ (force/induced velocity) measured at the soundpost position is simply given by the sum of the impedances at the soundpost position, $Z(\omega)_{\mathrm{top}} + Z(\omega)_{\mathrm{back}}$, of the individual plates with fixed outer edges. If one knows the waveforms of the individual coupled modes, it is relatively straightforward to evaluate the admittance at any other point on the two surfaces, and hence to evaluate the rotational admittance across the two feet of the bridge (*Woodhouse* [15.76]).

We have already described the important role of the bridge dynamics in the coupling between the strings and the vibrational modes of the instrument. For instruments of the violin family, the island region between the f-holes

Amazon
EX1013 Page 599

probably plays a rather similar role to the bridge, as it is via the vibrations of this central region that the larger-area radiating surfaces of the front plate are excited. At low frequencies this will be mainly by the lowest-order twisting and flexing modes of the central island region. It therefore seems likely that the dynamics of the central island region also contributes significantly to the BH hill feature and the resulting acoustical properties of the violin in the perceptually important frequency range of $\approx 2-4$ kHz, as recognised by Cremer and his colleagues [15.11].

Historically, the role of the soundpost and the coupling of plates through enclosed air resonances were first considered analytically using relatively simple mass–spring models with a limited number of degrees of freedom to mimic the first few resonances of the violin, as described in some detail by *Cremer* ([15.29], Chap. 10). Now that we can obtain detailed information about not only the frequencies, but also the shapes of the important structural modes of an instrument from finite-element calculations, holography and modal analysis, there is greater emphasis on analytic methods based on the observed set of coupled modes.

The Complete Instrument

Bowing, plucking or striking a string can excite every conceivable vibration of the supporting structure including, where appropriate, the neck, fingerboard, tailpiece and the partials of all strings both in front of and behind the bridge. Many of the whole-body lower-frequency modes can be visualised by considering all the possible ways in which a piece of soft foam, cut into the shape of the instrument with an attached foam neck and fingerboard, can be flexed and bent about its centre of mass.

Figure 15.59 illustrates the flexing, twisting and changes in volume of the shell of a freely supported violin for two prominent structural resonances computed by *Knott* [15.96] using finite-element analysis. However, not all modes involve a significant change in net volume of the shell, so that many of the lower-frequency modes are relatively inefficient acoustic radiators. Nevertheless, since almost all such modes involve significant bridge motion, they will be strongly excited by the player and will produce prominent resonant features in the input admittance at the point of string support on the bridge. They can therefore significantly perturb the vibrations of the string destroying the harmonicity of the string resonances and resulting playability of particular notes on the instrument, especially for bowed stringed instrument.

Helmholtz Resonance

Almost all hand-held stringed instruments and many larger ones such as the concert harp make use of a Helmholtz air resonance to boost the sound of their lowest notes, which are often well below the frequencies of the lowest strongly excited, acoustically efficient, structural resonances. For example, the lowest acoustically efficient body resonance on the violin is generally around 450 Hz, well above the bottom note G3 of the instrument at ≈ 196 Hz. Similarly, the first strong structural resonance on the classical acoustic guitar is ≈ 200 Hz, well above the lowest note of ≈ 82 Hz.

To boost the sound in the lower octave, a relatively large circular rose-hole is cut into the front plate of the guitar and two symmetrically facing f-holes are cut into the front plate of instruments of the violin family. The air inside the enclosed volume of the shell of such instruments vibrates in and out through these openings to form a Helmholtz resonator.

The frequency of an ideal Helmholtz cavity resonator of volume V, with a hole of area S in one of its rigid walls is given by

$$\omega_{\mathrm{H}} = \sqrt{\frac{\gamma P}{\rho}\frac{S}{L'V}} = c_0\sqrt{\frac{S}{L'V}}\,, \tag{15.88}$$

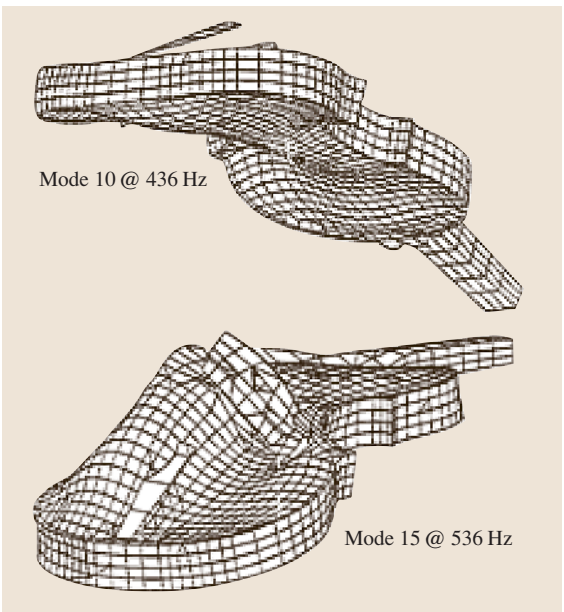


Fig. 15.59 Representative finite element simulations of the structural vibrations of a violin, with greatly exaggerated vibrational amplitudes (after *Knott* [15.96])

Amazon
EX1013 Page 600

Amazon EX1013 Page 601

where L' is the effective length of the open hole. For a circular hole of radius a, *Rayleigh* ([15.3] Vol. 2, Sect. 306) showed that $L' = \frac{\pi}{2}a$, while for an ellipse $L' \sim \frac{\pi}{2}(ab)^{1/2}$, provided the eccentricity is not too large. Noting that the effective length depends largely on area, *Cremer* ([15.29], Fig. 10.6) modelled the f-hole as an ellipse having the same width and area as the f-hole. The two f-holes act in parallel to give an air resonance for the violin at ≈ 270 Hz, at an interval of just over a fifth above the lowest open string. For the acoustic guitar, the circular rose-hole produces an air resonance around 100 Hz, which, like for the violin, is close to the frequency of the second-lowest open string on the instrument.

Any induced motion of the top and bottom plates that involves a net change in volume results in coupling to the Helmholtz mode. Such coupling will perturb the Helmholtz and body-mode frequencies, in just the same way that string resonances are perturbed by coupling to the body resonances (see *Cremer* [15.29], Sect. 10.3 for a detailed discussion of such coupling). Since the acoustically important coupled modes are at considerably higher frequencies than the Helmholtz resonance, the mutual perturbation is not very large. Because of such coupling, purists often object to describing this resonance as a Helmholtz resonance. Similar objections could apply equally well to string resonances, since they too are perturbed by their coupling to body modes. But, as already discussed, in many situations the normal modes largely retain the character of the individually coupled modes other then when their frequencies are close together and, even then, when the damping of either of the coupled modes is large compared to the splitting in frequencies induced by the coupling in the absence of damping (Fig. 15.46).

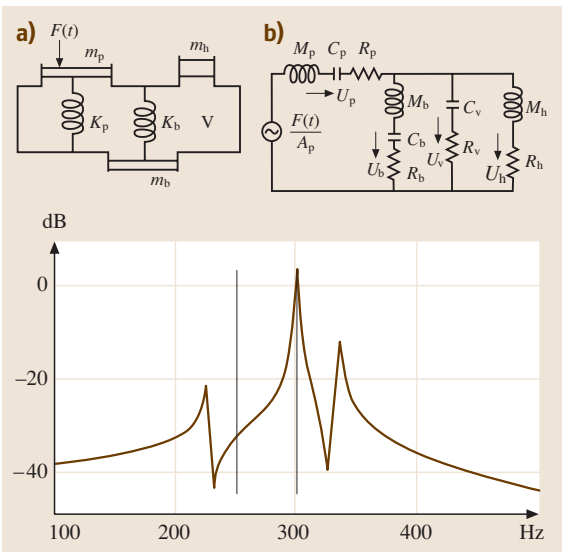


Fig. 15.60a,b The mechanical **(a)** and equivalent electrical **(b)** circuit for a three-mass model describing the vibrations of the front and back plates of a stringed instruments coupled via a Helmholtz resonance (after *Fletcher* and *Rossing* [15.5]). The modulus of the admittance at the top plate has been evaluated for identical front and back plates with uncoupled frequencies of 300 Hz, coupled via a Helmholtz air resonator at 250 Hz in the absence of coupling. The frequencies of the uncoupled air and body resonances are indicated by the *vertical lines*

Well below the Helmholtz resonance, any change in volume of the shell of the violin or guitar induced by the vibrating strings will be matched by an identical volume of air flowing out through the rose- or f-holes, with no net volume flow from the instrument as a whole. Since at low frequencies almost all the radiated sound is monopole radiation associated with the net flow of air across the whole surface of an instrument, little sound will be radiated. However, above the air resonance, the response of the air resonance will lag in phase by 180°, so that the flow from body and cavity will now be in phase, resulting in a net volume flow and strong acoustic radiation. The Helmholtz resonance serves the same purpose as mounting a loudspeaker in a bass-reflex cabinet, with the air cavity resonance boosting the intensity close to and above its resonant frequency.

A number of authors have considered the influence of the enclosed air on the lowest acoustically important modes of the violin (*Beldie* [15.97]) and guitar (*Meyer* [15.98], *Christensen* [15.99] and *Rossing* et al. [15.100]) using simple mechanical modes of interacting springs and masses with damping and their equivalent electric circuits. Figure 15.60 shows the mechanical and equivalent electrical circuits and resulting admittance curve for the top plate for the illustrative three-mass model used by Rossing et al., which accounts for the qualitative features of the first three most important resonances of a guitar body. To emphasise a number of important points, we have calculated the admittance for a cavity with identical front and back plates with uncoupled resonances at 300 Hz, coupled via a cavity Helmholtz resonance at 250 Hz. The closeness in frequencies of the coupled resonators has been chosen to emphasise the influence of the coupling on the modal frequencies.

Without concerning oneself with mathematical detail, one can immediately recognise an unshifted normal mode associated with the uncoupled body resonances

at 300 Hz, corresponding to the two plates vibrating in the same phase, with no volume change and hence no coupling to the air resonance. However, the coupling via the Helmholtz resonance splits the degenerate plate modes, to give a normal mode at a raised frequency, with the plates vibrating in opposite directions in a breathing mode. The coupling also decreases the frequency of the Helmholtz cavity resonance. The unperturbed mode may dominate the measured admittance and affect the playability of the instrument via its perturbation of string resonances. But, because there is no associated volume change, it will be an inefficient acoustic radiator. One should note that, because of the changes in phase of the air resonance on passing through resonance, it appears as a dispersive curve superimposed on the low-frequency wings of the stronger higher-frequency body resonances. The frequency of the excited normal mode is not the peak in the admittance curve (i. e. its modulus), as often assumed but is more nearly mid-way between the maximum and minimum, where its phase lags 90° relative to the phase of the higher frequency normal modes. Similarly, the upper body mode results in a dispersive feature in the opposite sense, as its phase changes from almost 180° to 0° relative to the unshifted normal mode. Very similar, but narrower, dispersive features are also observed in admittance-curve measurements from string resonances, unless they are purposely damped.

Cavity Modes

In addition to the Helmholtz air resonance, there will be many other cavity resonances of the air enclosed within the shell of stringed instruments, all of which can in principle radiate sound through the f- or rose-holes. Alternatively, the internal air resonances can radiate sound via the vibrations they induce in the shell of the instrument, as discussed in some detail by *Cremer* ([15.29], Sect. 11.4). Because of the relatively small height of the violin ribs, below around 4 kHz the cavity air modes are effectively two dimensional in character. Simple statistical arguments based on the overall volume of the violin cavity show that there are typically ≈ 28 resonances below this frequency, as observed in measurements by *Jansson* [15.101]. Whether or not such modes play a significant role in determining the tonal quality of an instrument remains a somewhat contentious issue. However, at a given frequency, the wavelengths of the flexural modes of the individual plates and the internal sound modes will not, in general, coincide. The mutual coupling and consequent perturbation of modes will therefore tend to be rather weak. Even if such coupling were to be significant, it is likely to be far smaller than the major changes in modal frequencies and shapes introduced by the f-holes and soundpost.

Finite-Element Analysis

To progress further in our understanding of the complex vibrations of instruments like the violin and guitar, it is necessary to include the coupled motions of every single part of the instrument and to consider the higher-order front and back plate modes, which will be strongly modified by their mutual coupling via the connecting ribs and, for the violin, the soundpost as well.

Such a task can be performed by numerical simulations of the vibrations of the complete structure using finite-element analysis (FEA). This involves modelling any plate or shell structure in terms of a large number of interconnected smaller elements of known shape and elastic properties. This division into smaller segments is known as tessellation. Provided the scale of the tessellation is much smaller than the acoustic wavelengths at the frequencies being considered, the motion of the structure as a whole can be described by the three-dimensional translations and rotations of the tessellated elements. The motion of each element can be related to the forces and couples acting on the adjoining faces of each three-dimensional (3-D) element. The problem is then reduced to the solution of N simultaneous equations proportional to the number of tessellated elements. Deriving the frequencies and mode shapes of the resulting normal modes of the system involves the inversion of a $N \times N$ matrix. Such calculations can be performed very efficiently on modern computer systems, though the computation time, proportional to N^2, can still be considerable for complex structures, particularly if a fine tessellation is used to evaluate the higher-frequency, shorter-wavelength, modes.

Figure 15.59 has already illustrated the potential complexity of the vibrational modes of a violin. The displacements have been greatly exaggerated for graphical illustration. In practice, the displacements of the plates are typically only a few microns, but can easily be sensed by placing the pad of a finger lightly on the vibrating surfaces. The first example shows a typical low-frequency mode involving the flexing and bending of every part of the violin, but with little change in its volume, so that it will radiate very little sound. The second example illustrates a mode involving a very strong asymmetrical vibration of the front plate, excited by the rocking action of the bridge with the soundpost inhibiting motion on the treble side of the instrument. Such a mode involves an appreciable change in volume of the

Amazon
EX1013 Page 602

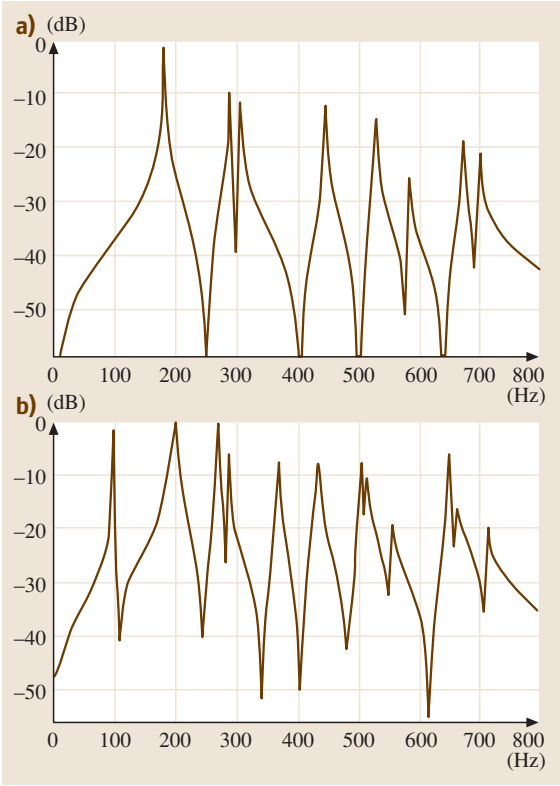


Fig. 15.61a,b FEA computations of admittance at bridge for a guitar with a 2.9 mm-thick front plate **(a)** in vacuo and **(b)** coupled to the air-cavity resonances (after *Derveaux* et al. [15.102])

shell-like structure, which will therefore radiate sound quite strongly.

One of the virtues of FEA is that the effects of changes in design of an instrument, or of the materials used in its construction, can be investigated systematically, without having to physically build a new instrument each time a change is made. For example, *Roberts* [15.103] has used FEA to investigate the changes in normal-mode frequencies of freely supported violin as a function of thickness and arching, the effects of cutting the f-holes and adding the bass bar, and the affect of the soundpost and mass of the ribs on the normal modes of the assembled body of the instrument, but without the neck and fingerboard. This enables a systematic correlation to be made between the modes and frequencies of the assembled instrument with the modes of the individual plates before assembly. Without the soundpost, the modes of the assembled violin were highly symmetric, with the bass-bar having only a marginal effect on the symmetry and frequency of modes. As expected, adding the soundpost removed the symmetry and changed the frequencies of almost all the modes, demonstrating the critical role of the soundpost and its exact position in determining the acoustic response of the violin.

Similar FEA investigations have been made of several other stringed instruments including the guitar (*Richardson* and *Roberts* [15.104]). Of special interest is the recent FEA simulation from first principles of the sound of a plucked guitar string by *Derveaux* et al. [15.102]. Their model includes interactions of the guitar plates, the plucked strings, the internal cavity air modes and the radiated sound. A DVD illustrating the methodology involved in such calculations [15.105] recently won an international prize, as an outstanding example of science communication. The effects of changing plate thickness, internal air resonances and radiation of sound on both admittance curves and the decay times and sound waveforms of plucked strings were investigated. Figure 15.61 compares the admittance curves at the guitar bridge computed for damped strings for a front-plate thickness of 2.9 mm first in vacuo and then in air. Note the addition of the low-frequency Helmholtz and higher-order cavity resonances in air and the perturbations of the structural resonances by coupling to the air modes.

15.2.7 Measurements

In this section we briefly consider the various methods used to measure the acoustical properties of stringed instruments, a number of which have already been referred to illustrate specific topics in the preceding section.

Admittance

The most common and easiest method used to investigate and characterise the acoustical properties of any stringed instrument is to measure the admittance $A(\omega)$ (velocity/force) at the point of string support on the bridge. *Fletcher* and *Rossing* [15.5] give examples of typical admittance curves for many stringed (and percussion) instruments including the violin, guitar, lute, clavichord and piano soundboard.

As described earlier, the admittance is in reality a complex tensor quantity, with the induced velocity dependent on and not necessarily parallel to the direction of the exciting force. In practice, most published admittance curves for the high-bridge instruments of the violin family show the amplitude and phase of the component of induced bridge velocity in the direction of an applied force parallel to the top plate. In contrast, for

Amazon
EX1013 Page 603

Amazon EX1013 Page 604

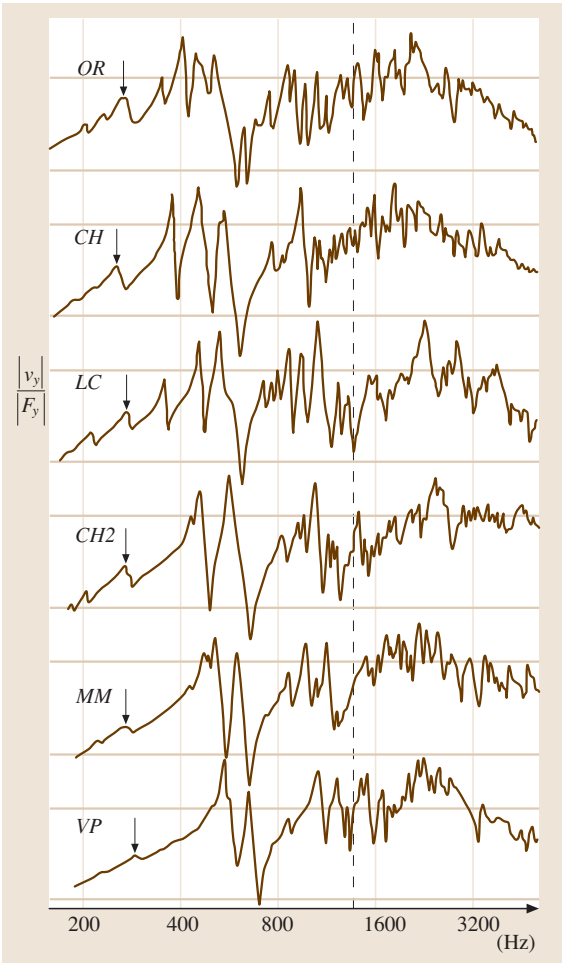


Fig. 15.62 Admittance measurements at bridge for six violins (after *Beldie* [15.97]). The *arrows* indicate the position of the Helmholtz air resonance. The horizontal lines are 20 dB markers

low-bridge instruments like the guitar, piano or harpsichord, the induced motion perpendicular to the top plate or soundboard is of primary interest.

The admittance at the bridge can be expressed in terms of the additive response of all the damped normal modes, which includes the mutual interactions of the plates of the instrument and all the component parts including, where appropriate, the neck, tailpiece, fingerboard, and all the strings. The admittance can then be written as

$$A(\omega) = \sum_n \frac{1}{m_n} \frac{\mathrm{i}\omega}{\omega_n^2 - \omega^2 + \mathrm{i}\omega\omega_n/Q_n} , \qquad (15.89)$$

where ω_n and Q_n are the frequency and Q-value of the nth normal mode and m_n is the associated effective mass at the point of measurement. The value of m_n depends on how well the normal mode is excited by a force at the point of measurement on the bridge. If, for example, the bridge on a guitar is at a position close to the nodes of particular normal modes, then the coupling will be weak and the corresponding effective mass will be rather large. Conversely, the low-frequency rocking action of the bridge on a bowed stringed instrument couples strongly into the breathing mode of the violin shell, so that the effective mass will be relatively low. The strength of this coupling plays an important role in determining the sound output from a particular instrument and also affects the playability of the bowed string and the sound of a plucked string.

In practice, by measuring the frequency response of the admittance, including both amplitude and phase, it is possible to decompose the admittance into the sum of the individual modal contributions and hence determine the effective mass, frequency and damping of the contributing normal modes. For the violin there are ≈ 100 identifiable modes below ≈ 4 kHz (*Bissinger* [15.106]), though not all of these are efficient acoustic radiators.

To avoid complications from the numerous sympathetic string resonances that can be excited, which includes all their higher-frequency partials, measurements are often made with all the strings damped by a piece of soft foam or a piece of card threaded between the strings. However, it should always be remembered that the damped strings still contribute significantly to the measured admittance. At low frequencies the strings still exert the same lateral restoring force on the bridge whether damped or not, while at high frequencies the damped strings present a resistive loading with their characteristic string impedances μc in parallel. When undamped, the strings present an additional impedance and transient response, which reflects the resonances of all the partials of the supported strings. This can make a significant difference to the sound of an instrument, notably when the sustaining pedal is depressed on a piano and in the ringing sound of any multi-stringed instrument, when a note is plucked or bowed and especially instruments like the theorbo and viola d'amore with freely vibrating sympathetic strings.

Figure 15.62 illustrates admittance measurements for six different violins by *Beldie* [15.97] reproduced from *Cremer* ([15.29], Fig. 10.1). The arrows indicate the position of the dispersive-shaped Helmholtz air resonance, which is the only predictable feature in such measurements, though its relative amplitude

varies significantly from one instrument to the next. Such measurements provide a fingerprint for an individual instrument, highlighting the very large number of almost randomly positioned resonances that can be excited, which must ultimately be responsible for the distinctive sound of an instrument. As described earlier, above ≈1500 Hz the admittance often exhibits an underlying BH peak at ≈2 kHz followed by a characteristic decrease at higher frequencies, which can be attributed to a resonance of the bridge/central island region [15.75, 76].

Although most instruments have very different acoustic fingerprints, Woodhouse (private communication), in a collaboration with the English violin maker David Rubio, demonstrated that it is possible to construct instruments with almost identical admittance characteristics, provided one uses closely matching wood from the same log, with a nearly perfect match of plate thickness and arching. The German violin maker *Martin Schleske* [15.107] also claims considerable success in producing *tonal copies* with almost identical acoustical properties to those of distinguished Cremonese instruments by grading the thickness and arching of the plates to reproduce both the input admittance and radiated sound.

In contrast, slavishly copying the dimensions of a master violin rarely produces an instrument with anything like the same tonal quality. This is easily understood in terms of the differing elastic and damping properties of the wood used to carve the plates, which remains a problem of great interest, but beyond the scope of this article.

Traditionally, the admittance is usually measured using a swept sinusoidal frequency source, often generated by a small magnet waxed to the bridge and driven by a sinusoidally varying magnetic field. The admittance can equally well be determined from the transient response $f(t)$ following a very short impulse to the bridge, since it is simply the Fourier transform,

$$A(\omega) = \int_0^\infty f(t)\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{d}t\,. \qquad (15.90)$$

If signal-to-noise ratio from a single measurement is insufficient, one can use a sequence of impulses or a noise source, which is equivalent to a random succession of short pulses. In addition to many professional systems, relatively inexpensive PC-based versions using sound cards have been developed for researchers and instrument makers, such as the WinMLS system by *Morset* [15.108].

Laser Holography

Admittance measurements at the bridge provide detailed information on the frequencies, damping and effective masses of the normal modes of vibration of an instrument, but provide no information on the nature of the modes excited. Laser holography, which is essentially the modern-day equivalent of Chladni plate measurements, enables one to visualise the vibrational modes of stringed and percussion instruments. In such measurements, photographs or real-time closed-circuit television images of the interference patterns of laser light reflected from a stationary mirror and from the vibrating object are recorded. Using photographic or electronic/software reconstruction of the original image from the recorded holograms, a 3-D image of the vibrating surface is formed with superimposed contours indicating lines of equal vibrational amplitude, as already illustrated for a number of prominent guitar modes in Fig. 15.56.

Recent developments in laser and electronic data-acquisition technology allow one to record such interferograms electronically and to display them in real time on a video monitor (for example, *Saldner* et al. [15.109]). To record the shapes of individual vibrational modes of an instrument excited by a sinusoidal force, care has to be taken to avoid contamination from neighbouring resonances, by judicious placing of the force transducer (e.g. placing it at a node of an unwanted mode). *Cremer* ([15.29], Chap. 12) reproduces an interesting set of holograms by *Jansson* et al. [15.110] for the front plate of a violin at various stages of its construction, before and after the f-holes are cut, before and after the bass-bar is added and with a soundpost supported on a rigid back plate. These highlight the major effect of the f-holes and soundpost on the modal shapes and frequencies, but the relatively small influence of the bass-bar, consistent with the FEA computations by *Roberts* [15.103] referred to earlier. However, the bass bar strengthens the coupling between the island area between the f-holes and the larger radiating surfaces of the top plate and therefore has a strong influence on the intensity of radiated energy.

With modern intense pulsed laser sources, one can also investigate the transient response of instruments using single pulses. For example, *Fletcher* and *Rossing* ([15.5], Fig.10.15) reproduce interferograms of the front and back plate of a violin by *Molin* et al. [15.111, 112], which illustrate flexural waves propagating out from the feet of the bridge on the front face and from the end of the soundpost on the back plate at intervals from 100–450 μs after the application of a sharp impulse at

Amazon
EX1013 Page 605

the bridge. Holograms can even be recorded while the instrument is being bowed [15.111, 112].

Modal Analysis

Modal analysis measurements have been extensively used to investigate the vibrational modes of the violin, guitar, the piano soundboards and many other stringed and percussion instruments (Chap. 28). Briefly, the method involves applying a known impulse at one point and measuring the response at a large number of other points on the surface of an instrument, which allows one to determine both the modal frequencies and the vibrations at all points on the surface. The Fourier transform of the impulse response is directly related to the nonlocal admittance, which in terms of the normal modes excited can be written as

$$A(r_1, r_2, \omega) = \mathrm{i}\omega \sum_n \frac{1}{m_n} \frac{\psi_n(r_1)\psi_n(r_2)}{\omega_n^2 - \omega^2 + \mathrm{i}\omega\omega_n/Q_n} , \quad (15.91)$$

where $\psi_n(r_1)\psi_n(r_2)$ is the product of the wavefunctions describing the displacements at the measurment and excitation points normalised to the product at the point of maximum displacement, and m_n is now the effective mass of the normal mode at its point of maximum amplitude of vibration (i. e. $KE_{\max} = 1/2 m_n \omega^2 \psi_n^2|_{\max}$). An FFT of the recorded transient response will give peaks in the frequency response, which can be decomposed into contributions from all the excited normal modes. By keeping the point of excitation fixed and moving the measurement point, one can record the amplitude and phase of the induced motion for a specific mode and, using the spatial dependence in (15.91), can map out the nodal waveform. Alternatively, one can keep the measurement point fixed and apply the impulse over the surface of the structure to derive similar information.

One of the first detailed modal analysis investigations of the violin was made by *Marshall* [15.93], who used a fixed measurement point on the top plate of the violin near the bass-side foot of the bridge with a force hammer providing calibrated impulses at a large number of points over the surface of the violin. From the FFT of the resultant transient responses, the amplitudes and phases of the excited normal modes of the violin involving all its component parts including the body shell, neck, fingerboard and tailpiece could be determined. Marshall identified and characterised around 30 normal modes below ≈1 kHz. Many of the modes involved the relatively low-frequency flexing and twisting of the instrument as a whole. However, because such modes involved little appreciable change in overall volume of the shell of the instrument structure, they resulted in little radiated sound. Nevertheless, it was suggested that such modes might well play an important role for the performer in determining the feel of the instrument and its playability.

In any physical measurement, the instrument has to be supported in some way. Rigid supports introduce additional boundary conditions, which can significantly perturb the normal modes of the instrument. Many measurements are made with the instrument supported by rubber bands, which provide static stability without significant perturbation of the higher-frequency structural modes. However, *Marshall* [15.93] showed that, when an instrument is held and supported by the player under the chin, the damping of many of the normal modes was significantly increased, which will clearly affect the sound of the instrument when played. This observation has also been confirmed in more recent modal analysis measurements by *Bissinger* [15.94] and by direct measurements of the decaying transient sound of a freely and conventionally supported violin by the present author [15.18].

Bissinger [15.106] has made extensive admittance, modal analysis and sound radiation measurements on a large number of instruments. Measurements were made using impulsive excitation at the bridge and a laser Doppler interferometer to record the induced velocities at over 550 points on the surface of the violin. Simultaneous measurements of the overall radiation and directivity were made using 266 microphone positions over a complete sphere. Figure 15.63 shows cross sections illustrating the displacements associated with four low frequency modes. The 159 Hz mode CD-ROM involves major vibrations of the neck, fingerboard, tailpiece and body of the instrument. The second example at 281 Hz CD-ROM illustrates the body displacements associated with the Helmholtz air resonance. The mode

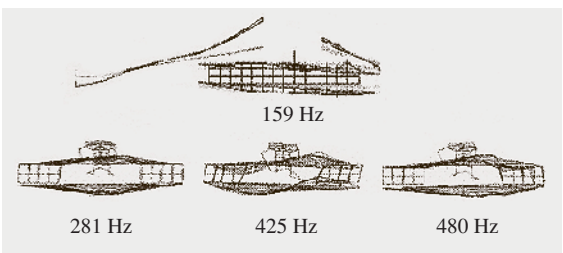


Fig. 15.63 Modal analysis measurements illustrating the displacements associated with four representative low-frequency modes of a violin (data provided by Bissinger)

Amazon
EX1013 Page 606

at 425 Hz CD-ROM illustrates a mode with asymmetric in-phase vibrations of the front and back plates, with little net volume change and hence little radiated sound, while the mode at 480 Hz CD-ROM is a strong breathing mode.

By combining the modal analysis and radiativity measurements, Bissinger has shown that the radiation efficiency of the plate modes (i. e. the fraction of sound energy radiated by the violin relative to a baffled piston of the same surface area having the same root mean square surface velocity displacement) rises to nearly 100% at a *critical frequency* of ≈2–4 kHz, when the wavelength of the flexural vibrations of the plates matches that of sound waves in air. Little apparent correlation was observed between the perceived quality of the measured violins and the frequencies and strengths of prominent structural resonances below ≈1 kHz or with the internal damping of the front and back plates. This runs contrary to the general view of violin makers that the front and back plates of a fine violin should be made of wood with a very long ringing time when tapped. Interestingly, the American violin maker Joseph Curtin has also observed that individual plates of old Italian violins often appear to be more heavily damped than their modern counterparts [15.113]. This is clearly an area that merits further research.

15.2.8 Radiation and Sound Quality

As already emphasised, at low frequencies, when the acoustic wavelength is smaller than the size of an instrument, the radiated sound is dominated by isotropic monopole radiation. As the frequency is increased, higher-order dipole and then quadrupole radiation become progressively important, while above the critical frequency, when the acoustic wavelength is shorter than the that of the flexural waves on the shell of an instrument, the radiation patterns become increasingly directional, so that it is no longer appropriate to consider the radiation in terms of a multipole expansions.

Fletcher and *Rossing* ([15.5], Fig. 10.30) reproduce measurements on both the violin and cello by *Meyer* [15.114], which highlight the increasing directionality of the sound produced with increasing frequency and the rather strong masking effect of the player at high frequencies. More recently, *Weinreich* and *Arnold* [15.115, 116] have made detailed theoretical and experimental studies of multipole radiation from the freely suspended violin at low frequencies (typically below 1 kHz). Interestingly, they made use of the principle

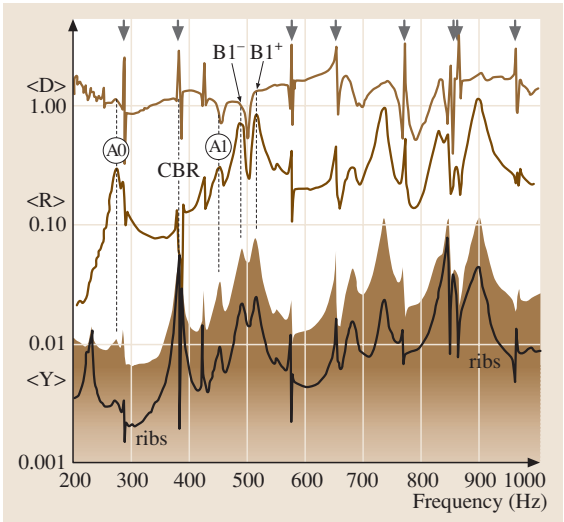


Fig. 15.64 Plots of the root-mean-square (rms) mobility $\langle Y \rangle$ (m/sN) averaged over the surface of the front and back plate (*solid shaded region*) and ribs (*white curve*) of the instrument, the radiativity $\langle R \rangle$ (Pa/N) averaged over a sphere, and directivity $\langle D \rangle$, the ratio of forward to backward radiation averaged over hemispheres. The *top arrows* represent the positions of the open-string resonances and their partials. A0 is the position of the Helmholtz air resonance and A1 the first internal cavity air resonance, CBR is a strong corpus bending mode and B1 and B2 are the two strong structural normal mode resonances of the coupled front and back plates (data kindly supplied by Bissinger)

of acoustic reciprocity, based on the fact that the amplitude of vibration at the top of the bridge produced by incoming sound waves is directly related to the sound radiated by a force applied to the violin at the same point. The violin was radiated by an array of loudspeakers to simulate incoming spherical or dipole sound fields and the induced velocity at the bridge recorded by a very light gramophone pick-up stylus.

Hill et al. [15.117] have used direct measurements to investigate the angular dependence of the sound radiated by a number of high-quality modern acoustic guitars with different cross-strutting, when excited by a sinusoidal force at the bridge. From such measurements, they were able to derive the fraction of sound radiated as the dominant monopole and dipole (with components in three directions) radiation, in addition to effective masses and Q-values, for a number of prominent modes up to ≈600 Hz. Significant differences were observed for the three different strutting schemes investigated.

Amazon
EX1013 Page 607

Bissinger [15.118] has made an extensive investigation of radiation from both freely supported and hand-held violins, including measurements above 1 kHz, where the multimodal radiation expansion is no longer appropriate. Bissinger correlates the sound radiated over a large number of points on a sphere surrounding the violin with measurements of the input admittance at the bridge and the induced surface velocities over the whole violin structure. Figure 15.64 shows a typical set of simultaneous measurements up to 1 kHz. Although the low-frequency Helmholtz resonance contributes strongly to the radiated sound, it results in a relatively small feature on the mobility curves for the body of the instrument (or on the measured admittance at the bridge, not shown). Bissinger was unable to find any significant correlation between the frequencies and Q-values of the prominent *signature* modes excited (see, for example, *Jansson* [15.119]) below $\approx$ 1 kHz and the perceived quality of the instruments investigated. Above $\approx$ 1 kHz the modes strongly overlap, so that it becomes more appropriate to compare the frequency averaged global features. The measurements show that the fraction of mode energy radiated increases monotonically from close to zero at low frequencies up to around almost 100% efficient at 4 kHz and above, where almost all the energy is lost by radiation rather than internal damping. The ultimate aim of these detailed modal analysis studies is to correlate the measured acoustical properties with the results obtained from finite-element analysis and to produce sufficient information about the acoustical properties that might allow a more realistic comparison between physical properties and the properties of an instrument judged from their perceived sound quality and playability.

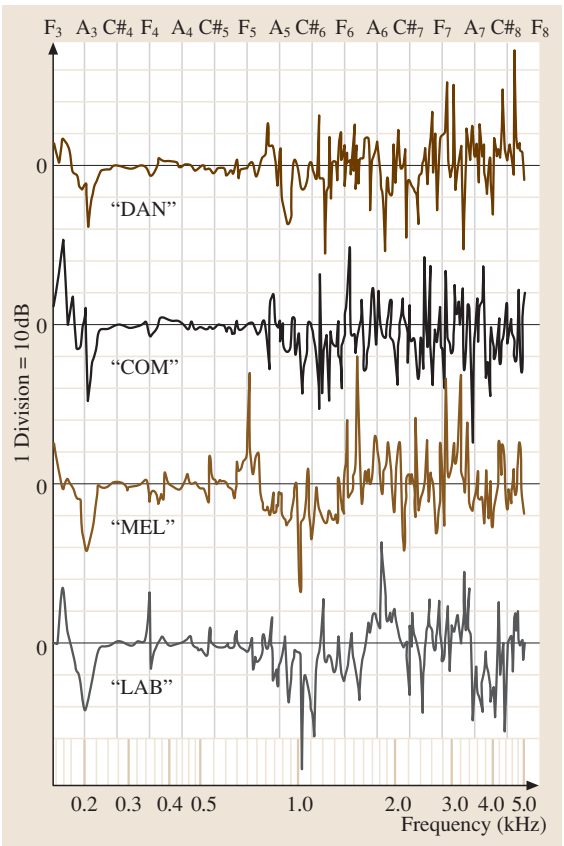


Fig. 15.65 Ratio of sound intensities along neck and perpendicular to front plate for four violins of widely different qualities (after *Weinreich* [15.120])

Directional Tone Colour

Although the acoustic power radiated averaged over a given frequency range is clearly an important signature of the sound of a particular instrument, the intensity at a particular frequency in the important auditory range above 1 kHz can vary wildly from note to note. This is illustrated in Fig. 15.65 by *Weinreich* [15.120], which compares the intensities of radiated sound in an anechoic chamber along the direction of the neck and perpendicular to the front plate for four violins of widely differing quality, with 0 dB representing an isotropic response. Above around 1 kHz, the wavelengths of flexural waves on the plates of the instrument become comparable with the wavelength of the radiated sound. This leads to strong diffraction effects in the radiated sound, which fluctuate wildly with direction as different modes are preferentially excited. At a particular point in the listening space, the spectral content of the bowed violin therefore varies markedly from note to note, as will the sound within a single note played with vibrato resulting in frequency modulation. The spectral content will also vary from position to position around the violin, especially if the player moves whilst playing. Weinreich has emphasised the importance of such effects in producing a greater sense of *presence* and *vibrancy* in the perceived sound from a violin than would be produced by a localised isotropic sound source, such as a small loudspeaker. Weinreich has coined the term *directional tone colour* to describe such effects. He has also designed a loudspeaker system based on the same principles, which gives a greater sense of realism to the sound of the recorded violin than a simple loudspeaker.

In addition to the intrinsic directionality of the violin, the time-delayed echoes from the surrounding walls

Amazon
EX1013 Page 608

of a performing space also have a major influence on the complexity of the sound waveforms produced by a violin (or any other instrument) played with vibrato, as first noted by *Meyer* [15.121]. This arises from the interference between the different frequencies associated with the prompt sound reaching the listener (or microphone) and the sound generated at earlier times reflected from the surrounding surfaces. As discussed by the present author (*Gough* [15.18]), the additional complexity is largely a dynamic effect associated with the time-delayed interference between signals of different frequencies rather than caused by the amplitude modulation of individual partials associated with the multi-resonant response of the violin, first highlighted by *Fletcher* and *Sanders* [15.122] and *Matthews* and *Kohut* [15.123].

Perceived Quality

No problem in musical acoustics has attracted more attention or interest than the attempts made over the last 150 or so years to explain the apparent superiority of old Cremonese violins, such as those made by Stradivarius and Guarnerius, over their later counterparts, which have often been near exact copies. Many explanations have been proposed – a magic recipe for the varnish, chemical treatment of the wood and finish of the plates prior to varnishing [15.124], the special quality of wood resulting from micro-climate changes [15.125], etc. However, despite the committed advocacy for particular explanations by individuals, there is, as yet, little agreement between researchers, players or dealers on the acoustical attributes that distinguish a fine Italian violin worth $1M or more from that of a $100 student instrument.

From a physicist's point of view, given wood of the same quality as that used by the old Italian makers, there is no rational reason why a modern violin should not be just as good from an acoustic point of view as the very best Italian instrument. We have already commented on Martin Schleske's attempts to replicate the sounds of fine Italian instruments, by making *tonal copies* having as near as possible the same acoustical properties [15.107]. In addition, we have also highlighted Dünnewald's attempt to correlate the physical properties of well over 200 violins with their acoustical properties [15.73], including the comparison of selected student, modern and master violins reproduced in Fig. 15.62. Such studies appear to show a correlation between the amount of sound radiation in the acoustically important range around 3–4 kHz. As we have emphasized, this is just the region where the resonant properties of the bridge have a major influence on the spectrum.

It must also be remembered that the changed design of the bridge, the increase in string tension, higher pitch, increased size of the bass-bar, neck and soundpost, and the use of metal-covered rather than gut strings have resulted in a modern instrument sounding very different from the instruments heard by the 17th and early 18th century maker and performer. Even amongst the violins of the most famous Cremonese luthiers, individual instruments have very different distinctive tones and degrees of playability, particular as judged by the player. The gold standard itself is therefore very elusive. What is currently and may always be lacking is reliable measurements on the individual plates and shells of a large number of really fine instruments. We still largely rely on a small number of measurements performed by Savart in the nineteenth century and a few measurements by Saunders in the 1950s on which to base scientific guidelines for modern violin makers.

Performance Acoustic

It should also be recognised that, when a violin (or any other instrument) is played, the performer excites not only the vibrational modes of the instrument but also the multi-resonant normal modes of the performance space. Whereas the sound heard by a violinist is dominated by the sound of the violin, for the listener the acoustics of the performance space can dominate the timbre and quality of the perceived sound. To distinguish between the intrinsic sound qualities of violins, comparisons should presumably best be made in a rather dry acoustic, even though such an acoustic is generally disliked by the performer (and listener), who appreciates the improvement in sound quality provided by a resonant room acoustic.

One cannot review progress towards our understanding of what makes a good violin without recognising the inspiration and enthusiastic leadership of Carleen Hutchins, the doyenne and founder of the Catgut Society and scientific school of violin making, which has attracted many followers world-wide. By matching the frequencies and shapes of the first few modes of free plates before they are assembled into the completed instrument (Fig. 15.55), the scientific school of violin makers clearly achieve a high degree of quality control, which goes some way towards compensating for the inherent inhomogeneities and variable elastic properties of the wood used to carve the plates. However, in practice, there is probably just as much variability in the sound of such instruments as there is in instruments made by more traditional methods, where makers tap, flex and bend the plates until they feel and sound about right, this being part of the traditional skills handed down

Amazon
EX1013 Page 609

from master to apprentice even today. That is certainly the way that the Italian masters must have worked, without the aid of any scientific measurements beyond the feel and the sound of the individual plates as they are flexed and tapped.

Scientific Scaling

The other interesting development inspired by Carleen Hutchins and her scientific coworkers Schelling and Saunders has been the development of the modern violin octet [15.126], a set of eight instruments designed according to physical scaling laws based on the premise that the violin is the most successful of all the bowed stringed instruments. The aim is to produce a consort of instruments all having what are considered to be the optimised acoustical properties of the violin. Each member of the family is therefore designed to have the frequencies of the main body and Helmholtz resonances in the same relationship to the open strings as that on the violin, where the Helmholtz air resonance strongly supports the fundamental of notes around the open D-string, while the main structural resonances support notes around the open A-string and the second and generally strongest partial of the lowest notes played on the G-string. Several sets of such instruments have been constructed and admired in performance, though not all musicians would wish to sacrifice the diversity and richness of sounds produced by the different traditional violin, viola, cello and double bass in a string quartet or orchestra. Nevertheless, the scaling methods have led to rather successful intermediate and small-sized instruments.

15.3 Wind Instruments

In this section we consider the acoustics of wind instruments. These are traditionally divided into the woodwind, played with a vibrating reed or by blowing air across an open hole or against a wedge, and brass instruments, usually made of thin-walled brass tubing and played by buzzing the lips inside a metal mouthpiece attached to the input end of the instrument.

In general, the playing pitch of woodwind instruments is based on the first two modes of the resonating air column, with the pitch changed by varying the effective length by opening and closing holes along its length. In contrast, brass players pitch notes based on a wide range of modes up to and some times beyond the 10th. The effective length of brass instruments can be changed by sliding interpenetrating cylindrical sections of tubing (e.g. the trombone) or by a series of valves, which connect in additional length of tubing (e.g. trumpet and French horn). The pitch of many other instruments, such as the organ, piano-accordion and harmonium, is determined by the resonances of a set of separate pipes or reeds to excite the full chromatic range of notes, rather like the individual strings on a piano.

A detailed discussion of the physics and acoustical properties underlying the production of sound in all types of wind instruments is given by *Fletcher* and *Rossing* [15.5], which includes a comprehensive list of references to the most important research literature prior to 1998. As in many fields of acoustics, *Helmholtz* [15.127] and *Rayleigh* [15.3] laid the foundations of our present-day understanding of the acoustics of wind instruments. In the early part of the 20th century, *Bouasse* [15.128] significantly advanced our understanding of the generation of sound by the vibrating reed. More recently, *Campbell* and *Greated* [15.129] have written an authoritative textbook on musical acoustics with a particular emphasis on musical aspects, including extensive information on wind and brass instruments. Recent reviews by *Nederveen* [15.130] and *Hirschberg* et al. [15.131] provide valuable introductions to recent research on both wind and brass instruments. Earlier texts by *Backus* [15.132] and *Benade* [15.133], both leading pioneers in research on wind-instrument acoustics, provide illuminating insights into the physics involved and provide many practical details about the instruments themselves. A recent issue of *Acta Acustica* [15.134] includes a number of useful review articles, especially on problems related to the generation of sound by vibrating reeds and air jets and on modern methods used to visualise the associated air motions. For a mathematical treatment of the physics underlying the acoustics of wind instruments, *Morse* and *Ingard* [15.135] remains the authoritative modern text. Other important review papers will be cited in the appropriate sections, and selected publications will be used to illustrate the text, without attempting to provide a comprehensive list of references.

We first summarise the essential physics of sound propagation in air and simple acoustic structures before considering the more complicated column shapes used for woodwind and brass instruments. An introduction is then given to the excitation of sound by vibrating lips and reeds, and by air jets blown over a sharp edge.

Amazon
EX1013 Page 610

The physical and acoustical properties of a number of woodwind and brass instruments will be included to illustrate the above topics.

A brief introduction to freely propagating sound in air was given in Sect. 15.1.3. In this section, we will be primarily concerned with the propagation of sound in the bores of wind and brass instruments, the excitation of standing-wave modes in such bores, the mechanics involved in the excitation of such modes and the resultant radiation of sound.

15.3.1 Resonances in Cylindrical Tubes

Standing waves in cylindrical tubes with closed or open ends provide the simplest introduction to the acoustics of wind instruments. For example, the flute can be considered as a first approximation as a cylindrical tube open at both ends, while the clarinet and trombone are closed at one end by the reed or the player's lips. For a cylindrical pipe open at both ends, wave solutions are of the general form

$$p_n(x,t) = A\sin(k_x x)\sin(\omega_n t)\,, \tag{15.92}$$

with the acoustic pressure zero at both ends. Neglecting end-corrections, open ends are therefore displacement antinodes and pressure nodes. These *boundary conditions* result in eigenmodes with $k_n = n\pi/L$ and $\omega_n = nc_0\pi/L$, where L is the length of the pipe and n is an integer.

Such modes are closely analogous to the transverse standing-wave solutions on a stretched string having n half-wavelengths along the length L and a *harmonic* set of frequencies $f_n = nc_0/L$, which are integral multiples of the *fundamental* (lowest) frequency $f_1 = c_0/2L$. When a cylindrical pipe open at both ends, such as a flute, is blown softly, the pitch is determined by the fundamental mode, but when it is *overblown* the frequency doubles, with the pitch stabilising on the second mode an octave above (audio CD-ROM).

A cylindrical pipe played by a reed or vibrating lips has a pressure antinode and displacement node at the playing end. This results in standing-wave solutions with an odd number of 1/4-wavelengths between the two ends, such that $k_n = n\pi/4L$, where n is now limited to odd integer values. The corresponding modal frequencies, $\omega_n = n\pi/4L$, are therefore in the ratios 1:3:5:7: etc. The lowest note on the cylindrical bore clarinet, closed at one end by the mouthpiece, is therefore an octave below the lowest note on a flute of the same length. Furthermore, when overblown, the clarinet sounds a note three times higher than the fundamental, musical interval of an octave plus a perfect fifth (audio CD-ROM). The weak intensity of the even-n-value modes in the spectrum accounts for the clarinet's characteristic hollow sound, particularly for the lowest notes on the instrument.

Real Instruments

For real wind and brass instruments, the idealised model of cylindrical tube resonators is strongly perturbed by a number of important factors. These include:

1. the shape of the internal bore of an instrument, which is often non-cylindrical including conical and often flared tubes with a flared bell on the radiating end,
2. the finite terminating impedance of the reed or mouthpiece used to excite the resonances, no longer providing a perfect displacement node,
3. radiation of sound from the end of the instrument, which is therefore no longer a perfect displacement antinode,
4. viscous and thermal losses to the walls of the instrument,
5. open and shut tone holes in the sides of wind instruments used to vary the pitch of the sounded note, and
6. bends and valves along the length of brass instruments, connecting additional lengths of tubing, which allow the player to play all the notes of a chromatic scale within the playing range of the instrument.

The skill of wind-instrument makers lies in their largely intuitive understanding of the way that changes in bore shape and similar factors affect the resonant modes of an instrument. This allows the design of instruments that retain, as closely as possible, a full set of harmonic resonances across the whole playing range of the instrument. This facilitates the stable production of continuous notes by the player, as the resulting harmonic set of Fourier components or partials coincide with the natural resonances of the instrument. For brass instruments with a flaring end this can often be achieved for all but the lowest natural resonance of the air column.

In discussing the acoustics of wind instruments with variable cross-sectional area S, the flow rate $U = Sv$ is a more useful parameter than the longitudinal particle velocity v. For example, the force acting on an element of air of length Δx along the bore length of an air column is then given by

$$-S\frac{\partial p}{\partial x}\Delta x = \rho S\frac{\partial v}{\partial t}\Delta x = \rho\frac{\partial U}{\partial t}\Delta x\,. \tag{15.93}$$

Amazon
EX1013 Page 611

For travelling waves, $e^{i(\omega t \pm kx)}$, this results in a ratio between the pressure and flow rate, defined as the tube impedance

$$Z = \frac{p}{U} = \pm \frac{\rho c_0}{S} , \qquad (15.94)$$

where the plus and minus signs refer to waves travelling in the positive and negative x-directions, respectively. There is a very close analogy with an electrical transmission line, with pressure and flow rate the analogue of voltage and current, as discussed later. Because the impedance is inversely proportional to area, it can be appreciably higher at the input end of a brass or wind instrument than at its flared output end. The flared bore of a brass instrument or the horn on an old wind-up gramophone can therefore be considered as an acoustic transformer, which improves the match between the high impedance of the vibrating source of sound to the much lower impedance presented by air at the end of the instrument. There is clearly an optimum matching, which enhances the radialed sound without serious degredation of the excited resonant modes.

Acoustic Radiation

In elementary textbook treatments, the pressure at the end of an open pipe is assumed to be zero and the flow rate a maximum, so that $Z_{\text{closed}} = p/U = 0$. However, in practice, the oscillatory motion of the air extends somewhat beyond the open end, providing a pulsating source that radiates sound, as described by *Rayleigh* ([15.3] Vol. 1, Sect. 313). Such effects can be described by a complex terminating load impedance, $Z_L = R + jx$. Figure 15.66 shows the real (*radiation resistance*) R and imaginary (inertial *end-correction*) x components of Z_L as a function of ka, where a is the radius of the open-ended pipe. The impedance is normalised to the tube impedance $\rho c_0/\pi a^2$.

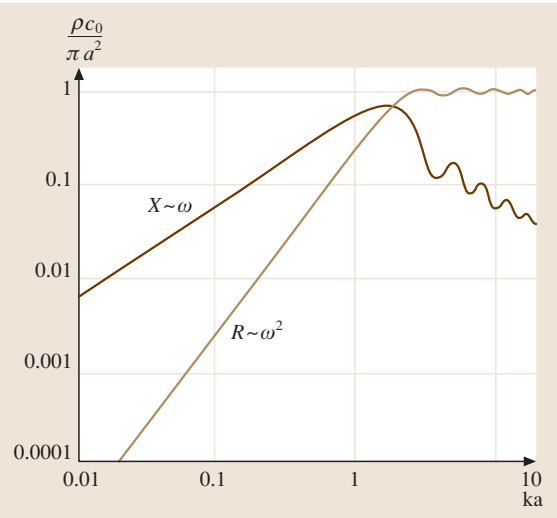


Fig. 15.66 Real and imaginary components of f, the impedance at the unbaffled open end of a cylindrical tube of radius a, in units of $\rho c_0/\pi a^2$, as a function of ka (after *Beranek* [15.136])

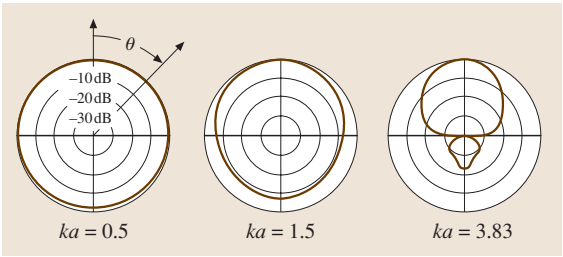


Fig. 15.67 Polar plots of the intensity radiation from the end of a cylindrical pipe of radius a for representative ka values, calculated by *Levine* and *Schwinger* [15.137]. The radial gradations are in units of 10 dB. The intensities in the forward direction ($\theta = 0$) relative to those of an isotropic source are 1.1, 4.8 and 11.8 dB (*Beranek* [15.136])

When $ka \ll 1$, the reactive component is proportional to ka and corresponds to an effective increase in the tube length or end-correction of $0.61a$. At low frequencies, the real part of the impedance represents the radiation resistance $R_{\text{rad}} = \rho c/4S(ka)^2$. In this regime, the sound will be radiated isotropically as a monopole source of strength $U e^{i\omega t}$, illustrated in Fig. 15.67 by the polar plots of sound intensity as a function of angle and frequency.

When ka is of the order of and greater than unity, the real part of the impedance approaches that of a plane wave acting across the same area as that of the tube. Almost all the energy incident on the end of the tube is then radiated and little is reflected. For $ka \gg 1$, sound would be radiated from the end of the pipe as a beam of sound waves. The transition from isotropic to highly directional sound radiation is illustrated for a sequence of ka values in Fig. 15.67. The ripples in the impedance in Fig. 15.66 arise from diffraction effects, when the wavelength becomes comparable with the tube diameter.

For all woodwind and brass instruments, there is therefore a crossover frequency $f_c \sim c_0/2\pi a$, below which incident sound waves are reflected at the open end to form standing waves. Above f_c waves generated by the reed or vibrating lips will be radiated from the ends of the instrument strongly with very little reflec-

Amazon
EX1013 Page 612

tion. Narrow-bore instruments can have a large number of resonant modes below the cut-off, while instruments with a large output bell, like many brass instruments, have far fewer.

For a narrow-bore cylindrical bore wind instrument with end radius $a \approx 1$ cm, the cut-off frequency ($ka \approx 1$) is ≈ 5.5 kHz. Below this frequency the instrument will support a number of relatively weakly damped resonant modes, which will radiate isotropically from the ends of the instrument or from open holes cut in its sides. In contrast, for brass instruments the detailed shape and size ot the flared end-bell determines the cut-off frequency. The large size of the bell leads to an increase in intensity of the higher partials and hence brilliance of tone-color, especially when the bell is pointed directly towards the listener. For French horns, much of the higher-frequency sound is therefore projected backwards relative to the player, unless there is a strongly reflecting surface behind.

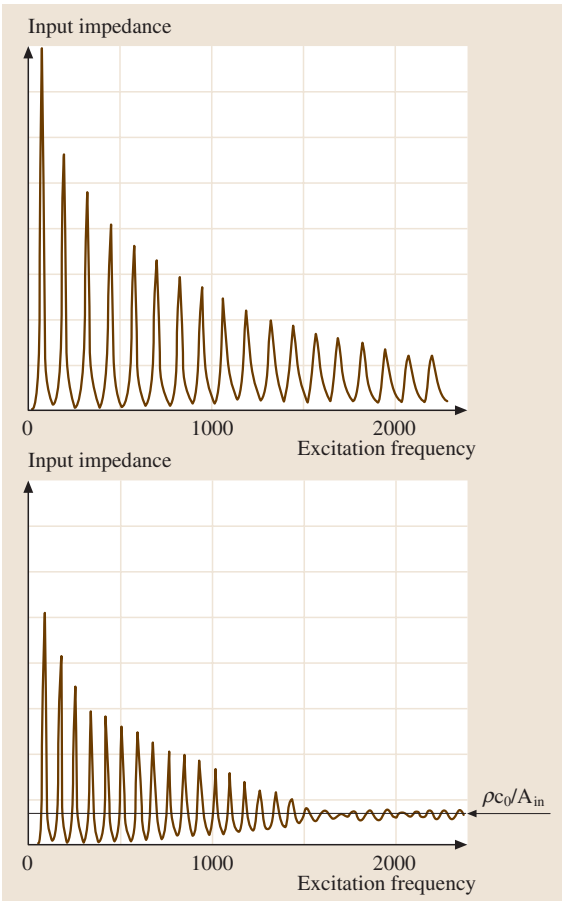


Fig. 15.68 Input impedance of a length of 1 cm diameter trumpet tubing with and without a bell attached to the output end (after *Benade* [15.133])

For $ka \ll 1$, the open end of a musical instrument acts as an isotropic monopole source with radiated power P given by

$$P = U_{\mathrm{rms}}^2 R_{\mathrm{rad}} = \omega^2 \frac{\rho}{8\pi c} (S\omega\xi)^2 \ . \qquad (15.95)$$

For a given vibrational displacement, the radiated power therefore increases with the fourth power of both frequency and radius. This very strong dependence on size explains why brass instruments tend to have rather large bells and why high-fidelity (HI-FI) *woofer* speakers and the horns of public address loudspeakers tend to be rather large. Conversely, it explains why the sound of small loudspeakers, such as those used in PC notebooks, fall off very rapidly below a few hundred Hz.

Acoustic radiation will lower the height and increase the width of resonances in a cylindrical tube. The resulting Q-values can be determined from

$$Q = \omega \frac{\text{stored energy}}{\text{radiated energy}} = \omega \frac{\frac{1}{4}\rho \mathrm{SL}\omega^2\xi^2}{\omega^4 (\rho/8\pi c) S^2\xi^2} = 2\pi cL/\omega S \ . \qquad (15.96)$$

Narrow-bore instruments will therefore have larger Q-values and narrower resonances than wide-bore instruments such as brass instruments, where the flared end-sections enhance the radiated energy at the expense of increasing the net losses.

The increased damping introduced by radiation from the end of an instrument is illustrated in Fig. 15.68, which compares the resonances of a length of 1 cm-diameter trumpet tubing, first with a normal open end and then with a bell attached (*Benade* [15.132], Fig. 20.4). Attaching a bell to such a tube dramatically increases the radiated sound from the higher partials and perceived intensity, but at the expense of a cut-off frequency at around $\approx$1.5 kHz and a significant broadening of the resonances at lower frequencies. Audio CD-ROM demonstrates the sound of a mouthpiece-blown length of hose pipe with and without a conical chemical filter funnel attached to its end.

Viscous and Thermal Losses

In addition to radiation losses, there can be significant losses from viscous damping and heat transfer to the walls, as discussed in detail in *Fletcher* and *Rossing* ([15.5], Sect. 8.2). Although simple models for waves

Amazon
EX1013 Page 613

propagating along tubes assume a constant particle displacement across the whole cross-section, in reality the air on the surface of the walls remains at rest. The particle velocity only attains the assumed plane-wave constant value over a boundary-layer distance of δ_η from the walls. This is determined by the viscosity η, where $\delta_\eta = (\eta/\omega\rho)^{1/2}$, which can be expressed as $\sim 1.6/f^{1/2}$ mm for air at room temperature. At 100 Hz, $\delta_\eta \sim 0.16$ mm. which is relatively small in comparison with typical wind-instrument pipe diameters. Nevertheless, it introduces a significant amount of additional damping of the lower-frequency resonances of wind and brass instruments.

The viscous losses lead to an attenuation of sound waves, which can be described by adding an imaginary component to the k value such that $k' = k - \mathrm{i}\alpha$. Waves therefore propagate as $\mathrm{e}^{-\alpha x}\,\mathrm{e}^{\mathrm{i}(\omega t - kx)}$, with an attenuation coefficient

$$\alpha = \frac{1}{ac_0}\sqrt{\frac{\eta\omega}{2\rho}} = \frac{k\delta_\eta}{a}\,. \qquad (15.97)$$

In addition, heat can flow from the sinusoidally varying adiabatic temperature fluctuations of the vibrating column of air into the walls of the cylinder. At acoustic frequencies, this takes place over the thermal diffusion boundary length $\delta_\theta = (\kappa/\omega\rho C_\mathrm{p})^{1/2}$, where κ is the thermal conductivity and C_p is the heat capacity of the gas at constant pressure. In practice, $\delta_\theta \sim \delta_\eta$, as anticipated from simple kinetic theory (for air, the values differ by only 20%). Viscous and heating losses are therefore comparable in size giving an effective damping factor for air at room temperature, $\alpha = 2.210^4 k^{1/2}/a\ \mathrm{m}^{-1}$ (*Fletcher* and *Rossing* [15.5], Sect. 8.2). The ratio of the real to imaginary components of k determines the damping and effective Q-value of the acoustic resonances from wall losses alone, with $Q_\mathrm{walls} = k/2\alpha$.

The combination of radiation and wall losses leads to an effective Q_total of the resonant modes given by

$$\frac{1}{Q_\mathrm{total}} = \frac{1}{Q_\mathrm{radiation}} + \frac{1}{Q_\mathrm{wall\text{-}damping}}\,. \qquad (15.98)$$

Because of the different frequency dependencies, wall damping tends to be the strongest cause of damping of the lowest-frequency resonances of an instrument. It can also be significant in the narrow-bore tubes and *crooks* used to attach reeds to wind instruments.

Input Impedance

The method used to characterize the acoustical properties of a wind or brass instrument is to measure the input impedance $Z_\mathrm{in} = p_\mathrm{in}/U_\mathrm{in}$ at the mouthpiece or reed end of the instrument. Such measurements are frequently made using the capillary tube method. This involves connecting an oscillating source of pressure fluctuations to the input of the instrument through a narrow-bore tube. This maintains a constant oscillating flow of air into the instrument, which is largely independent of the frequency-dependent induced pressure fluctuations at the input of the instrument. Several examples of such measurements, similar to those for the length of trumpet tubing (Fig. 15.68), for woodwind and brass instruments are shown and discussed by *Backus* [15.132] and *Benade* [15.133], who pioneered such measurements, and in *Fletcher* and *Rossing* ([15.5], Chap. 15). Alternatively, a piezoelectric driver coupled to the end of the instrument can provide a known source of acoustic volume flow.

The input impedance of a cylindrical tube is a function of both the tube impedance $Z_0 = \rho c_0/S$ and the terminating impedance Z_L at its end. It can be calculated using standard transmission-line theory, which takes into account the amplitude and phases of the reflected waves from the terminating load. The reflection and transmission coefficients R and T for a sound wave impinging on a terminating load Z_L are given by

$$R = \frac{Z_\mathrm{L} - Z_0}{Z_\mathrm{L} + Z_0} \quad \text{and} \quad T = \frac{2Z_\mathrm{L}}{Z_\mathrm{L} + Z_0}\,. \qquad (15.99)$$

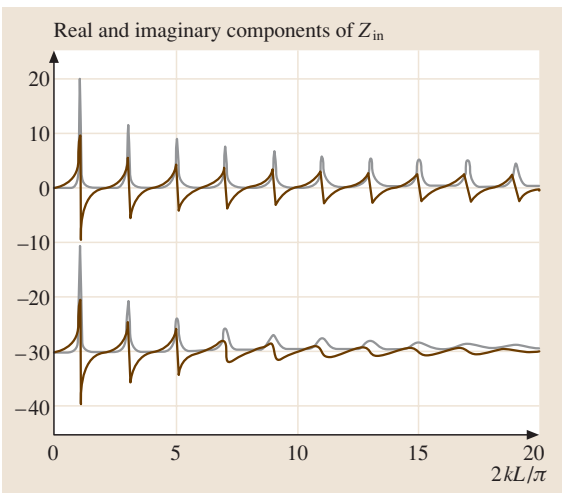


Fig. 15.69a,b Real and imaginary components of the input impedance in units of Z_0 as a function of $2kL/\pi$ for **(a)** an ideally open-ended, $Z_\mathrm{L} = 0$, cylindrical tube with wall losses varying as $k^{1/2}$, and **(b)** the same components shifted downwards for a pipe with radiation resistance proportional to k^2 also included

Amazon
EX1013 Page 614

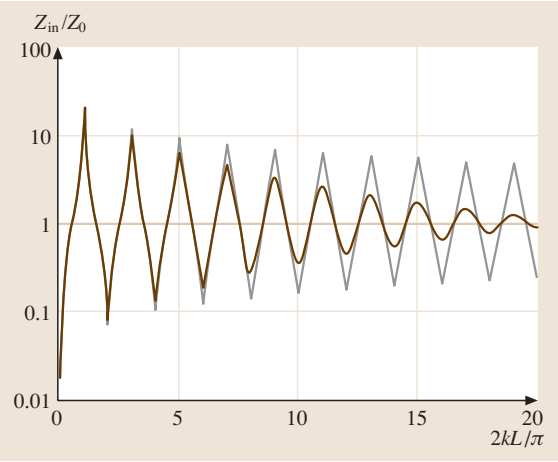


Fig. 15.70 The modulus of the input impedance plotted on a logarithmic scale of cylindrical pipe with wall damping alone and with additional radiative damping as a function of k in units of $\pi/2$

For a cylindrical tube of length L, the input impedance is given by

$$Z_{in} = Z_0 \frac{Z_L \cos kl + iZ_0 \sin kl}{iZ_L \sin kl + Z_0 \cos kl}, \qquad (15.100)$$

with complex k values $k - i\alpha$, if wall losses need to be included.

In Fig. 15.69 we have plotted the kL dependence of the real and imaginary components of the input impedance of an open-ended cylindrical pipe. The upper plot includes wall losses alone proportional to $k^{1/2}$ while the lower plot includes losses from the end of the instrument with radiation resistance $\mathrm{Re}(Z_L)$ varying as k^2. The input impedance is high when $k_n = nc_0/4L$, where n is an odd integer. The input impedance is a minimum when n is even.

It is instructive to consider the magnitude of the input impedances on a logarithmic scale, as shown in Fig. 15.70. When plotted in this way, the resonances and anti-resonances are symmetrically placed about the tube impedance Z_0. The magnitude of the impedance of the nth resonance is $Q_n Z_0$, where Q_n is the quality factor of the excited mode. In contrast the anti-resonances have values of Z_0/Q. The widths $\Delta\omega/\omega$ of the modes at half or double intensity are given by $1/Q_n$.

For efficient transfer of sound, the input impedance of a wind instrument has to match the impedance of the sound generator. For instruments like the flute, recorder and pan pipes, sound is excited by imposed fluctuations in air pressure across an open hole, so that the generator impedance is small. The resonances of such instruments are therefore located at the minima of the anti-resonance impedance dips, corresponding to the evenly spaced resonances, $f_n = nc_0/2L$, of a cylindrical tube with both ends open. In contrast, for all the brass instruments and woodwind instruments played by a reed, the playing end of the tube is closed by a relatively massive generator (the lips or reed). Resonances then occur at the peaks of the input impedance with frequencies $f_n = nc_0/4L$, where n is now an odd integer, corresponding to the resonances of a tube closed at one end. If we had plotted the magnitude of the input admittance, $A(\omega) = 1/Z(\omega)$, instead of the impedance, the positions of the resonances and anti-resonances would have been reversed. The resonant modes of a double-open-ended wind instrument therefore occur at the peaks of the input admittance, whereas the resonant modes of wind or brass instruments played with a reed or mouthpiece are at the peaks of the input impedance. This is a general property of wind instruments, whatever the size or shape of their internal bores.

15.3.2 Non-Cylindrical Tubes

Although there are simple wind instruments with cylindrical bores along their whole length, the vast majority of modern instruments and many ancient and ethnologically important instruments have internal bores that flare out towards their ends. One of the principle reasons for such flares is that they act as acoustic transformers, which help to match the high impedance at the mouthpiece to the much lower radiation impedance of the larger-area radiating output end. However, increasing the fraction of sound radiated decreases the amplitude of the reflected waves and hence the height and sharpness of the natural resonances of the air resonances. In addition, the shape of the bore can strongly influence the frequencies of the resonating air column, which destroys the harmonicity of the modes. This makes it more difficult for the player to produce a continuous note that is rich in partials, since any repetitive waveform requires the excitation of a harmonic set of frequencies.

Conical Tube

We first consider sound propagation in a conical tube, approximating to the internal bore of the oboe, saxophone, cornet, renaissance cornett and bugle. If side-wall interactions are neglected, the solutions for wave propagation in a conical tube are identical to those of spherical wave

Amazon
EX1013 Page 615

propagating from a central point. Such waves satisfy the wave equation, which may be written in spherical coordinates as

$$\nabla^2 (rp) = \frac{1}{c_0^2} \frac{\partial^2 (rp)}{\partial t^2} \,. \tag{15.101}$$

We therefore have standing-wave solutions for rp that are very similar to those of a cylindrical tube, with

$$p = C \frac{\sin kr}{r} \, \mathrm{e}^{\mathrm{i}\omega t} \,. \tag{15.102}$$

Note that the pressure remains finite at the apex of the cone, $r = 0$, where $\sin(kr)/r \to k$. For a conical tube with a pressure node $p = 0$ at the open end, we therefore have standing wave modes with $k_n L = n\pi$ and $f_n = nc_0/2L$, where n is any integer. The frequencies of the excited modes are therefore identical to the modes of a cylindrical tube of the same length that is open at both ends. The lowest note at $f_1 = c_0/2L$ for a conical tube instrument with a reed at one end (e.g. the oboe and saxophone) is therefore an octave above a reed instrument of the same length with a cylindrical bore (e.g. the clarinet) with a fundamental frequency of $c_0/4L$.

The flow velocity U is determined by the acceleration of the air resulting from the spatial variation of the pressure, so that

$$\rho \frac{\partial U}{\partial t} = \frac{\partial (r^2 p)}{\partial r} = C \, (\sin kr + kr \cos kr) \, \mathrm{e}^{\mathrm{i}\omega t} \,. \tag{15.103}$$

Figure 15.71 illustrates the pressure and flow velocity for the $n = 5$ mode of a conical tube. Unlike the modes of cylindrical tube, the nodes of U no longer coincide with the peaks in p, which is especially apparent for the first few cycles along the tube. Furthermore, the amplitude fluctuations increase with distance r from the apex ($\sim r$), whilst the fluctuations in pressure decrease $\approx 1/r$. A conical section therefore acts as an acoustic transformer helping to match the high impedance at the input mouthpiece end to the low impedance at the output radiating end.

Attaching a mouthpiece or reed to the end of a conical tube requires truncation of the cone, which will clearly perturb the frequencies of the harmonic modes. However, using a mouthpiece or reed unit having the same internal volume as the volume of the truncated section removed will leave the frequencies of the lowest modes unchanged. Only when the acoustic wavelength becomes comparable with the length of truncated section will the perturbation be large.

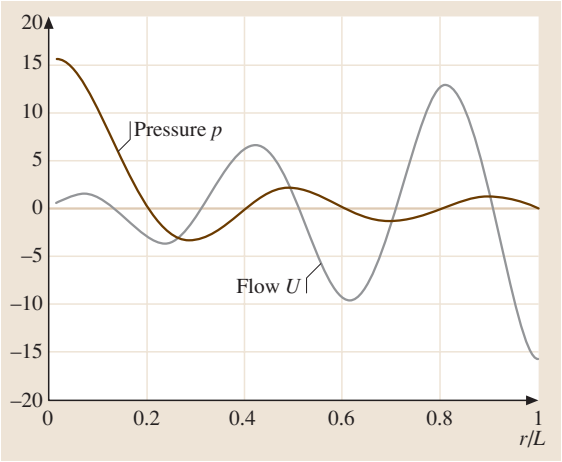


Fig. 15.71 Pressure and flow velocity of the $n = 5$ mode along the length of a conical tube

Fletcher and *Rossing* ([15.5], Sect. 8.7) consider the physics of the truncated conical tube and give the input impedance derived by *Olson* [15.139]

$$Z_{\mathrm{in}} = \frac{\rho c_0}{S_1} \frac{\mathrm{i} Z_{\mathrm{L}} \left(\frac{\sin(kL - \theta_2)}{\sin \theta_2} \right) + \left(\frac{\rho c_0}{S_2} \right) \sin kL}{Z_{\mathrm{L}} \frac{\sin(kL + \theta_1 - \theta_2)}{\sin \theta_1 \sin \theta_2} + j \left(\frac{\rho c_0}{S_2} \right) \frac{\sin(kl + \theta_1)}{\sin \theta_1}} \,, \tag{15.104}$$

where x_1 and x_2 are the distances of the two ends from the apex of the truncated conical section. The length $L = x_2 - x_1$, the end areas are S_1 and S_2, with $\theta_1 = \tan^{-1} kx_1$ and $\theta_2 = tan^{-1} kx_2$.

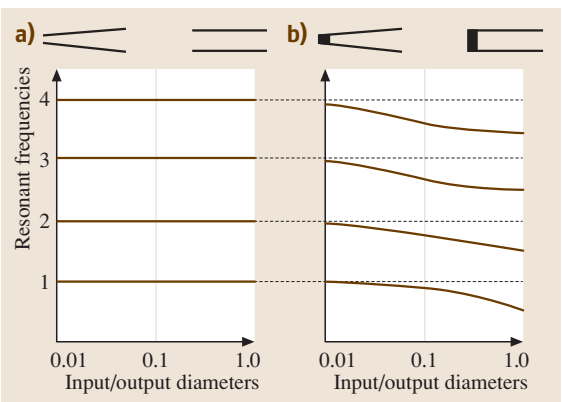


Fig. 15.72a,b The first four resonant frequencies of truncated cones with **(a)** both ends open, and **(b)** the input end closed, as a function of the ratio of their input to output diameters (after *Ayers* et al. [15.138])

Amazon
EX1013 Page 616

For a cone with $Z_L = 0$ at the open end, (15.104) reduces to

$$Z_{in} = -j\frac{\rho c_0}{S_1}\frac{\sin kL \sin\theta_1}{\sin(kL+\theta_1)}, \quad (15.105)$$

which is zero for $kL = n\pi$. The resonant frequencies of truncated cones with both ends open are therefore independent of cone angle and are the same as the equally spaced harmonic modes of a cylinder of the same length with both ends open, as shown in Fig. 15.72a. In contrast, the resonant frequencies of a truncated cone with one end closed (e.g. by the reed of an oboe or saxophone or mouthpiece of a bugle) are strongly dependent on the cone angle or ratio of input to output diameter, as shown in Fig. 15.72b, adapted from *Ayers* et al. [15.138]. As the ratio of input to output diameters of a truncated cone increases, the modes change from the evenly spaced harmonics of an open-ended cylinder of the same length, to the odd harmonics of a cylinder closed at one end. In the transitional regime, the frequencies of the modes are no longer harmonically related. This has a significant effect on the playability of the instrument, as the upper harmonics are no longer coincident with the Fourier components of a continuously sounded note. However, for an instrument such as the oboe, with a rather small truncated cone length, the perturbation of the upper modes is relatively small, as can be seen from Fig. 15.71b.

Cylindrical and non-truncated conical tubes are the only tubes that can produce a harmonically related set of resonant modes, independent of their length. Hence, when a hole is opened in the side walls of such a tube, to reduce the effective length and hence pitch of the note played, to first order, the harmonicity of the modes is retained. This assumes a node at the open hole, which will not be strictly correct, as discussed later in Sect. 15.3.3.

In reality, the bores of wind instruments are rarely exactly cylindrical or conical along their whole length. Moreover, many wind instruments have a small flare at the end to enhance the radiated sound, while others, like the cor anglais and oboe d'amore, have an egg-shaped cavity resonator towards their ends, which contributes to their characteristic timbre or tone colour. Table 15.6 lists representative wind instruments that are at least approximately based on cylindrical and conical bore shapes. The modern organ makes use of almost every conceivable combination of closed- and open-ended cylindrical and conical pipes.

Hybrid Tubes

Although many brass instruments include considerable lengths of cylindrical section, they generally have a fairly long, gently flared, end-section terminated by a very strongly flared output bell to enhance the radiated sound. The shape of such flares can be optimized to preserve the near harmonicity of the resonant modes, as described in the following section.

One can use (15.104) to model the input impedance of a flared tube of any shape, by approximating the shape by a number of short truncated conical sections joined together. Starting from the radiating end, one evaluates the input impedance of each cone in turn and uses it to provide the terminating impedance for the next, until one reached the mouthpiece end.

A weakness of all such models is the assumed plane wavefront across the air column, whereas it must always be perpendicular to the walls and belled outwards in any rapidly flaring region. We will return to this problem later.

Typical brass instruments, like the trumpet and trombone, have bores that are approximately cylindrical for around half their length followed by a gently flared section and end bell, while others have initial conical sections, like the bugle and horn. The affect on the resonant frequencies of the first six modes of adding a truncated conical section to a length of cylindrical tubing is shown as a function their relative lengths in Fig. 15.73, from *Fletcher* and *Rossing* ([15.5], Fig. 8.9). Note the major deviations from harmonicity of the resonant modes, apart from when the two sections are of nearly equal lengths. These results highlight the com-

Table 15.6 Instruments approximately based on cylindrical and conical air columns

		
$f_n = nc_0/2L$ n even and odd	$f_n = nc_0/4L$ n odd	$f_n = nc_0/2L$ n even and odd
Flute Recorders Shakuhachi Organ flue pipes (e.g. diapason)	Clarinet Crumhorn Pan pipes Organ flue pipes (e.g. bourdon) Organ reed pipes (e.g. clarinet)	Oboe Bassoon Saxophone Cornett Serpent Organ reed pipes (e.g. trumpet)

Amazon
EX1013 Page 617

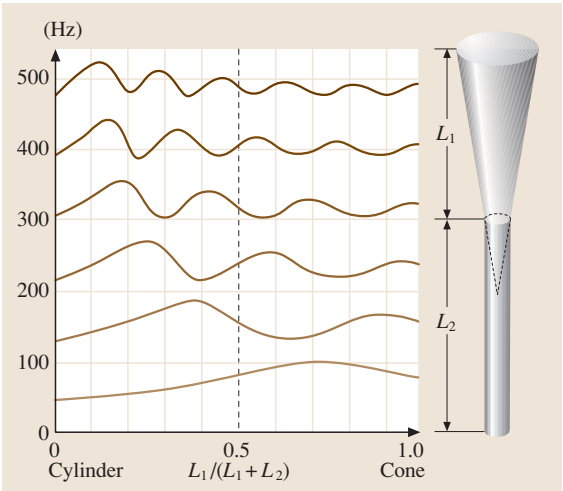


Fig. 15.73 The frequencies of the first six modes of a compound horn formed from different fractional lengths of cylindrical and conical section (after *Fletcher* and *Rossing* [15.5])

plexity involved, when adding flaring sections to brass instruments to increase the radiated sound.

Horn Equation

Physical insight into the influence of bore shape on the modes of typical brass instruments is given by the *horn equation* introduced by *Webster* [15.141], though similar models date back to the time of *Bernoulli* (*Rossing* and *Fletcher* [15.5], Sect. 8.6). In its simplest form, the horn equation can be written as

$$\frac{1}{S}\frac{\partial}{\partial x}\left(S\frac{\partial p}{\partial x}\right)=\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2}\,, \qquad (15.106)$$

where $S(x)$ is the cross-sectional area of the horn at a distance x along its length. Provided the flare is not too

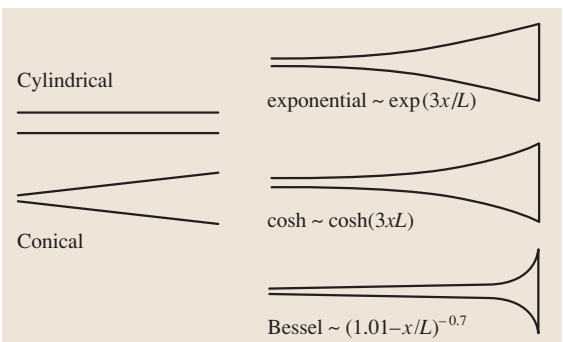


Fig. 15.74 Analytic horn shapes

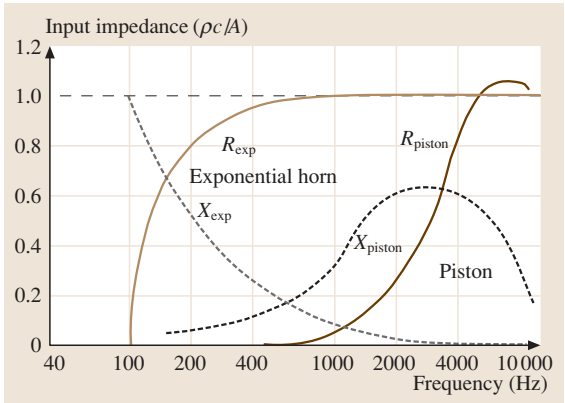


Fig. 15.75 Comparison of input impedance at the input throat of an infinitely long exponential horn and a piston of the same area set into an infinite baffle (after *Kinsler* et al. [15.140])

large, the above plane-wave approximation gives a good approximation to the exact solutions and preserves the essential physics involved.

If we make the substitution $p=\psi S^{1/2}$ and look for solutions varying as $\psi(x)\mathrm{e}^{\mathrm{i}\omega t}$, the horn equation can be expressed as

$$\frac{\partial^2\psi}{\partial x^2}+\left[\left(\frac{\omega}{c_0}\right)^2-\frac{1}{a}\frac{\partial^2 a}{\partial x^2}\right]\psi=0\,, \qquad (15.107)$$

where the radius $a(x)$ is now a function of position along the length. The above equation is closely related to the Schrödinger wave equation in quantum mechanics, with $1/a\;\partial^2 a/\partial x^2$ the analogue of potential energy and $-\partial^2\psi/\partial x^2$ the analogue of kinetic energy $-h^2/2m\;\partial^2\psi/\partial x^2$, where m is the mass of the particle and h is Planck's constant. One can look for solutions of the form $\mathrm{e}^{\mathrm{i}(\omega t\pm kx)}$. At any point along the horn at radius x the radius of curvature of the horn walls, $R=(\partial^2 a/\partial x^2)^{-1}$, so that

$$k^2=\left(\frac{\omega}{c_0}\right)^2-\frac{1}{aR}\,. \qquad (15.108)$$

If $\omega>\omega_\mathrm{c}=c_0/(aR)^{1/2}$, k is real, so that unattenuated travelling and standing-wave solutions exist. However, when $\omega<\omega_\mathrm{c}$, k is imaginary and waves no longer propagate, but are exponentially damped as $\mathrm{e}^{-x/\delta}\mathrm{e}^{\mathrm{i}\omega t}$ with a decay length of $c_0/\left(\omega_\mathrm{c}^2-\omega^2\right)^{1/2}$.

The propagation of sound waves in a horn is therefore directly analogous to the propagation of particle waves in a spatially varying potential. If the curvature

Amazon
EX1013 Page 618

is sufficiently large sound waves will be reflected before they reach the end of the instrument. However, just like particle waves in a potential well, sound waves can still tunnel through the potential barrier and radiate into free space at the end of the flared section. For a horn with a rapidly increasing flare, the reflection point occurs when the wavelength $\lambda^2 \sim (2\pi)Ra$. The effective length of an instrument with a flared horn on its end is therefore shorter for low-frequency modes than for the higher-frequency modes. This is illustrated schematically in Fig. 15.77 for resonant modes of a flared Bessel horn, which will be considered in more detail in the next section.

Exponential Horn

We now consider solutions of the horn equation, for a number of special shapes that closely describe sections of the internal bore of typical brass instrument. Cylindrical and conical section horns are special solutions with $(1/a)\partial^2 a/\partial x^2 = 0$, so that ψ satisfies the simple dispersionless wave equation. Figure 15.74 illustrates a number of other horn shapes described by analytic functions.

The radii of exponential and cosh function horns vary exponentially as $A\,\mathrm{e}^{mx}$ and $A\cosh(mx)$, respectively, so that $(1/a)\partial^2 a/\partial x^2 = m^2$. The $\cosh mx$ function provides a smooth connection to a cylindrical tube at the input end. For both shapes, the horn equation can then be written as

$$\frac{\partial^2\psi}{\partial x^2} + \left[\left(\frac{\omega}{c_0}\right)^2 - m^2\right]\psi = 0\,, \qquad (15.109)$$

which has travelling solutions for the sound pressure $p = \psi/S^{1/2}$, where

$$p(x) = \mathrm{e}^{-mx}\,\mathrm{e}^{\mathrm{i}\left(\omega t - \sqrt{k^2 - m^2 x^2}\right)}\,, \qquad (15.110)$$

and $k = \omega/c_0$. Above a critical cut-off frequency, $f_c = c_0 m/2$, waves can propagate freely along the air column with a dispersive phase velocity of $c_0/\sqrt{1-(\omega_c/\omega)^2}$, while below the cut-off frequency the waves are exponentially damped. The cut-off frequency occurs when the free-space wavelength is approximately six times the length for the radius to increase by the exponential factor e.

Figure 15.75 compares the input resistance and reactance of an infinite exponential horn with that of a baffled piston having the same input area (*Kinsler* et al. [15.140] Fig. 14.19). The plots are for an exponential horn with $m = 3.7\,\mathrm{m}^{-1}$, which corresponds to a cut-off frequency of ≈100 Hz, and a baffled piston having the same radius of 2 cm as the throat of the exponential horn. Above ≈400 Hz, there is very little difference between the impedance of an infinitely long horn and a horn with a finite length of ≈1.5 m or longer, though below this frequency reflections cause additional fluctuations around the plotted values. Below the cut-off frequency, no acoustic energy can be radiated. Above the cut off the input resistance rises rather rapidly towards its limiting 100% radiating value. The exponential horn with a piston source at its throat is therefore a much more efficient radiator of sound than a baffled piston at all frequencies above the cut-off frequency.

The exponential horn illustrates how the flared cross section of brass instruments enhances the radiation of sound, though brass instruments are never based on exponential horns, otherwise no resonant modes could be set up. However, exponential horns were widely used in the early days of the gramophone. In the absence of any electronic amplification, they amplified the sound produced by the input diaphragm excited by the pick-up stylus on the recording cylinder or disc. They are still widely used in powerful public address systems. Such horns can also be used in reverse, as very efficient detectors of sound, with a microphone placed at the apex of the horn.

Bessel Horn

We now consider more realistic horns with a rapidly flaring end bell, which can often be modelled by what are known as Bessel horn shapes, with the radius varying

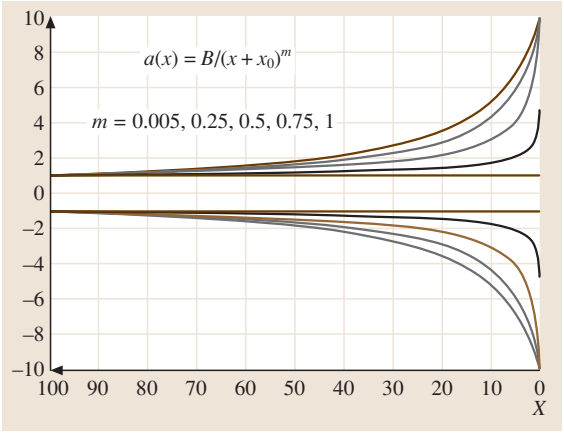


Fig. 15.76 Bessel horns representing the rapid outward flare of the bell on the end of a brass instrument, for a sequence of m values giving a ratio of input to output diameters of 10

Amazon
EX1013 Page 619

as $1/x^m$ from their open end. Typical flared horn shapes are shown in Fig. 15.76 for various values of m, where the horn functions $A/(x+x_0)^m$ have been normalised by suitable choice of A and x_0, to model horns with input and output values 1 and 10. Increasing the value of m increases the rapidity with which the flare opens out at the end.

Again assuming the plane-wave approximation, the horn equation can be written as

$$\frac{\partial^2\psi}{\partial x^2}+\left[\left(\frac{\omega}{c_0}\right)^2-\frac{m(m+1)}{x^2}\right]\psi=0\,, \qquad (15.111)$$

with solutions

$$\psi(kx)=x^{1/2}J_{m+1/2}(kx) \qquad (15.112)$$

and pressure varying from the end as

$$p(kx)=\frac{1}{x^{1/2}}J_{m+1/2}(kx)\,, \qquad (15.113)$$

where $J_{m+1/2}(kx)$ is a Bessel function of order $m+1/2$, giving the name to such horns.

In the plane-wave approximation, the sharpness and height of the barrier to wave propagation arising from the curvature could result in total reflection of the incident waves, so that no sound would be emitted from the end of the instrument. In reality, the curvature of the waveform will smear out any singularity in the horn function over a distance somewhat smaller than the output bell radius. Nevertheless, despite its limitations, the plane-wave model provides an instructive description of the influence of a rapidly flaring bell on a brass instrument. This is illustrated in Fig. 15.77 for the fundamental and fourth modes of a Bessel horn with $m=1/2$, with the pressure $p(x)$ varying from the output end as $xJ_1(kx)$.

The most important point to note is the way that the flare pushes the effective node of the incident sine-wave solutions (extended into the flared section as dashed curves) away from the end of the instrument. The effective length is therefore shortened and resonant frequencies increased, the effect being largest for the lower frequency modes. The flare and general outward curvature of the horn cross section therefore destroys the harmonicity of the modal frequencies. This is a completely general result for any horn with a flared end. In practice, the nodal positions will also be affected by the curvature of the wavefront, which will further perturb the modal frequencies, but without changing the above qualitative behaviour.

Benade ([15.133], Sect. 20.5) notes that, from the early 17th century, trumpets and trombones have been designed with strongly flaring bell corresponding to m values of 0.5–0.65, while French horns have bells with a less sudden flare with m values of 0.7–0.9.

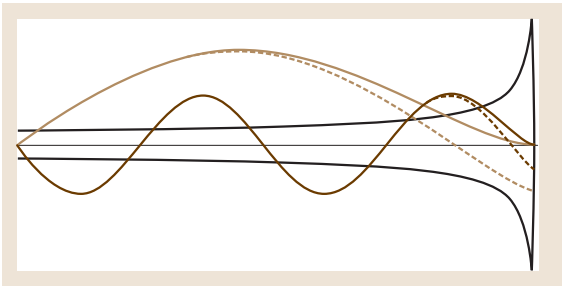


Fig. 15.77 Fundamental and fourth mode of an $m=1/2$ Bessel horn, illustrating the increase in wavelength and resulting shift inwards of the effective nodal position. The *dashed lines* illustrate the extrapolated sine-wave solutions from well inside the bore. The plot is of $p(x)/x$

From Fig. 15.77, it is easy to see how the player can significantly affect the pitch of a note on the French horn, by moving the supporting hand up into the bell of the instrument, which is referred to as hand-stopping. The pitch can be lowered by around a semitone, by placing the downwardly cupped hand and wrist against the top of the flared bell, effectively reducing the flare and increasing the effective length of the instrument. Alternatively, the pitch can be raised by a similar amount when the hand almost completely closes the inner bore. This leads to a major perturbation of the boundary conditions, effectively shortening the air column and reducing the output sound. The increase in frequency can be explained by the player using an almost unchanged embouchure to excite a higher frequency mode of the shortened tube. Because of the increased reflection of sound by the presence of the hand, the resultant sound although quieter is also much richer in higher partials (Fig. 15.113). Both effects are illustrated in Audio example CD-ROM.

Flared Horns

Use is made of the dependence of modal frequencies on the curvature of horn shapes to design brass instruments with a set of resonances, which closely approximate to a harmonic series with frequencies $f_n=nf_1$. This should be contrasted with the modes of a cylindrical tube closed at one end by the mouthpiece, which would only involve the odd-n integer modes with frequencies of $f_n=c/2L$.

Historically, this has been achieved by empirical methods, with makers adjusting the shapes of brass instrument to give as near perfect a set of harmonic

Amazon
EX1013 Page 620

resonances as possible, as illustrated in Fig. 15.78 from *Backus* ([15.132] Chap. 12). A harmonic set of modes enhances the playability of the instrument, as the partials of any continuously blown note then coincide with the natural resonances of the instrument. However, the fundamental mode is always significantly flatter than required and is therefore not normally used in musical performance. Nevertheless, the brass player can still sound a *pedal note* corresponding to the *virtual fundamental* of the higher harmonics by exciting a repetitive waveform involving the higher harmonics, but with only a weak Fourier component at the pitch of the sounded note.

The way that this is achieved is shown schematically in Fig. 15.79 starting from the odd-n resonances

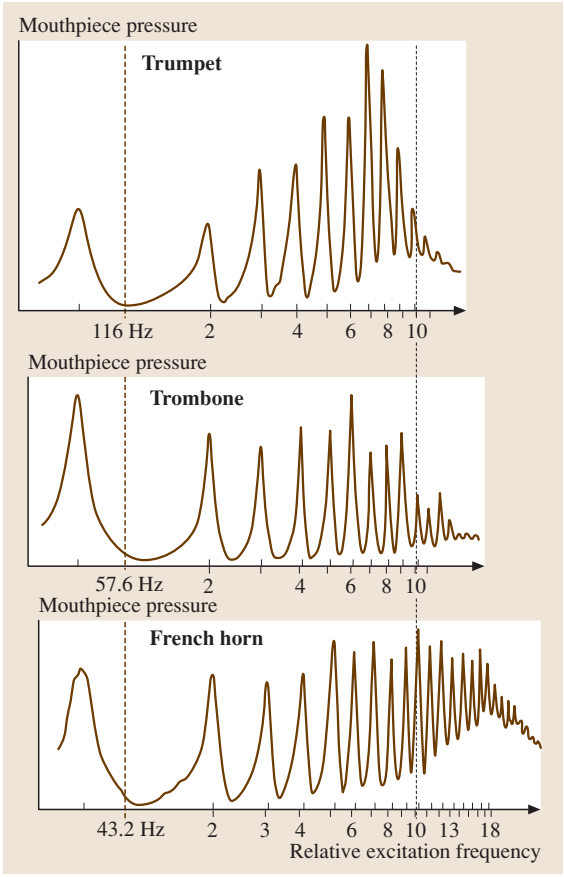


Fig. 15.78 Input impedances of the trumpet, trombone and French horn with the positions of the resonant peaks marked above the axis to show their position relative to a harmonic series based on the second harmonic of the instrument (after *Backus* [15.132])

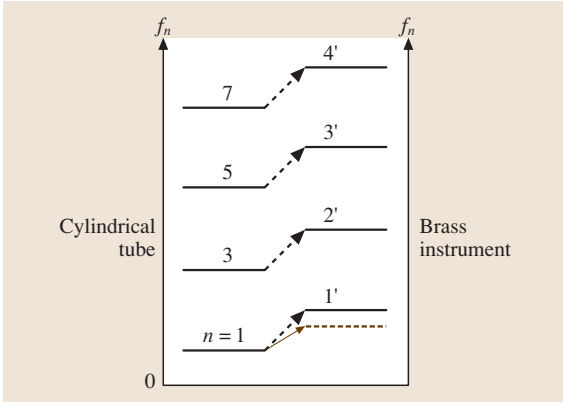


Fig. 15.79 The transformation of the odd-n modes of a cylindrical air column closed at one end to the near harmonic, all integer, n' modes of a flared brass instruments. The *lower dashed line* indicates schematically what can be achieved in practice for the lowest

of a cylindrical tube closed at one end by the mouthpiece to an appropriately flared horn of the same length. In practice, one can achieve a nearly perfect set of harmonic resonances, midway between the odd-integer modes of a cylindrical tube closed at one end, for all but the fundamental mode, which cannot be shifted upwards by a sufficient amount to form a harmonic fundamental, stopping and of the new set of modes.

Benade ([15.133], Sect. 20.5) has given an empirical expression for the frequencies of the partials of Bessel horns closed at the mouthpiece end, which closely describes these perturbations,

$$\frac{f_n'}{f_n} \sim \left(1 + 0.637 \frac{\sqrt{m(m+1)}}{2n-1}\right) , \qquad (15.114)$$

where the $(2n-1)$ in the denominator emphasising the preferential raising in frequency of the lower-frequency modes. This gives frequencies for the first six modes of a Bessel function horn with $m = 0.7$ are in the ratios 0.94, 2.00, 3.06, 4.12, 5.18 and 6.24, normalised to the $n = 2$ mode. These should be compared with the ideal 1, 2, 3, 4, 5, 6 ratios. Apart from the lowest note, which is a semitone flat, the higher modes are less than a semitone sharp compared with their ideal values.

Perturbation Models

Perturbation theory can be used to describe how changes in bore shape perturb the resonant modes of brass and woodwind instruments. *Fletcher* and *Rossing* ([15.5], Sect. 8.10) show that the change of frequency of a reso-

Amazon
EX1013 Page 621

nant mode $\Delta\omega$ resulting from small distributed changes $\Delta S(x)$ in bore area $S(x)$ is given by

$$\frac{\Delta\omega_n}{\omega_n} = -\frac{1}{2}\left(\frac{c_0}{\omega_n}\right) \int_0^L \left[\frac{\partial}{\partial x}\left(\frac{\Delta S(x)}{S(x)}\right) p_n \frac{\partial p_n}{\partial x}\,\mathrm{d}x\right] \Big/ \int_0^L \left[S(x) p_n^2\,\mathrm{d}x\right] . \quad (15.115)$$

An alternative equivalent derivation uses Rayleigh's harmonic balance argument and equates the peak kinetic energy to the peak potential energy. To first order, the perturbation is assumed to leave the shape of the modal wavefunction unchanged. The kinetic and potential energy stored in a particular resonant mode can be expressed in terms of the local kinetic $\frac{1}{2}\rho\omega_n^2\xi_n^2$ and strain $\frac{1}{2}\gamma P_0(\partial\xi_n/\partial x)^2$ energy densities. For simplicity, we consider the perturbation of the nth resonant mode of a cylindrical air column open at one end, with particle displacement $\xi_n \approx \sin(n\pi x/L)\cos(\omega_n t)$, where n is an odd integer. Equating the peak kinetic and potential energy over the perturbed bore of the cylinder, we can then write

$$\omega_n'^2 \int_0^L \rho\,[S+\Delta S(x)]\sin^2(kx)\,\mathrm{d}x = \gamma P_0 k_n^2 \int_0^L [S+\Delta S(x)]\cos^2(kx)\,\mathrm{d}x , \quad (15.116)$$

where ω_n' is the perturbed frequency. This can be rewritten as

$$\frac{\omega_n'^2}{\omega_n^2} = \int_0^L [S+\Delta S(x)]\cos^2(kx)\,\mathrm{d}x \Big/ \int_0^L [S+\Delta S(x)]\sin^2(kx)\,\mathrm{d}x . \quad (15.117)$$

Because the perturbations are assumed to be small, we can rearrange (15.117) to give the fractional change in frequency

$$\frac{\Delta\omega_n}{\omega_n} = \frac{1}{L}\int_0^L \frac{\Delta S(x)}{S}\left(\cos^2 kx - \sin^2 kx\right)\mathrm{d}x . \quad (15.118)$$

Hence, if the tube is increased in area close to a displacement antinode, where the particle flow is large (low pressure), the modal frequency will increase, whereas the frequency will decrease, if constricted close to a nodal position (large pressure) (*Benade* [15.133], Sect. 22.3). This result can be generalised to a tube of any shape. Hence, by changing the radius over an extended region close to a node or antinode, the frequencies of a particular mode can be either raised or lowered, but at the expense of similar perturbations to other modes. Considerable art and experience is therefore needed to correct for the inharmonicity of several modes simultaneously.

Electric Circuit Analogues

It is often instructive to consider acoustical systems in terms of equivalent electric circuit analogues, where voltage V and electrical current I can represent the acoustic pressure p and flow along a pipe U. For example, a volume of air with flow velocity U in a pipe of area S and length l has a pressure drop $(\rho l/S)\partial U/\partial t$ across its length, which is equivalent to the voltage $L\partial I/\partial t$ across an inductor in an electrical circuit. Likewise, the rate of pressure rise, $\partial p/\partial t = \gamma P_0 U/V$, as gas flows into a volume V, is equivalent to the rate of voltage rise, $\partial V/\partial t = I/C$ across a capacitance $C \equiv V/\gamma P_0$.

As a simple example, we re-derive the Helmholtz resonance frequency, previously considered in relation to the principal air resonance of the air inside a violin or guitar body (Sect. 15.2.4), but equally important, as we will show later, in describing the resonance of air within the mouthpiece of brass instruments.

In its simplest form, the Helmholtz resonator consists of a closed volume V with an attached cylindrical pipe of length l and area S attached, through which the air vibrates in and out of the volume. All dimensions are assumed small compared to the acoustic wavelength, so that the pressure p in the volume and the flow in the pipe U can be assumed to be spatially uniform. The volume acts as an acoustic capacitance $C = V/\gamma P_0$, which resonates with the acoustic inductance $L = \rho l/S$ of the air in the neck. The resonant frequency is therefore given by

$$\omega_{\mathrm{Helmholtz}} = \frac{1}{\sqrt{LC}} = \sqrt{\frac{S}{\rho l}\frac{\gamma P_0}{V}} = c_0\sqrt{\frac{S}{lV}} , \quad (15.119)$$

as derived earlier.

Any enclosed air volume with holes in its containing walls acts as a Helmholtz resonator, with an

Amazon
EX1013 Page 622

effective *kinetic inductance* of the hole region equivalent to a tube of the same diameter with an effective length of wall thickness plus ≈ 0.61 hole radius (*Kinsler* et al. [15.140] Sect. 9.2). This is the familiar end-correction for an open-ended pipe (*Kinsler* et al. [15.140] Sect. 9.2). Open holes of different diameters will therefore give resonances corresponding to different musical tones. The ocarina is a very simple musical instrument based on such resonances, in which typically four or five holes with different areas can be opened and closed in combination, to give a full range of notes on a chosen musical scale (audio CD-ROM). Because the sound is based on the single resonance of a Helmholtz resonator, there are no simply related higher-frequency modes that can be excited. Ocarinas appear in many ancient and ethnic cultures around the world and are often sold as ceramic toys.

Acoustic Transmission Line

There is also a close equivalence between acoustic waves in wind instruments and electrical waves on transmission lines, with an acoustic pipe having an equivalent inductance $L_0 = \rho/S$ and capacitance $C_0 = S/\gamma P_0$ per unit length. For a transmission line the wave velocity is therefore $c_0 = \sqrt{1/L_0C_0} = \sqrt{\gamma P_0/\rho}$ and characteristic impedance $Z_0 = \sqrt{L_0/C_0} = \rho c_0/S$, as expected. The input impedance of a transmission line as a function of its characteristic impedance and terminating load is given by (15.100).

Valves and Bends

To enable brass instruments to play all the notes of the chromatic scale, short lengths of coiled-up tubing are connected in series with the main bore by a series of piston- or lever-operated air valves. The constriction of air flow through the air channels within the valve structures and the bends in the tubing, used to reduce the size of the instruments to a convenient size for the player to support, will clearly present discontinuities in the acoustic impedance of the air bore and will lead to reflections. Such reflections will influence the feel of the instrument for the player exciting the instrument via the mouthpiece and will also perturb the frequencies of the resonant modes of the instrument.

If the discontinuities are short in size relative to the acoustic wavelengths involved, the discontinuity can be considered as a discrete (localised) lumped circuit element. Using our electromechanical equivalent, a short, constricted channel through a valve can be represented as an inductance $\rho L_{\text{valve}}/S_{\text{valve}}$ in series with the acoustic transmission line, or an equivalent additional extra length of bore tubing $L_{\text{valve}}S_{\text{tube}}/S_{\text{valve}}$ of cross section S_{tube}. For all frequencies such that $kL_{\text{valve}} \ll 1$, the valve simply increases the length of the acoustic air column slightly and the frequencies of all the lower modes by the same fractional amount. Only at very high frequencies, outside the normal playing range, will the constricted air channel significantly change the modal frequencies.

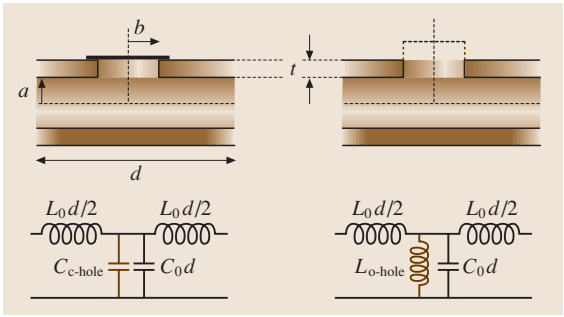


Fig. 15.80 Equivalent circuits for a short length d of cylindrical pipe containing a closed and an open tone hole, shunting the acoustic transmission line with a capacitance and inductance, respectively

When a straight length of cylindrical tube of radius a is connected to the same size tubing but bent into a circle of radius R, there will a small change in the acoustic impedance and velocity of sound waves, which arises because the forces acting on each element induces rotational in additional to linear motion. The presence of bends will lead to reflections and slight perturbations of resonant frequencies, though these effects will again be relatively small. *Nederveen* [15.142] showed that fractional increase in phase velocity and decrease in wave impedance of a rectangular duct is given by the factor $F^{1/2}$, where

$$F = \frac{B^2}{2} \Big/ \left[1 - \left(1 - B^2\right)^{1/2}\right] = 1 - B^2/4 \quad \text{for} \quad B \ll 1 \,, \tag{15.120}$$

$B = a/R$, a is the half-width of the duct and R its radius. *Keefe* and *Benade* [15.143] subsequently generalised this result to a bent circular tube, with its radius r replacing a.

Finger Holes

In many woodwind instruments, tone holes can be opened or closed to change the effective resonating length of an air column and hence pitch of the sounded note. The holes can be closed by the pad of the finger or a hinged felt-covered pad operated with levers.

Amazon
EX1013 Page 623

To a first approximation, opening a side hole creates a pressure node at that position, shortening the effective length of the instrument and raising the modal frequencies. However, as described in *Fletcher* and *Rossing* ([15.5], Sect. 15.2) and in detail by *Benade* ([15.133], Chaps. 21 and 22), the influence of the side holes is in practice strongly dependent on the hole size, position and frequency, as summarised below.

At low frequencies, when the acoustic wavelength is considerably longer than the size and spacing of the tone holes, one can account for the effect of the tone holes by considering their equivalent capacitance when closed and their inductance when open, as illustrated schematically in Fig. 15.80.

Because the walls of wind instruments and particularly woodwind instruments have a significant thickness, the tone holes when shut introduce additional small volumes distributed along the length of the vibrating air column. Each closed hole will introduce an additional volume and equivalent capacitance $C_{\text{c-hole}} = \pi b^2/\gamma P_0$, which will perturb the frequencies of the individual partials upwards or downwards by a small amount that will depend on its position relative to the pressure and displacement nodes and the closed volume of the hole. In severe cases, the perturbations can be as large as a few per cent (one semitone is 6%), which requires compensating changes in bore diameter along the length of the instrument, to retain the harmonicity of the partials. However, this is essentially a problem that depends on geometrical factors involving the air column alone. Once solved, like all acoustic problems involving the shape and detailed design, instruments can be mass-produced with almost identical acoustic properties, quite unlike the problems that arise for stringed instruments.

The more interesting situation is when the holes are opened, introducing a pressure node at the exit of the tone hole and shortening the effective acoustical length of the instrument. An open hole can be considered as an inductance, $L \sim \rho(t+0.6b)/\pi b^2$, where the effective length of the hole is increased by the unflanged hole end-correction. Neglecting radiation losses from the hole [*Fletcher* and *Rossing* [15.5], (15.21, 22)], the effective impedance Z^* of an open-ended cylindrical pipe of length l and radius a shunted by the inductive impedance of a circular hole of radius b set into the wall of thickness t is given by

$$\frac{1}{Z^*} \sim \frac{\pi b^2}{\mathrm{i}\omega\rho(t+0.6b)} + \frac{\pi a^2}{\mathrm{i}\rho c_0 \tan kl} = \frac{\pi a^2}{\mathrm{i}\rho c_0 \tan kl'} \,. \tag{15.121}$$

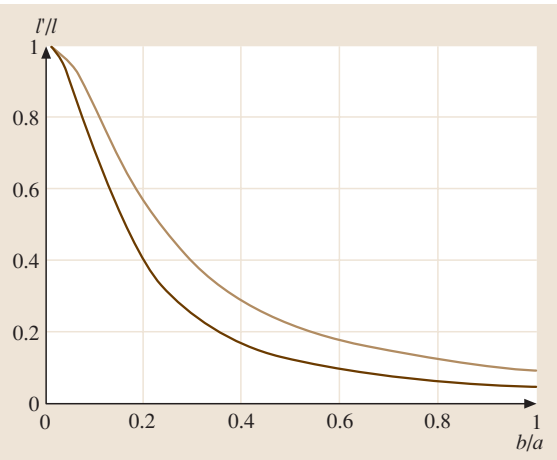


Fig. 15.81 Low-frequency ($kl \ll 1$) fractional reduction of effective length of a cylindrical end-pipe as a function of hole to cylinder radius, for additional lengths of 10 (lower curve) and 20 (upper curve) times the tube radius in length. The side wall thickness is 0.4 times the tube radius

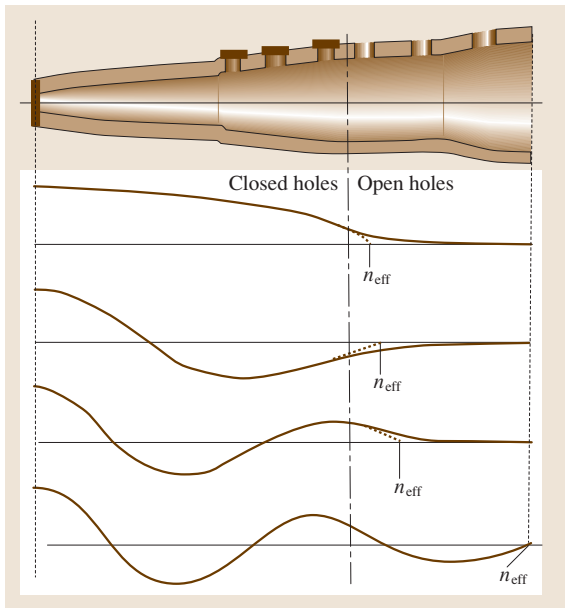


Fig. 15.82 Schematic representation of the influence of open holes on the first four partials of a woodwind instrument, with the effective length indicated by the intercept n_{eff} on the axis of the extrapolated incident wave (after *Benade* [15.133])

Amazon
EX1013 Page 624

Thus can be expressed in terms of an impedance of an effectively reduced length l'.

$$\text{For } kl \ll 1\,, \quad \frac{l'}{l} = \left[1 + \frac{t + 0.6b}{l}\left(\frac{a}{b}\right)^2\right]^{-1}\,. \tag{15.122}$$

The change in effective length introduced by the open hole depends strongly on its area relative to that of the cylinder, the thickness of the wall and its length from the end. This gives the instrument designer a large amount of flexibility in the positioning of individual holes on an instrument. Figure 15.81 illustrates the dependence of the effective pipe length on the ratio of hole to cylinder radii for two lengths of pipe between the hole and end of the instrument.

Not surprisingly, a very small hole with $b/a \ll 1$ has a relatively small effect on the effective length of an instrument. In contrast, a hole with the same diameter as that of the cylinder shortens the effective added length to about one hole diameter.

In practice, there will often be several holes open beyond the first open tone hole, all of which can affect the pitch of the higher partials.

Consider a regular array of open tone holes spaced a distance d apart. The shunting kinetic inductance of each open hole is in parallel with the capacitance associated with the volume of pipe between the holes. At low frequencies, such that $\omega \ll 1/\sqrt{L_{\text{hole}} C_0 d}$, the impedance is dominated by the hole inductance, so that each hole attenuates any incident wave by approximately the ratio

$$\sim L_{\text{hole}}/\left(L_{\text{hole}} + L_0 d\right) = \left[1 + \frac{d}{t + 1.5b}\left(\frac{a}{b}\right)^2\right]^{-1}\,, \tag{15.123}$$

where L_0 is the inductance of the pipe per unit length. Incident waves are therefore attenuated with an effective node just beyond the actual hole as discussed above.

However, for frequencies such that $\omega \gg 1/\sqrt{L_{\text{hole}} C_0 d}$, the impedance of the shunting hole inductance is much larger than that of the capacitance of the air column, so that the propagating properties of the incident waves is little affected be the presence of the open hole. There is therefore a *crossover* or *cut-off* frequency

$$\omega \approx 1/\sqrt{L_{\text{hole}} C_0 d} = c_0 \frac{a}{b}\left(\frac{1}{t_{\text{eff}} d}\right)^{1/2}\,, \tag{15.124}$$

below which the incident waves are reflected to give a pressure node just beyond the first hole of the array and above which waves propagate increasingly freely through the array to the open end of the instrument.

Figure 15.82 (*Benade* [15.133], Fig. 21.1) illustrates the effect of an array of open holes on the first few partials of a typical woodwind instrument, highlighting the increase in acoustic length of the instrument (indicated by the intercept of the extrapolated incident waveform)

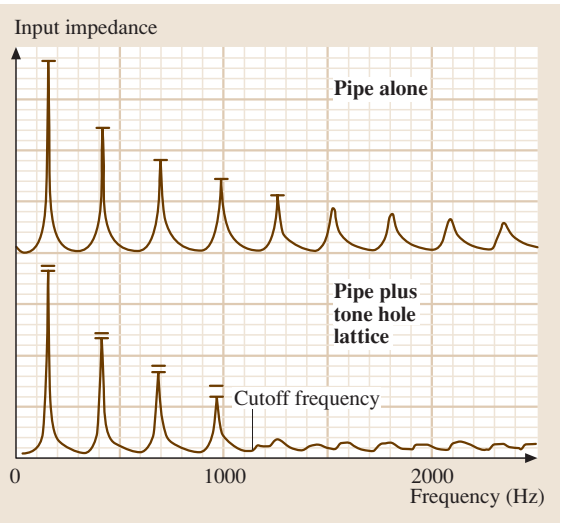


Fig. 15.83 Illustration of the cut-off-frequency effect, when adding an addition length of tubing with an array of open tone holes (after *Benade* [15.133])

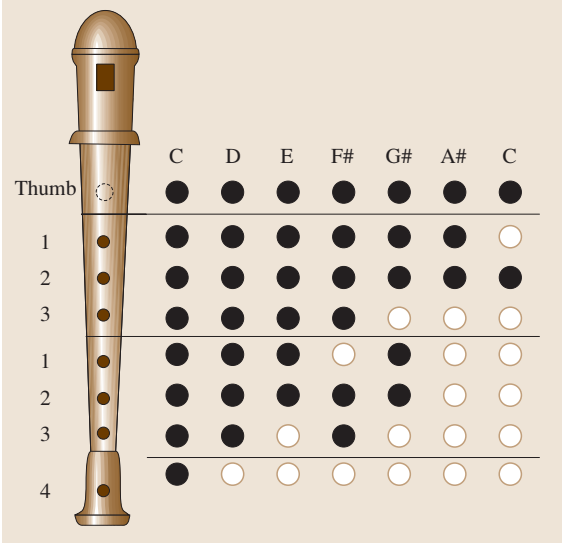


Fig. 15.84 Soprano recorder fingering for the first seven notes of a whole-tone scale (after *Fletcher* and *Rossing* [15.5])

Amazon
EX1013 Page 625

with increasing frequency. The dependence of the effective length of the acoustic air column on frequency is therefore rather similar to the influence of the flare on the partials of a brass instrument.

A consequence of the greater penetration at high frequencies of the acoustic wave through the array of open tone holes is the greater attenuation of such waves by radiation and the consequent reduction in the amplitude of the higher resonant modes in measurements of the input impedance. This is illustrated in Fig. 15.83 for a length of clarinet tubing first without and then with an added section containing an array of equally spaced tone holes (*Benade* [15.133], Fig. 21.3).

Benade ([15.133], Sect. 21.1) states that "specifying the cut-off frequency for a woodwind instrument is tantamount to describing almost the whole of its musical personality" – assuming the proper tuning and correct alignment of resonances for good oscillation. His measured values of the cut-off frequency for the upper partials of classical and baroque instruments are 1200–2400 Hz for oboes, 400–500 Hz for bassoons, and 1500–1800 Hz for clarinets.

Cross-Fingering

The notes of an ascending scale can be played by successively opening tone holes starting from the far end of the instrument. In addition, by overblowing, the player can excite notes in the *second register* based on the second mode. As remarked earlier, instruments like the flute and oboe overblow at the octave, whereas the clarinet overblows at the twelfth (an octave plus a perfect fifth). To sound all the semitones of the western classical scale on the flute or oboe would therefore require 12 tone holes and the clarinet 20 – rather more than the fingers on the two hands! In practice, the player generally uses only three fingers on the left hand and four on the right to open and close the finger holes. The thumb on the left hand is frequently used to open a small *register* hole near the mouthpiece, which aids the excitation of the overblown notes in the higher register.

In practice, *cross-* or *fork-fingering* enables all the notes of the chromatic scale to be played using the seven available fingers and combinations of open and closed tone holes. This is illustrated in Fig. 15.84 for the baroque recorder (*Fletcher* and *Rossing* [15.5], Fig. 16.21). The bottom two notes can be sharpened by a semitone by half-covering the lower two holes and the overblown notes an octave above are played with the thumb hole either fully open or half closed. Cross-fingering makes use of the fact that the standing waves

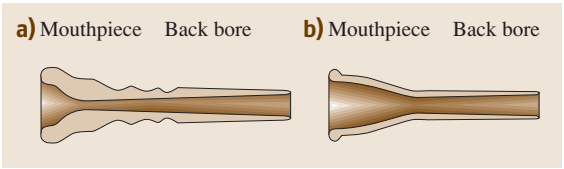


Fig. 15.85a,b Cross sections of **(a)** trumpet mouthpiece and **(b)** horn mouthpiece (after *Backus* [15.132])

set up in a pipe extend an appreciable distance into an array of open tone holes (Fig. 15.82), so that opening and closing holes beyond the first open hole can have an appreciable influence on the effective length of the resonating air column.

Modern woodwind instruments use a series of interconnected levers operated by individual keys, which facilitates the ease with which the various hole-opening combinations can be made.

Radiated Sound

Although the reactive loading of an open hole determines the effective length of the resonant air column, particularly at low frequencies, it does not follow that all the sound is radiated from the open tone holes. Indeed, since the intensity of the radiated sound depends on $(ka)^2$, very little sound will be radiated by a small hole relative to the much wider opening at the end of an instrument. The loss in intensity of sound passing an open side hole may therefore, in large part, be compensated by the much larger radiating area at end of the instrument. This also explains why the characteristic hollow sound quality of a cor anglais, derived in part from the egg-shaped resonating cavity near its end, is retained, even when the tone holes are opened on the mouthpiece side of the cavity.

In practice, the sound from the end and open tone holes of a woodwind instrument act as independent monopole sources. When the acoustic wavelength becomes comparable with the hole spacing, interesting interference effects in the output sound can occur contributing to strongly directional radiation patterns, as discussed by *Benade* ([15.133], Sect. 21.4). Similarly, reciprocity allows one to make use of such interference effects to produce a highly directional microphone by placing a microphone at the end of a cylindrical tube with an array of open side holes.

Brass Mouthpiece

Brass instruments are played using a mouthpiece insert, against which the lips are pressed and forced to vibrate by the passage of air between them. The mouthpiece not

Amazon
EX1013 Page 626

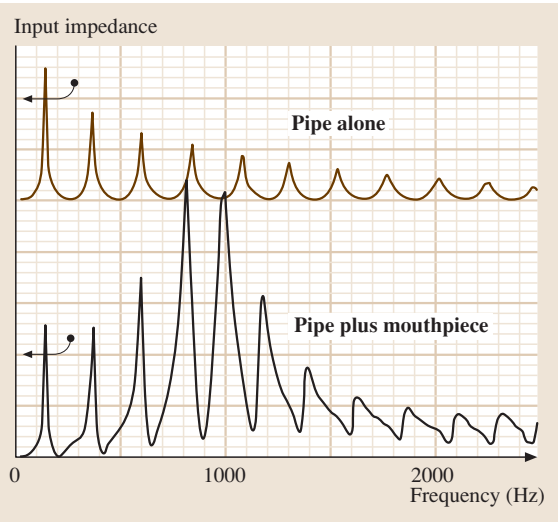


Fig. 15.86 Input impedance of a length of cylindrical trumpet pipe with and without a mouthpiece attached (after *Benade* [15.133])

only enables the player to vibrate their lips over a wide range of frequency, but also provides a very important acoustic function in significantly boosting the amplitude

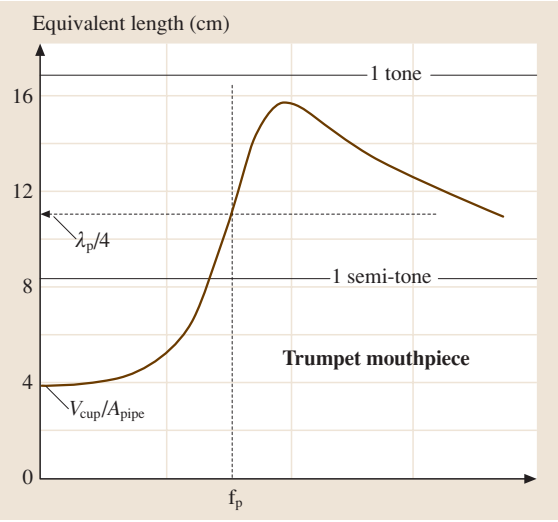


Fig. 15.87 The amount by which a trumpet tube of length 137 cm would have to be lengthened to compensate for the lowering in frequency of the instrument's resonant frequencies when a mouthpiece is attached to the input (after *Benade* [15.133]). The changes in length to give a semitone and a whole-tone change in frequency are indicated by the *horizontal lines*

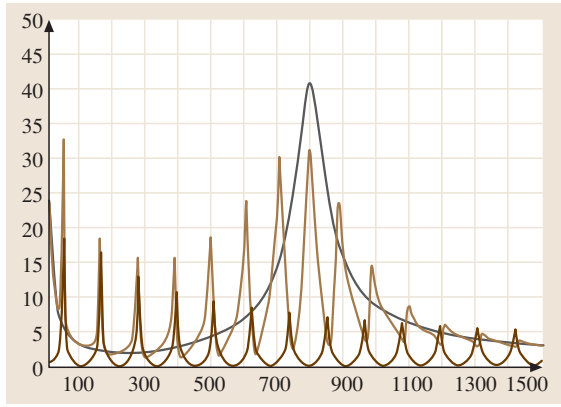


Fig. 15.88 The calculated impedance of an 800 Hz Helmholtz mouthpiece (*dark brown*), an attached pipe (*black*) and the combination of mouthpiece and pipe (*light brown*)

of the higher partials, helping to give brass instruments their bright and powerful sounds.

Typical mouthpiece shapes are shown in Fig. 15.85. Mouthpieces can be characterized by the mouthpiece volume and the *popping* frequency characterizing the Helmholtz resonator comprising the mouthpiece volume and backbore. The popping frequency can easily be estimated from the sound produced when the mouthpiece is slapped against the open palm of the hand (audio CD-ROM).

By adjusting the tension in the lips, the shape of the lips within the mouthpiece (the embouchure), and the flow of air between the lips via the pressure in the mouth, the skilled brass player forces the lip to vibrate at the required frequency of the note to be played. This can easily be demonstrated by making a pitched buzzing sound with the lips compressed against the rim of the mouthpiece cup. The circular rim constrains the lateral motion of the lips making it far easier to produce stable high notes. A brass player can sound all the notes on an instrument by simply blowing into the mouthpiece alone, but the mouthpiece alone produces relatively little volume. The instrument both stabilises the playing frequencies and increases the coupling between the vibrating lips and radiated sound.

Figure 15.86 illustrates the enhancement in the input impedance around the popping frequency, when a mouthpiece is attached to the input of a cylindrical pipe, as measured by *Benade* [15.133]. Benade showed that the influence of the mouthpiece on the acoustical characteristics of a brass instrument is, to a first approximation, independent of the internal bore shape and

Amazon
EX1013 Page 627

can be characterized by just two parameters, the internal volume of the mouthpiece and the popping frequency.

Benade also measured the perturbation of the resonant frequencies of an instrument by the addition of a mouthpiece, as illustrated in Fig. 15.87. At low frequencies, the mouthpiece simply extends the effective input end of a terminated tube by an equivalent length of tubing having the same internal volume as the mouthpiece. In the measurements shown, Benade removed lengths of the attached tube to keep the resonant frequencies unchanged on adding the mouthpiece. However, since the fractional changes in frequency are small, the measurements are almost identical to the effective increase in length from the addition of the mouthpiece.

At the mouthpiece *popping* frequency (typically in the range 500 Hz to 1 kHz depending on the mouthpiece and instrument considered), the effective increase in length is $\lambda/4$. This can result in decreases in resonant frequencies by as much as a tone, which could have a significant influence on the harmonicity, and hence the playability, of an instrument. The effective length continues to increase above the popping frequency before decreasing at higher frequencies. In many brass instruments, such as the trumpet, there is also a longer transitional conical section (*the lead pipe*) between the narrow bore of the mouthpiece and the larger-diameter main tubing. This reduces the influence of the mouthpiece on the tuning of individual resonances and the overall formant structure of resonances.

It is straightforward to write down the input impedance inside the cup of a mouthpiece attached to an instrument using an equivalent electrical circuit. The volume within the cup is represented by a capacitance C in parallel with the inductance L and resistance R of air flowing through the backbore, which is in series with the input impedance of the instrument itself, so that

$$Z_{\text{in}} = \frac{1}{\mathrm{i}\omega C}\,\frac{\mathrm{i}\omega L + R + Z_{\text{horn}}}{(1/\mathrm{i}\omega C) + \mathrm{i}\omega L + R + Z_{\text{horn}}}\,. \qquad (15.125)$$

Figure 15.88 shows the calculated input impedance of an 800 Hz Helmholtz mouthpiece resonator, of volume 5 cm^3 with a narrow-backbore neck section resulting in a Q-value of 10, before and after attachment to a cylindrical pipe of length 1.5 m and radius 1 cm, radiating into free space at its open end. The input impedance of the pipe alone is also shown. Note the marked increase in heights and strong frequency shifts of the partials in the neighbourhood of the mouthpiece resonance. As anticipated from our previous treatment of coupled resonators in the section on stringed instruments, the addition of the mouthpiece introduces an additional normal mode resonance in the vicinity of the Helmholtz resonance. In addition, it lowers the frequency of all the resonant modes below the popping frequency and increases the frequency of all the modes above.

Above the mouthpiece resonance, the input impedance is dominated by the inertial input impedance of the mouthpiece. The resonances of the air column are superimposed on this response and exhibit the familiar dispersive features already noted for narrow violin string resonances superimposed on the much broader body resonances. The calculated behaviour is very similar to the measured input admittance of typical brass instrument (Fig. 15.78) as extended to instruments with realistic bore shapes by *Caussé*, *Kergomard* and *Lurton* [15.144].

15.3.3 Reed Excitation

In the next sections, we consider the excitation of sound by: (a) the single and double reeds used for many woodwind instruments and selected organ pipes, (b) the vibrating lips in the mouthpiece of brass instrument, and (c) air jets used for the flute, certain organ stops and many ethnic instruments such as pan pipes.

Reed Types

Figure 15.89 shows a number of reed types used in woodwind and brass instruments (*Fletcher* and *Rossing* [15.5], Figs. 13.1, 7).

Helmholtz [15.127] classified two main types of reed: *inward-striking reeds*, which are forced shut by an overpressure within the mouth, and *outward-striking reeds*, forced open by an overpressure. Modern authors often prefer to call such reeds inward-closing and

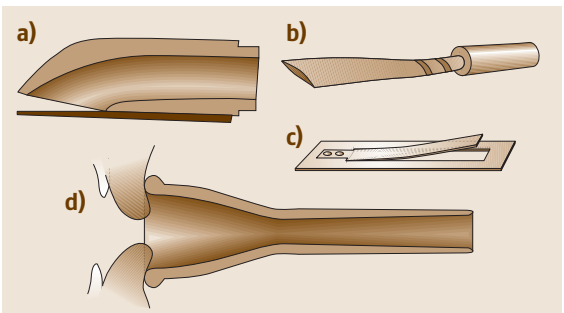


Fig. 15.89a–d Examples of wind and brass instrument reeds: **(a)** a single reed (clarinet), **(b)** a double reed (oboe), **(c)** a cantilever reed (harmonium) and **(d)** the mouthpiece lip-reed (horn) (after *Fletcher* and *Rossing* [15.5])

Amazon
EX1013 Page 628

outward-opening or swinging-door reeds. In addition there are reeds that are pulled shut by the decreased Bernoulli pressure created by the flow of air between them. Such reeds are often referred to as sideways-striking or sliding-door reeds.

A more formal classification (*Fletcher* and *Rossing* [15.5], Sect. 13.3) characterises such reeds by a doublet symbol (σ_1, σ_2), where the values of $\sigma_{1,2} = \pm 1$ describe the action of over- and under-pressures at the input and output ends of the reed. When the valve is forced open by an overpressure at either end, $\sigma_{1,2} = +1$; if forced open by an under-pressure, $\sigma_{1,2} = -1$. The force tending to open the valve can then be written as $(\sigma_1 p_1 S_1 + \sigma_2 p_2 S_2)$, where $S_{1,2}$ and $p_{1,2}$ are the areas and pressures at the reed input and output. The operation of reeds can therefore be classified as $(+, -)$, $(-, +)$,$(-, -)$ or $(+, +)$. Single and double woodwind reeds are inward-striking $(-, +)$ valves, while the vibrating lips in a mouthpiece and the vocal cords involve both outward-swinging $(+, -)$ and sideways-striking $(+, +)$ actions.

Figure 15.90 summarises the steady-state and dynamic flow characteristics of the above reeds for typical operating pressures across the valve, $\Delta p = p_m - p_{ins}$,

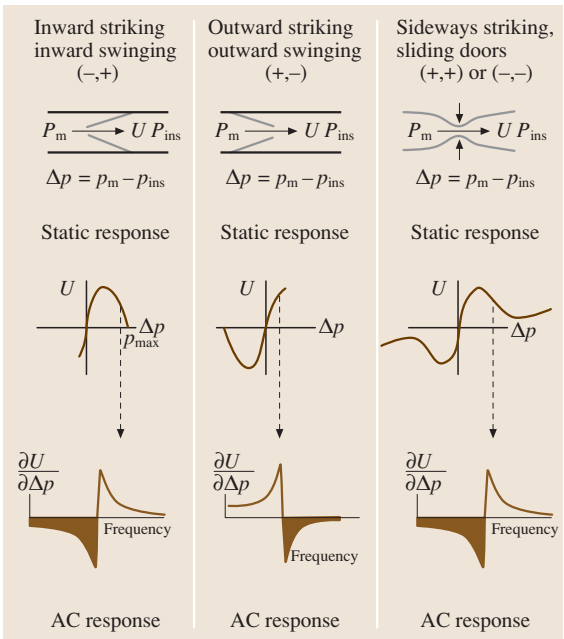


Fig. 15.90 Main classifications of vibrating reeds summarising reed operation, nomenclature and the associated static and *ac* conductance, with the negative resistance frequency regimes indicated by *solid shading*

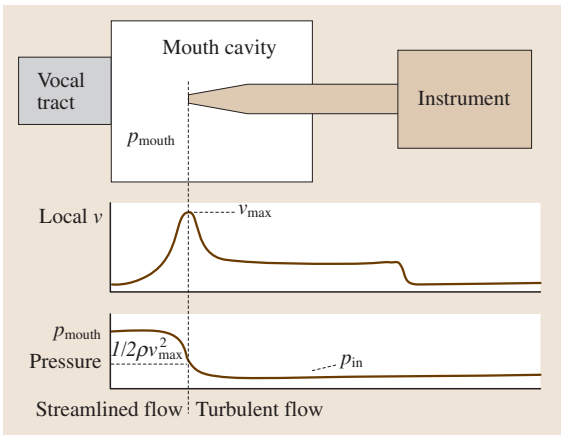


Fig. 15.91 Schematic representation of vocal tract, mouth cavity, reed and instrument, illustrating the variation of local velocity and pressure for air flowing into and along the reed and attached instrument

where p_m and p_{ins} are the input and output pressures in the mouth and instrument input, respectively. For the inward-swinging $(-, +)$ reed, the flow rate initially increases for a small pressure difference across the valve, but then decreases as the difference in pressures tends to close the valve, leading to complete closure above a certain pressure difference p_{max}. Before closure, there is an extended range of pressures where the flow rate decreases for increasing pressure difference across the reed. This is equivalent to an input with a negative resistance to flow. This results in positive feedback exciting resonances of any attached air column, provided the feedback is sufficient to overcome viscous, thermal and radiation losses.

It is less obvious why the outward-swinging $(-, +)$ reed can give positive feedback, because the steady-state flow velocity always increases with increasing pressure across the valve. However, this is only true at low frequencies below the mechanical resonance of the reed. Above its resonant frequency, the reed will move in anti-phase with any sinusoidally varying fluctuations in pressure. This will result in a regime of negative resistance and the resonant excitation of any attached air column, as discussed by *Fletcher* et al. [15.145].

Sideways-striking $(+, +)$ or $(-, -)$ reeds behave rather like inward-striking reeds, with an extended region of negative conductance. However, such reeds will never cut off the flow completely, so that for large pressure differences the dynamic conductance again becomes positive, as indicated in Fig. 15.90.

Amazon
EX1013 Page 629

Bernoulli Pressures

Figure 15.91 schematically illustrates the variation of flow velocity and pressure as air flows from the mouth into the reed and attached air column. To solve the detailed dynamic response from first principles for a specific reed geometry would require massive computer modelling facilities. Fortunately, the physics involved is reasonably well understood, so that relatively simple models can be used to reproduce reed characteristics rather well, as illustrated for the clarinet reed in the next section.

The operation of all reed generators is controlled by the spatial variations in Bernoulli pressure exerted by the air flowing across the reed surfaces. Such variations in P arise because, within any region of streamlined flow with velocity v, $P + \frac{1}{2}\rho v^2$ remains constant. Hence the pressure will be lowered on any surface over which the air is flowing. The flow of air is determined by the specific reed assembly geometry and the nonlinear Navier–Stokes equation, which also includes the effects of viscous damping.

After passing through the narrow reed constriction, the air emerges as a jet, which breaks down into turbulent motion on the downstream side of the reed. The turbulence leads to a rapid lateral mixing of the air, so that the flow is no longer streamlined. As a result, the pressure on the downstream end of the reed opening remains low and fails to recover to the initial pressure inside the mouth. The double reeds used for playing the oboe, bassoon and bagpipe chanter are mounted on a relatively long, narrow tube connected to the wider bore of the instrument. Turbulent flow in this region could contribute significantly to the flow characteristics, though recent measurements by *Almeida* et al. [15.146] have shown that such effects are less important than initially envisaged, as discussed later.

Single Reed

We first consider the clarinet reed, which is one of the simplest and most extensively studied of all woodwind reeds (*Benade* [15.133], Sect. 21.2) *Fletcher* and *Rossing* ([15.5], Chap. 13 and recent investigations by *Dalmont* and collaborators [15.147, 148]). Figure 15.92 shows a cross section of a clarinet mouthpiece, defining the physical parameters of a highly simplified but surprisingly realistic model.

The lungs are assumed to supply a steady flow of air U, which maintains a steady pressure P_{mouth} within the mouth. Air flows towards the narrow entrance or lip of the reed through which it passes with velocity

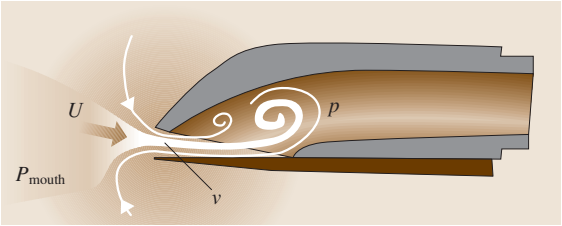


Fig. 15.92 Cross section of air flow through a clarinet mouthpiece and reed assembly, illustrating the streamlined flow into the gap with jet formation and turbulence on exiting the reed entrance

v. Because the air flowing into the reed is streamlined, the pressure drops by $\frac{1}{2}\rho v^2$ on entering the reed, while the much slower-moving air on the outer surfaces of the reed leaves the pressure on the outer reed surfaces largely unchanged. The air is then assumed to stream through the narrow gap of the reed to form an outward-going jet, which breaks up into vortices and turbulent flow on the far side of the input constriction, with no further change in overall pressure p in the relatively wide channel on the downstream side of the reed entrance.

The resulting pressure difference $\frac{1}{2}\rho v^2$ across the reed forces the reed back towards its closing position on the curved *lay* of the mouthpiece, indicated by the dashed line in Fig. 15.92. The pressure difference Δp is assumed to reduce the area of reed opening from S_0 to $S_0(1 - \Delta p/p_{\max})$, where $p_{\max}$ is the pressure difference required to close the reed. The net flow of air through

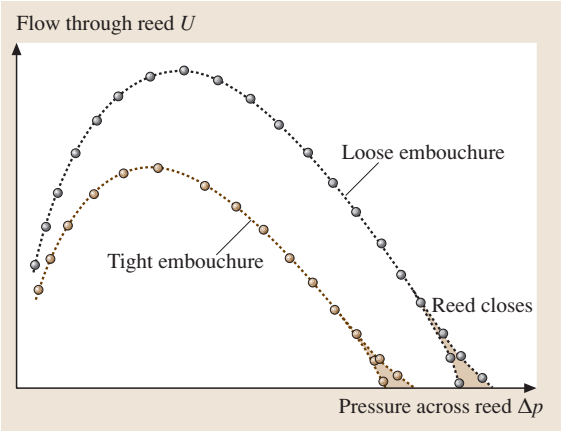


Fig. 15.93 Quasistatic flow through a clarinet single reed as a function of pressure across it illustrating the influence of the player's embouchure on the shape (after *Benade* [15.133])

Amazon
EX1013 Page 630

the reed is therefore given by

$$U(\Delta p) = \alpha(\Delta p)^{1/2}(1 - \Delta p/p_{\max}) \,. \qquad (15.126)$$

The player can control these characteristic by varying the position and pressure of the lips on the reed, which is referred to as the *embouchure*. A lower pressure is required to close the reed, if the reed is already partially closed by pressing the lips against the reed to constrict the entrance gap.

The flow rate U as a function of static pressure across a clarinet reed is illustrated by the measurements of *Backus* and *Nederween* [15.149] redrawn in Fig. 15.93. Apart from a small region near closure, where the exact details of the closing geometry and viscous losses may also be important, the shape of these curves and later measurements by *Dalmont* et al. [15.148, 150], which exhibit a small amount of hysteresis from viscoelastic effects on increasing and decreasing pressure, are in excellent agreement with the above model. The measurements also illustrate how the player is able to control the flow characteristics by changing the pressure of the lips on the reed.

The *reed equation* can be written in the universal form

$$\frac{U\left(\frac{\Delta p}{p_{\max}}\right)}{U_{\max}} = \frac{3^{3/2}}{2}\left(\frac{\Delta p}{p_{\max}}\right)^{1/2}\left(1 - \frac{\Delta p}{p_{\max}}\right) , \qquad (15.127)$$

with just two adjustable parameters: $U_{\max}$ the maximum flow rate and $p_{\max}$, the static pressure required to force the reed completely shut. The maximum flow occurs when $\Delta p/p_{\max} = 1/3$.

Double Reeds

Instruments like the oboe, bassoon and bagpipe chanters use double reeds, which close against each other with a relatively long and narrow constricted air channel on the downstream side before entering the instrument. The turbulent air motion in the constricted air passage would result in an additional turbulent resistance proportional to the flow velocity squared, which would add to the pressure difference across the reed. This could result in strongly hysteretic re-entrant static velocity flow characteristics as a function of the total pressure across the reed and lead pipe (see, for example, *Wijnands* and *Hirschberg* [15.151]).

A recent comparison of the flow-pressure characteristics of oboe and bassoon double reeds and a clarinet single reed, Fig. 15.94, by *Almeida* [15.146]) shows no evidence for re-entrant double-reed features. Nevertheless, the measurements are strongly hysteretic, because of changes in the properties of the reeds (elasticity and mass), as they absorb and desorb moisture from the damp air passing through them. In the measurements the static pressure was slowly increased from zero to its maximum value and then back again. Under normal playing conditions, one might expect to play on a non-hysteretic operating characteristic somewhere between the two extremes of the hysteretic static measurements. Thus, although the shape of the flow-pressure curves for the double reeds differs significantly from those of the clarinet single reed, the general form is qualitatively similar, with a region of dynamic negative resistance above the peak flow. The strongly moisture dependent properties of reeds are very familiar to the player, who has to moisten and "play-in" a reed before it is ready for performance.

There therefore appears to be no fundamental difference between the way single and double reeds operate. Indeed, the sound of an oboe is apparently scarcely changed, when played with a miniature clarinet reed mouthpiece instead of a conventional double reed (Campbell and Gilbert, private communication).

Dynamic Characteristics

Fletcher [15.152] extended the quasistatic model by assuming the reed could be described as a simple mass–

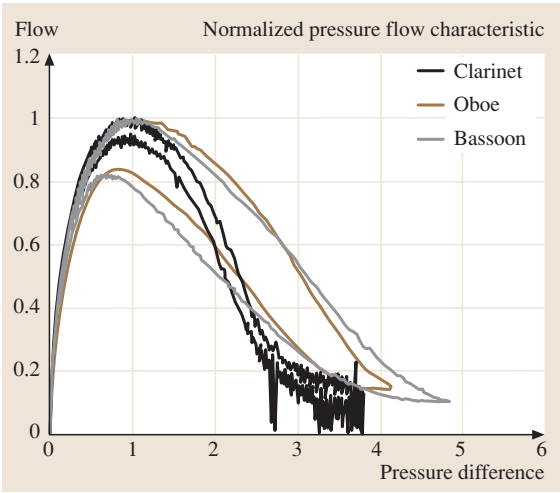


Fig. 15.94 A comparison of the normalised, hysteric static pressure/flow characteristics of single (clarinet) and double (oboe and bassoon) reeds measured on first increasing and then decreasing the flow rate of moist air through the reeds (after [15.146])

Amazon
EX1013 Page 631

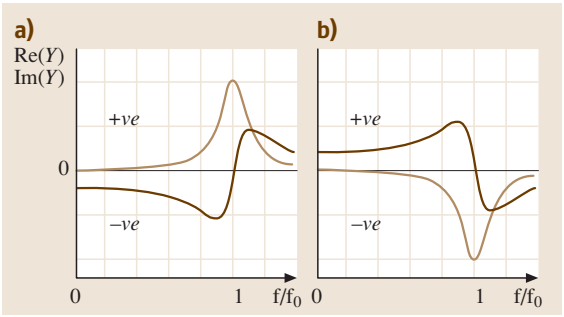


Fig. 15.95a,b Real (*dark brown*) and imaginary (*light brown*) components of the reed admittance $Y(\omega)$ for **(a)** an inward-closing reed in the negative dynamic conductance regime and for **(b)** an outward-opening reed, as a function of frequency normalised to the resonant frequency of the reed for reeds having Q-values of 5

spring resonator resulting in a dynamic conductance of

$$Y(\omega) = Y(0)\frac{1}{1-(\omega/\omega_0)^2 + \mathrm{i}\omega/\omega_0 Q} \qquad (15.128)$$

where $Y(0) = \partial U/\partial\,(\Delta p)|_{\omega=0}$ is the quasistatic flow conductance and the denominator describes the dynamic resonant response of the reed. The Q-value is determined by viscous and mechanical losses in the reed.

The resistive and reactive components of $Y(\omega)_{\mathrm{reed}}$ given by (15.128) are plotted in Fig. 15.95a,b for an inward-closing reed $(-,+)$, in the negative flow conductance regime above the velocity flow maximum, and for the outward-closing $(+,-)$ reed. As discussed qualitatively above, the negative input dynamic conductance of the inward-striking reed remains negative at all frequencies below its resonant frequency, whereas the conductance of the outward-opening reed only becomes negative above its resonant frequency.

For the oscillations of any attached air column to grow, feedback theory requires that

$$\begin{aligned} &\mathrm{Im}(Y_\mathrm{r} + Y_\mathrm{p}) = 0 \\ &\mathrm{Re}(Y_\mathrm{r} + Y_\mathrm{p}) < 0\,, \end{aligned} \qquad (15.129)$$

where Y_p and Y_r are the admittances of the pipe and reed, respectively. The negative dynamic conductance of the reed must therefore be sufficiently small to overcome the losses in the instrument. Furthermore, the reactive components of the reed conductance will perturb the frequencies of the attached air column.

Fletcher and *Rossing* ([15.5], Chap. 13) give an extended discussion of the dynamics of reed generators including polar plots of admittance curves for typical outward and inward-striking reed generators as a function of blowing pressure.

For the inward-striking reeds of the clarinet, oboe and bassoon, the real part of the reed admittance is negative below the resonant frequency of the reed. For the oboe this is typically around 3 kHz, above the pitch of the reed attached to its staple (joining section) alone (CD-ROM). However, when attached to an instrument, the negative conductance will excite the lower-frequency natural resonances of the attached tube (CD-ROM). In this regime, the reactive load presented by the reed is relatively small and positive and equivalent to a capacitive or spring loading at the input end of the attached pipe. This results in a slight increase in the effective length of the pipe and a slight lowering of the frequencies of the resonating air column.

Free reeds, like the vibrating brass cantilevers used in the mouth organ, harmonium and certain organ reed pipes (Fig. 15.89c), are rather weakly damped inward-closing $(-,+)$ reeds (*Fletcher* and *Rossing* [15.5], Sect. 13.4). The reed is initially open. High pressure on one side or suction on the other (as in the harmonium or American organ) forces the reed back into the aperture, controlling the air flow. Like the clarinet reed, above a certain applied pressure the reed will close and restrict the flow resulting in a negative conductance regime. If the reed is forced right through the aperture, it becomes an outward-opening $(+,-)$ reed with a positive conductance. Because of its low damping, the blown-closed reed tends to vibrate at a frequency rather close to its resonant frequency. In practice, as soon as a threshold pressure is reached that is significantly below the maximum in the static characteristics, a harmonium reed starts to vibrate with a rather large sinusoidal amplitude (typically ≈4 mm) resulting in highly non-sinusoidal flow of air through the aperture (*Koopman* et al. [15.153]). For such high-Q-value mechanical resonators, the vibrational frequency is strongly controlled by the resonant frequency of the reed itself, so that a separate reed is needed for each note, as in the piano accordion, harmonium, mouth organ and reed organ pipes. The reeds in a mouth organ are arranged in pairs in line with the flow of air. They are individually excited by overpressure and suction.

In contrast, the dynamic conductance of an outward-opening reed $(+,-)$ is only negative above its resonant frequency. The conductance then decreases rather rapidly with increasing frequency, so that there may only be a relatively narrow range of frequencies above resonance over which oscillations can occur. The vibrating lips provide a possible example of such a reed,

Amazon
EX1013 Page 632

with an increase in steady-state pressure always increasing the static flow through them. Above their resonant frequency, the dynamic conductance becomes negative and could excite oscillations in an attached pipe. In such a regime, the reactive component of the reed admittance is negative. This corresponds to an inductive or inertial load, which will shorten the effective length of the air column and increase its resonant frequencies. The influence of the reed on the resonant frequencies of an attached instrument therefore provides a valuable clue to the way in which a valve is operating, as we will discuss later in relation to the vibrations of the lips in a brass-instrument mouthpiece.

Small-Amplitude Oscillations

We now consider the stability of the oscillations excited by the negative dynamic conductance of the reed. In particular, it is interesting to consider whether the oscillations, once initiated, are stabilised or grow quickly in amplitude into a highly nonlinear regime. Surprisingly, this depends on the bore shape of the attached air column, as discussed by *Dalmont* et al. [15.154]. Several authors have investigated such problems, including *Backus* [15.155] using simple theoretical models and measurements, *Fletcher* [15.156] using analytic models, *Schumacher* [15.157,158] in the time rather than the frequency domain, and *Gilbert* et al. [15.154,159] using a harmonic balance approach, which we will briefly outline in the following section. Recent overviews of the nonlinear dynamics of both wind and brass instruments have been published by *Campbell* [15.160] and by *Dalmont* et al. [15.154].

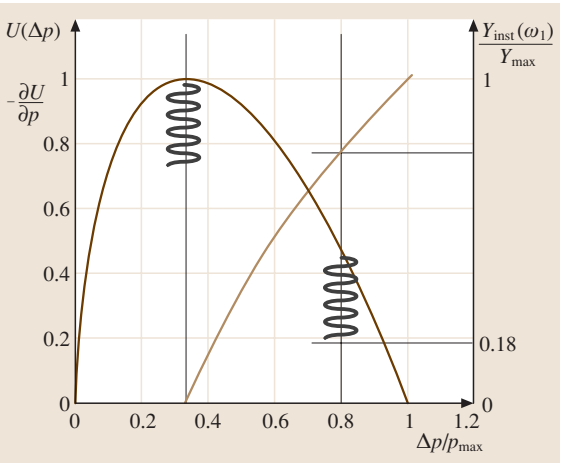


Fig. 15.96 Plot of normalised flow rate and differential negative conductance of an inward-striking reed valve as a function of pressure across the reed normalised to the pressure required for closure. The intersection of the negative conductance curve with the real part of the input admittance $Y_{inst}(\omega)$ of the instrument determines the pressure in the mouthpiece for the onset of oscillations, illustrated schematically for $Y_{inst}(\omega) \approx 0.78 Y_{max}$

We first consider the excitation of small-amplitude oscillations based on the reed equation, which is replotted in Fig. 15.96 as a universal curve together with the negative dynamic admittance or conductance, $-\partial U/\partial p$ above the flow-rate maximum.

The onset of self-oscillations occurs when the sum of the real and the imaginary parts of the admittance of the reed and attached instrument are both zero (15.129). If losses in the reed and the attached instrument were negligible, resonances of the air column would be excited as soon as the mouthpiece pressure exceeded $\frac{1}{3} p_{max}$. However, when losses are included, the negative conductance of the reed has to be sufficiently large to overcome the losses in the instrument. The onset then occurs at a higher pressure, as illustrated schematically in Fig. 15.96.

The onset of oscillations depends not only on the mouthpiece pressure but also on the properties of the reed, such as the initial air gap and its flexibility, which will depend on its thickness and elastic properties. The elastic properties also change on the take up of moisture during playing. It is not surprising that wind players take great care in selecting their reeds. Furthermore, notes are generally tongued. This involves pressing the tongue against the lip of the reed to stop air passing through it, so that the pressure builds up to a level well above that required to just excite the note. When the tongue is removed, the note sounds almost immediately giving a much more precise definition to the start of a note.

The transition to the oscillatory state can be considered using the method of small variations and harmonic balance (*Gilbert* et al. [15.159]). For a given mouth pressure defining the overall flow rate, oscillations of the flow rate u can be written as a Taylor expansion of the accompanying small pressure fluctuations p at the output of the reed, such that

$$u = Ap + Bp^2 + Cp^3 + \dots . \qquad (15.130)$$

From the reed equation plotted in Fig. 15.96, the coefficient A is positive for mouthpiece pressures above $p_{max}/3$, while B and C are positive and negative respectively. We look for a periodic solution with Fourier components that are integer multiples of the fundamental

Amazon
EX1013 Page 633

frequency ω, so that

$$p(t) = \sum p_n \, \mathrm{e}^{\mathrm{i}n\omega t} \, , \tag{15.131}$$

with a corresponding oscillatory flow $u(t)$ superimposed on the static flow U,

$$u(t) = \sum u_n \, \mathrm{e}^{\mathrm{i}n\omega t} \, . \tag{15.132}$$

At the input to the instrument, the oscillatory flow can be expressed in terms of the Fourier components of the input pressure and admittance, so that

$$u(t) = \sum p_n Y(n\omega) \, \mathrm{e}^{\mathrm{i}n\omega t} \, . \tag{15.133}$$

Using the method of harmonic balance, we equate the coefficients of the Fourier components in (15.130) with those in having substituted for the pressure. The first three Fourier components are then given by

$$\begin{aligned} p_1 &= \pm \left(\frac{Y_1 - A}{2B^2/(Y_2 - A) + 3C} \right) \, , \\ p_2 &= \left(\frac{B}{Y_2 - A} \right) p_1^2 \, , \\ p_3 &= \left(\frac{C}{Y_3 - A} \right)^{1/2} p_1^3 \, . \end{aligned} \tag{15.134}$$

For a cylindrical-bore instrument like the clarinet at low frequencies, only the odd-n partials will be strongly excited, so that Y_2 is very large. The amplitude of the fundamental component is then given by $\pm[(Y_1 - A)/C]^{1/2}$, with a vanishingly small second harmonic p_2. The amplitude p_1 of small oscillations is then stabilised by the cubic coefficient C. Stable, small-amplitude oscillations can therefore be excited as soon as the negative conductance of the reed exceeds the combined admittance of the instrument and any additional losses in the reed itself.

Because the transition is continuous, p_1 rises smoothly from zero, taking either positive or negative values (simply solutions with opposite phases). The transition is therefore referred to as a *direct bifurcation*. The player can vary the pressure in the mouthpiece and the pressure of the lips on the reed to vary the coefficients A and C and hence can control the amplitude of the excited sound continuously from a very quiet to a loud sound, as often exploited by the skilled clarinet player.

In contrast, for a conical-bore instrument, the amplitude of the fundamental,

$$p_1 = \pm \left(\frac{(Y_1 - A)(Y_2 - A)}{2B^2 + 3C(Y_2 - A)} \right)^{1/2} \, , \tag{15.135}$$

involves the admittance of both the fundamental and second partial. On smoothly increasing A by increasing the pressure on the reed, *Grand* et al. [15.161] showed that there can again be a direct smooth bifurcation to small-amplitude oscillations, if $2B^2 > -3C(Y_2 - Y_1)$. However, if this condition is not met, there will be an indirect transition, with a sudden jump to a finite-amplitude oscillating state. This gives rise to the hysteresis in the amplitude as the mouth pressure is first increased and then decreased. This means that the player may have to exert a larger pressure to sound the note initially, but can then relax the pressure to produce a rather quieter sound. It may also explain why it is more difficult to initiate a very quiet note on the saxophone with a conical bore than it is on the clarinet with a cylindrical bore.

For a direct bifurcation transition, the small nonlinearities in the dynamic conductance will result in a spectrum of partials with amplitudes varying as p_1^n, where p_1 is the amplitude of the fundamental component excited. The spectral content or *timbre* of wind and brass instruments, as discussed later, therefore changes with increasing amplitude. This is illustrated by measurements of the amplitude dependence of the partials of a trumpet, clarinet and oboe by *Benade* ([15.133], Fig. 21.6c), which are reproduced for the trumpet in Fig. 15.104. For the largely cylindrical-bore trumpet and clarinet, nonlinearity results in partials varying as p_1^n over quite a large range of amplitudes. However, for the oboe, with its conical bore, the relative increase in strength of the partials is rather more complicated (*Benade* [15.133], Sect. 21.3). Eventually, the small-amplitude approximation will always break down, with a transition to a strongly nonlinear regime. Benade associates this transition with a change in timbre and responsiveness of the instrument for the player.

Large-Amplitude Oscillations

For a *lossless* cylindrical-bore instrument with only odd integer partials, the large-amplitude solutions are particularly simple. The flow of air from the lungs and pressure in the mouth is assumed to remain constant resulting in an average flow rate through the instrument. The pressure at the exit of the reed then switches periodically from a high-pressure to a low-pressure state, with equal amplitudes above and below the mean mouth pressure, spending equal times in each. The net acoustic energy fed into the resonating air column per cycle is therefore zero, $U \int p(t) \mathrm{d}t = 0$, as required for a lossless system.

Such a solution can easily be understood in terms of the excess-pressure wave propagating to the open

Amazon
EX1013 Page 634

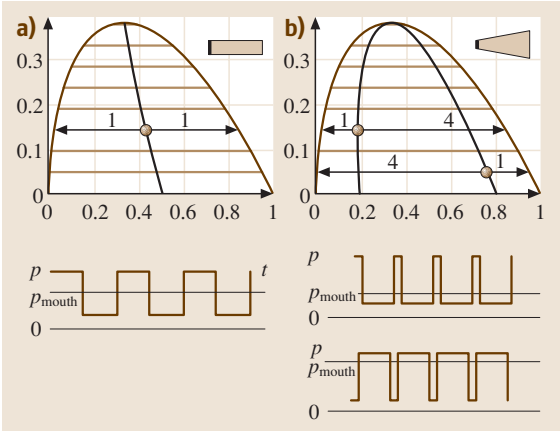


Fig. 15.97a,b Large-amplitude Helmholtz pressure fluctuation of **(a)** a cylinder and **(b)** a truncated cone with length to apex of 1/4 of its length, illustrating the dependence of fluctuation amplitudes as a function of mouthpiece pressure. For the cylinder, the mouthpiece pressure is single valued for a given flow rate, but for the truncated cone there are two possible solutions referred to as the standard and inverted Helmholtz solutions

end of the instrument, where it is reflected with change of sign. On return to the reed it reverses the pressure difference across the reed, which switches to the reduced pressure state. The subsequent reflection of the reduced pressure wave then switches the reed back to its original high-pressure state and the process repeats indefinitely, with a periodic time of $4L/c_0$, as expected.

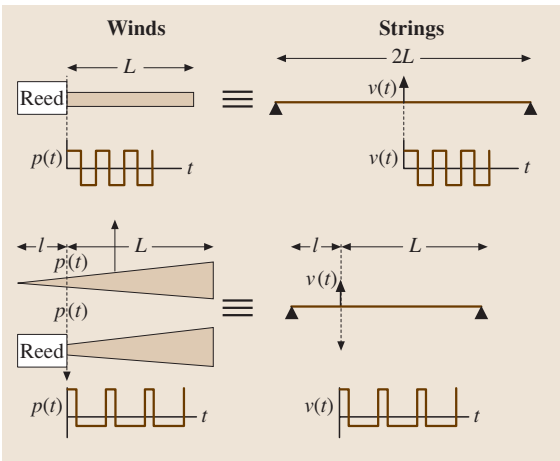


Fig. 15.98 Analogy between large-amplitude pressure waves in the bores of wind instruments and the transverse velocity of Helmholtz waves on a bowed stretched string, where the reed position is equivalent to the bowing position

The dependence of the square-wave pressure fluctuations on the applied pressure can be obtained by the simple graphical construction illustrated in Fig. 15.97. The locus of the static pressure required to excite square-wave pressure fluctuations above and below the mouth pressure is shown by the solid line drawn from $p_{mouth} = 1/3$ to $1/2\,p_{max}$, which bisects the high and low pressures for a given flow rate. If losses are taken into account, the horizontal lines are replaced by load lines with a downward slope given by the real part of the instrument's input admittance (*Fletcher* and *Rossing* [15.5], Fig. 15.9). At large amplitudes, the solutions can then involve periods during which the reed is completely closed. The transition from small-amplitude to large-amplitude solutions is clearly of musical importance, as it changes the sound of an instrument, and remains an active area of research [15.154].

Analogy with Bowed String

In recent years, an interesting analogy has been noted between the large-amplitude pressure fluctuations of a vibrating air column in a cylindrical or truncated conical tube and simple Helmholtz waves excited on a bowed string (*Dalmont* and *Kergomard* [15.162]). For example, the square-wave pressure fluctuations at the output of the reed attached to a cylindrical tube are analogous to the velocity of the bowed Helmholtz transverse waves of a string bowed at its centre, illustrated schematically in Fig. 15.98. Helmholtz waves could equally well be excited on a bowed string by a transducer with a square-wave velocity output placed halfway along the length of the string, in just the same way that the reed with a square-wave pressure output excites Helmholtz sound pressure waves into a cylinder, which acts like half the string length.

The analogy is particularly useful in discussing the large-amplitude pressure fluctuations in conical-bore instruments such as the oboe or saxophone. As described earlier, the conical tube has the same set of resonances as a cylindrical tube that is open at both ends. Therefore, in addition to having the same set of standing-wave sinusoidal solutions for the transverse oscillations of a stretched string, a conical tube can also support Helmholtz wave solutions. For the bowed string, the closer one gets to either end of the string the larger becomes the mark-to-space ratio between the regions of high to low transverse velocity. The same is also true for the switched fluctuations in pressure in a lossless conical tube, shown schematically in Fig. 15.97. Hence, if one

Amazon
EX1013 Page 635

truncates a conical tube with a vibrating reed system, the resonant modes of the remaining air column will be unchanged, provided the vibrating reed produces the same pressure fluctuations that would otherwise have been produced by the reflected Helmholtz waves returning from the removed apex end of the cone. Hence a conical tube, truncated by a vibrating reed at a distance l from the apex, can support Helmholtz wave solutions in the remaining length L. To produce such a wave the reed has to generate a rectangular pressure wave with a mark-to-space ratio and pressure fluctuations about the mean in the ratio $L : l$, as illustrated in Fig. 15.98, for a truncated cone with $L/l = 4$.

The period of the Helmholtz wave solutions of a conical bore instrument modelled as a truncated cone will therefore be $2(L+l)/c_0$, with a spectrum including all the harmonics $f_n = nc_0/2(L+l)$, other than those with integer values of $n = (L+l)$. To determine the amplitude of the rectangular pressure wave as a function of mouthpiece pressure, a graphical construction can be used similar to that used for the cylindrical tube, except that the pressures have now to be in the ratio L/l, as indicated in Fig. 15.97b. For the lossless large-amplitude modes of a truncated cone, there are two possible solutions involving high and low mouth pressures, which are known as the standard and inverted Helmholtz solutions, respectively.

Any complete model of a reed driven instrument must include losses and departures from harmonicity of an instrument's partials. This leads to a rounding of the rectangular edges of the Helmholtz waveforms and additional structure, in much the same way that bowed string waveforms are perturbed by frictional forces and non-ideal reflections at the end-supports (Fig. 15.34). Figure 15.99 shows a typical pressure waveform input for the conical-bore saxophone, which is compared with the Helmholtz waveform predicted for an ideal lossless system (*Dalmont* et al. [15.162]).

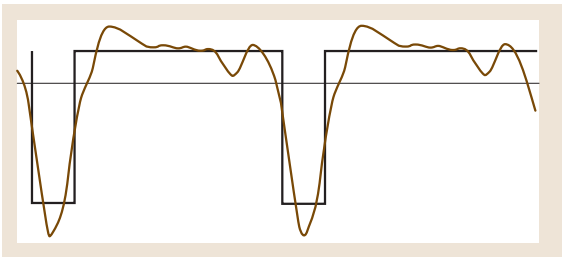


Fig. 15.99 Measured pressure waveform at input to a saxophone compared with the Helmholtz waveform expected for a truncated cone (after *Dalmont* et al. [15.162])

A completely realistic model for the excitation of sound in wind instruments must also include coupling to the vocal tracts (*Backus* [15.163] and *Scavone* [15.164]), since the assumption of a constant flow rate and constant mouth pressure is clearly over-simplistic.

Register Key

This analysis implicitly assumes that the reed excites the fundamental mode of the attached instrument. In practice, the reed will generally excite the partial with the lowest admittance corresponding to the highest peak in impedance measurements. For most instruments this is usually the fundamental resonance. However, the amplitude of the fundamental can be reduced relative to the higher resonances by opening a small hole, the register hole, positioned between the reed and first hole used in normal fingering of the instrument. Because of the difference in wavelengths and position of nodes, opening the register hole preferentially reduces the Q-value and shifts the frequency and amplitude of the fundamental relative to the higher partials. This allows the player to excite the upper register of notes based on the second mode, which in the case of the conical-bore saxophone, oboe and bassoon is an octave above the fundamental but is an octave and a perfect fifth above the fundamental for the cylindrical-bore clarinet. The lower and upper registers of the clarinet are sometimes referred to as the chalumeau and chanter registers, after the earlier instruments from which the clarinet was derived.

Figure 15.100 illustrates the lowering in amplitude and shift in resonant frequency of the fundamental mode on opening the register hole for the note E3 on a clarinet, which leaves the upper partials relatively unchanged (*Backus* [15.132]). The measurements also show the sig-

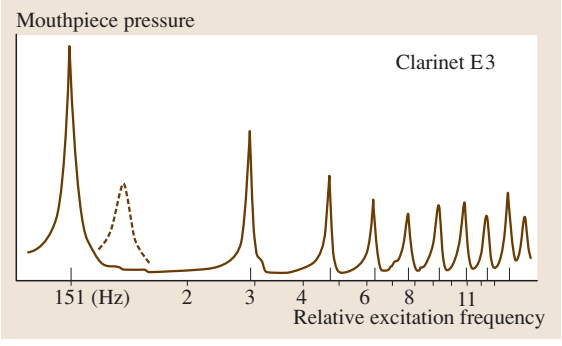


Fig. 15.100 Resonance curves for the note E3 on a clarinet, showing the shift of the lowest partial on opening the register hole (after *Backus* [15.132])

Amazon
EX1013 Page 636

nificant departures from the 1, 3, 5, 7 harmonicity of the resonant modes of the instrument, which act as a warning not to take ideal models for the harmonicity of modes in real wind instruments too literally. Fletcher has developed a mode-locking model to account for the excitation of periodic waveforms on instruments with inharmonic partials [15.165].

15.3.4 Brass-Mouthpiece Excitation

The excitation of sound by the vibrating lips in the mouthpiece of a brass instrument cannot be described by any of the above simple models alone, which consider the reed as a simple mass–spring resonator. As we will show, the vibrations of the lips are three-dimensional and much more complicated. As a result, in some regimes the lips behave rather like outward-swinging-door valves, as first proposed by Helmholtz, and Bernoulli pressure operated sliding-door reeds in others. In addition the air flow is also affected by three-dimensional wave-like vibrations on the surface of the lips.

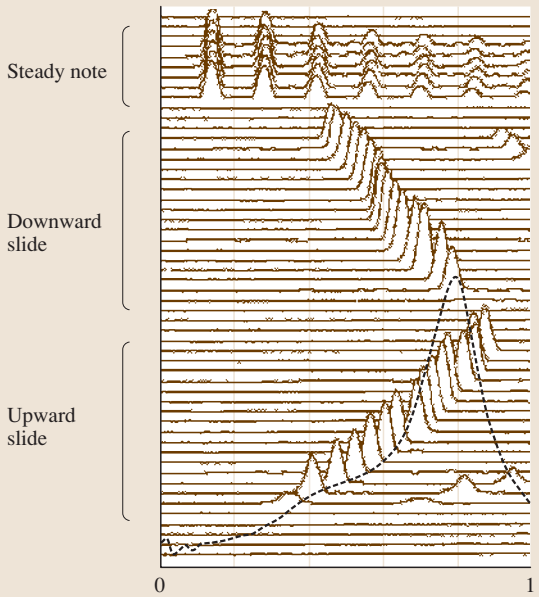


Fig. 15.101 Time sequence (*from bottom to top*) of spectra of the sound produced by a player "buzzing" into a trumpet mouthpiece, first for an upward slide in frequency, then for a downward slide and finally for a steady low note at high intensity showing the excitation of a note with many Fourier components. The *dashed line* is the spectrum of the *popping* note excited by slapping the open end of the mouthpiece against the palm of the hand (after *Ayers* [15.166])

When playing brass instruments, the lips are firmly pressed against the rim of the mouthpiece with the lips pouted inwards. Pitched notes are produced by blowing air through the tightly clenched lips to produce a buzzing sound. The excitation mechanism can easily be demonstrated by buzzing the lips alone, though it is difficult to produce a very wide range of pitched sounds However, if the lips are buzzed when pressed against the rim of a mouthpiece, the input rim provides an additional constraint on the motion of the lips, which makes it much easier to produce pitched notes over a wide range of frequencies (CD-ROM). The audio demonstrates the "popping" sounds of trumpet and horn mouthpieces followed by the sound of the player buzzing the mouthpiece alone up to a pitch close to the popping frequency and back again. Attaching the mouthpiece to an instrument locks the oscillations to the various possible resonances of the instrument (CD-ROM).

Figure 15.101 shows a series of time-sequence plots of spectra of the sound produced by a player "buzzing" into a trumpet mouthpiece (*Ayers* [15.166]), which acts rather like a simple Helmholtz resonator. In the lower sequence, the player excites well-defined pitched notes from low frequencies up to slightly above the mouthpiece *popping frequency* (see Sect. 15.3.3). The middle set of traces shows the spectrum as the player starts at a high frequency and lowers the pitch. The upper traces shows the spectrum of a loudly sounded, low-frequency, note, illustrating the rich spectrum of harmonics produced by the strongly nonlinear sound-generation processes involved (see, for example, *Elliot* and *Bowsher* [15.167]).

These measurements on the mouthpiece alone strongly suggest that the lips can generate periodic fluctuations at frequencies up to, but not significantly beyond, the resonant frequency of any coupled acoustic resonator. This was confirmed in an investigation of lip-reed excitation using a simple single-resonant-frequency Helmholtz resonator by *Chen* and *Weinreich* [15.168], who used a microphone and loudspeaker in a feedback loop to vary the Q-value of the Helmholtz resonator played using a normal mouthpiece. They concluded that a player could adjust the way they vibrated their lips in the mouthpiece to produce notes that were slightly higher or lower than the Helmholtz frequency, though the most natural playing parameters generated frequencies in the range 20–350 Hz below the resonator frequency.

Amazon
EX1013 Page 637

Attached Mouthpiece

Ayers [15.166] also compared the frequencies produced in the mouthpiece before and immediately after attachment of the instrument. In these measurements the player first excited a pitched note in the mouthpiece with the instrument effectively decoupled by opening a large hole in its bore close to the mouthpiece. The hole was then closed and the immediate change in frequency measured before the player had time to make any adjustments to the embouchure. The results of such measurements are shown in Fig. 15.102, where the diagonal line represents the pitched notes before the instrument was connected and the discontinuous solid line through the triangular points are the modified frequencies for the same embouchure with the instrument connected.

At the higher frequencies, the jumps between successive branches are from just above the resonant frequency of one partial to just below the resonant frequency of the next, with a monotonic increase in frequency on tightening the embouchure in between. However, for the first two branches, the instrument resonances initially have a relatively small effect on the frequencies excited by the mouthpiece alone until such frequencies approach a particular partial frequency. The frequency then approaches the resonant frequency of the instrument before jumping to a frequency well below the next partial and the sequence repeats. The difference in behaviour of the lower and higher branches suggests that more than one type of lip-reed action is involved.

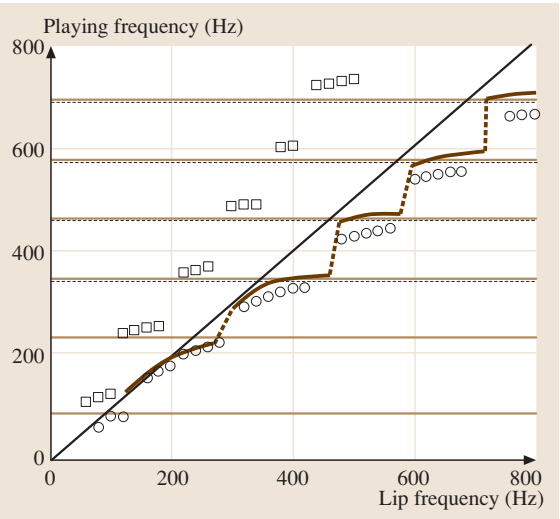


Fig. 15.102 Frequencies produced by a trumpet mouthpiece without (*diagonal line*) and with (*broken line*) the instrument strongly coupled using an unchanged embouchure under the same playing conditions. The *solid horizontal lines* are the resonant frequencies of the assembled trumpet. The *squares* and *circles* are predictions for the Helmholtz outwardly opening-door and sliding-door models computed by *Adachi* and *Sato* [15.169] for a trumpet with slightly lower-frequency resonant modes indicated by the *dashed horizontal lines* (after *Ayers* [15.166])

Comparison with computational models by *Adachi* and *Sato* [15.169] appear to rule out the outward-swinging-door model first proposed by Helmholtz, indicated by the squares in Fig. 15.101, as the predicted frequencies are always well above those of the instrument's partials. The computed predictions for the Bernoulli sliding door model, indicated by the circles, are in better agreement with measurements, but with predicted frequencies rather lower than those observed and never rising above the resonant frequencies of the instrument, in contrast to the observed behaviour for the higher modes excited.

Any model for the lip-reed sound generator has to explain all such measurements. Such measurements also highlight the way that a brass player can adjust the embouchure and pressure acting on the lips in the mouthpiece to change the frequency of the excited mode. On tightening the embouchure and pressure the player can progressively excite successive modes and can "lip" the pitch of the note up and down by surprisingly large amounts, used with great expressive effect by jazz trumpeters.

Vibrating Lips

In practice, the production of sound by the vibrating lips inside a mouthpiece is a highly complex three-dimensional problem closely analogous to the production of sound by the vocal folds – see, for example,



Fig. 15.103 High-speed photograph clips showing one cycle of lip vibration in a trombone mouthpiece

Amazon
EX1013 Page 638

Titze [15.170]. A complete model would involve solving the coupled solutions of the Navier–Stokes equations describing the flow of air from the mouth, through the lips and into the mouthpiece, and the three-dimensional motions of the soft tissues of the lips induced by the Bernoulli pressures acting on their surfaces.

Stroboscopic and ultra-fast photography of the brass players lips while playing reveal highly complex three-dimensional motions (*Coppley* and *Strong* [15.171], *Yoshikawa* and *Muto* [15.172]) suggest that the upper lip is primarily involved in the generation of sound. Figure 15.103 and the video clip |CD-ROM provided by Murray Campbell show high-speed photography clips of such motion. The lips inside the mouthpiece open and shut rather like the gasping opening and shutting of the mouth of a goldfish, but speeded up several hundred times. Points on the surface of the upper lip exhibit cyclic orbital motions involving the in-quadrature motions of the upper lip parallel and perpendicular to the flow. To model such motion clearly requires at least two independent mass–spring systems to account for the induced motions of the lips along and perpendicular to the flow (*Adachi* and *Sato* [15.169]). In addition, there is a pulsating wave-like motion along the surface of the lips in the direction of air flow, with the rear portion of lips moving in anti-phase with the front. *Yoshikawa* and *Muto* [15.172] identify such motion as strongly damped Rayleigh surface waves travelling through the mucous tissue of the upper lip.

The simplest possible model to describe such motion therefore requires at least three interacting mass–spring elements; one to describe the lip motion along the direction of flow, and two to describe the motions of the front and back surfaces of the lips. But even then, the model will still only be an approximation to the three-dimensional bulk tissue motions involved. Not surprisingly, research into the lip-reed sound-excitation mechanism remains a problem of considerable interest.

Artificial Lips

To achieve a better understanding of lip dynamics and its effect on the sound produced by brass instrument, several groups have developed artificial lips to excite brass instrument (e.g. *Gilbert* et al. [15.173] and *Cullen* et al. [15.174]). These can be used to investigate instruments under well-controlled and reproducible experimental conditions. Typically, the lips are simulated by two slightly separated thin-walled (0.3 mm thickness) latex tubes filled with water under a controlled pressure. The tubes are rigidly supported from behind so that the internal pressure forces the lips together. The tubes are placed across an opening in an otherwise hermetically sealed unit that represents the mouth and throat cavities. Air is fed into the mouth cavity at a constant flow rate. A fixed mouthpiece is then pushed against the artificial lips with a measured force. By varying this force, the applied pressure and the pressure within the artificial lips, the experimenter can simulate the various ways in which a player can control the dynamics of the lips (the *embouchure*) to produce different sounding notes. Despite the considerable simplification in comparison with the dynamics of real lips, the sound of brass instruments played by artificial lips is extremely close to that produced by a real player. Such systems enable acoustical studies to be made on brass instruments with a much greater degree of flexibility, reproducibility and stability than can be achieved by a player. Using a fixed mouthpiece, the playing characteristics of different attached instruments can easily be compared.

Nonlinear Sound Excitation

When played very quietly, brass instruments can produce sounds that are quasi-sinusoidal with relatively weak higher harmonics. However, as previously noted for vi-

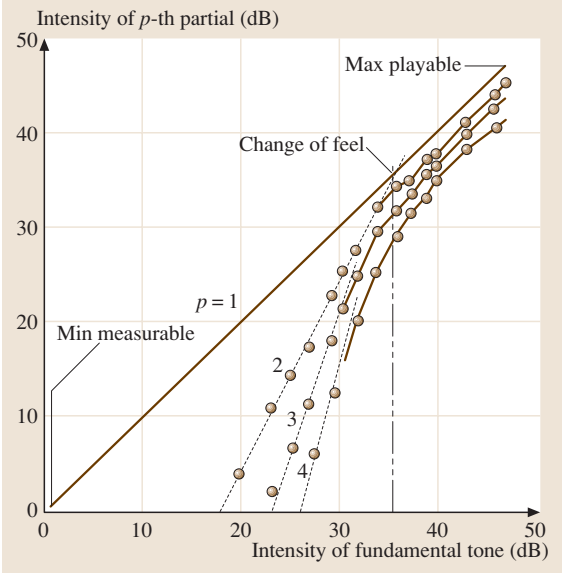


Fig. 15.104 Intensity of the first four partials of the note C4 on a trumpet as a function of the intensity of the first partial, measured from the minimum to the maximum playing intensity (after *Benade* [15.133] Fig. 21.8). On this logarithmic scale, the dashed lines through the measurements for the second, third and fourth Fourier components have slopes 2, 3 and 4, respectively

Amazon
EX1013 Page 639

brating reeds, any nonlinearity will lead to the generation of harmonics of the fundamental frequency ω at frequencies 2ω, 3ω, etc., with initially amplitudes increasing as $|p_0(\omega)|^n$, where $p_0(\omega)$ is the amplitude of the fundamental. However, in the strongly nonlinear region at high amplitudes, all partials become important and increase in much the same way with increasing driving force (*Fletcher* [15.175] and *Fletcher* and *Rossing* [15.5], Sects. 14.6 and 14.7). Such effects are illustrated in Fig. 15.104 for measurements on a B-flat trumpet by *Benade* and *Worman* ([15.133], Sect. 21.3). The spectral content and resulting brilliance of the sound or timbre of a trumpet, or any other brass instrument, therefore depends on the intensity with which the instrument is played. Benade noted a change in the sound and feel of an instrument by the player in the transition region between the power-law dependence of the Fourier component and the high-amplitude regime, where the harmonic ratios remain almost constant. Similar characteristics were observed for the clarinet though rather different characteristics for the oboe.

Examples of the strongly non-sinusoidal periodic fluctuations of the pressure and flow velocity within the mouthpiece for two loudly played notes on a trombone are shown in Fig. 15.105 (*Elliot* and *Bowsher* [15.167]). *Fletcher* and *Rossing* ([15.5], Sect. 14.7) discuss such waveforms in terms of the lips operating slightly above

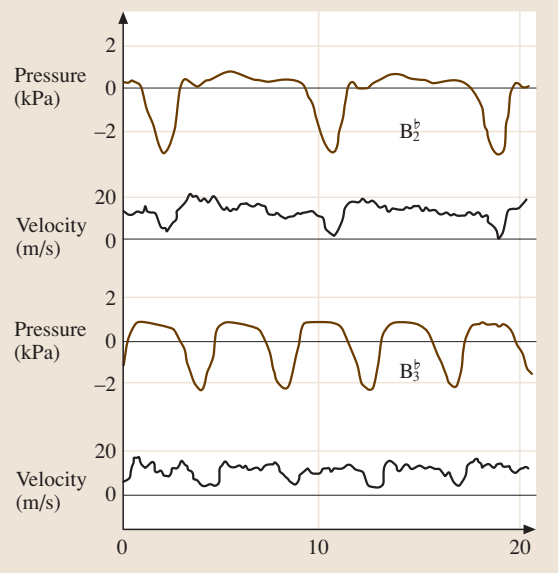


Fig. 15.105 Non-sinusoidal pressure fluctuations within the mouthpiece for two notes played at large amplitudes on the trombone (after *Elliot* and *Bowsher* [15.167])

the resonant frequency of their outward-swinging-door resonant frequencies.

The nonlinearity of the lip-reed excitation mechanism enables the player to vibrate the lips at the frequency of the missing fundamental of the quasi-harmonic series of modes of brass instruments, illustrated in Fig. 15.38. This is referred to as the *pedal note* and is an octave below the lowest mode normally excited on the instrument. The lips vibrate at the pedal-note frequency but only excite the quasi-harmonic $n = 2, 3, 4, \ldots$ modes.

The pressure fluctuations in Fig. 15.104 of $\approx 3\,\mathrm{kPa}$ correspond to a sound intensity of nearly 160 dB. As the static pressure in the mouthpiece is only a few percent above atmospheric pressure ($10^5\,\mathrm{Pa}$), such pressure excursions are a significant fraction of the excess static pressure. Even larger-amplitude pressure fluctuations can be excited on the trumpet and trombone when played really loudly, to produce a brassy sound. *Long* [15.176] has recorded pressure levels in a trumpet mouthpiece as high as 175 dB, corresponding to pressure fluctuations of $\approx 20\,\mathrm{kPa}$.

At such high amplitudes, one can no longer neglect the change in density of a gas when considering its acceleration under the influence of the pressure gradient. To a first approximation, the wave equation then becomes

$$\frac{\partial^2 \xi}{\partial x^2} = \left(1 + \frac{\partial \xi}{\partial x}\right) \frac{1}{c_0^2} \frac{\partial^2 \xi}{\partial t^2} \,. \tag{15.136}$$

The speed of sound will now depend on both frequency and wave shape, with the velocity varying as

$$c' = c_0 \left\langle \left(1 + \frac{\partial \xi}{\partial x}\right)^{-1/2} \right\rangle_{x,t} \sim c_0 \left[1 + \alpha\,(k\xi)^2\right] \,, \tag{15.137}$$

where the averaging is taken over both time and wavelength giving a value for $\alpha \approx 1/8$. In practice, for waves propagating down a tube, other terms involving momentum transport, viscosity and heat transfer also have to be included in any exact solution. However, the essential physics remains unchanged. The net effect of the nonlinearity is to progressively increase the slope of the leading edge of any wave propagating along the tube. A propagating sine wave is then transformed into a sawtooth waveform, or shockwave, with an infinitely steep leading edge, at sufficiently large distances along the tube.

Such waves have indeed been observed in trumpet and trombone bores, which are long and relatively narrow. The effect is illustrated in Fig. 15.106 by the measurements of *Hirschberg* et al. [15.177], for waves

Amazon
EX1013 Page 640

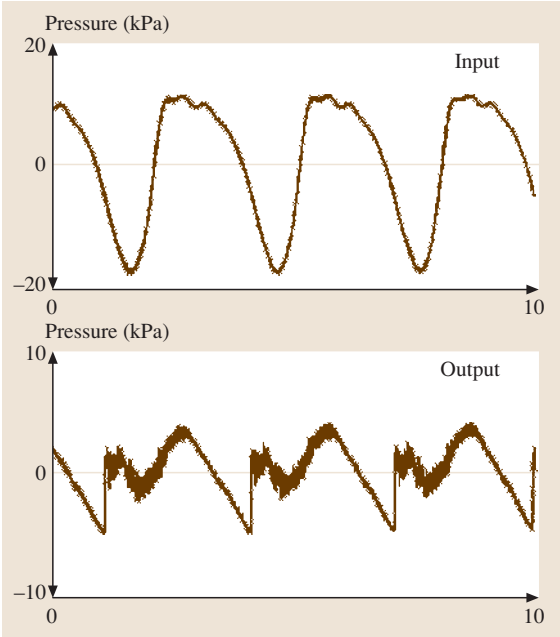


Fig. 15.106 Internal acoustic pressure in a trombone played at a dynamic level fortissimo (ff). *Upper curve*: pressure at the input to the air column in the mouthpiece. *Lower curve*: pressure at the output of the slide section, showing the characteristic profile of a shock wave (after *Hirschberg* et al. [15.177])

propagating along a trombone tube. The sound intensity of ≈ 175 dB is considerably higher than the intensities illustrated in Fig. 15.99. As predicted, the sharpness of the leading edge of the waveform progressively increased on propagating along the bore. The discontinuity in waveform of the fully developed shockwave dramatically increases the intensities of the higher harmonics of a continuously played note and gives the trumpet and trombone (and trompette organ pipes) their characteristic brassy sound at very high intensities (CD-ROM). The high-frequency components of all such sounds will make them highly directional. *Campbell* [15.160] reviews nonlinear effects in woodwind and brass instruments, with many references to recent research.

To achieve such brassy sounds, the instrument must have a sufficiently long length of relatively narrow pipe, like the trumpet and trombone, in which the pressure fluctuations remain high and have time to build up into a shock wave. Shockwaves are far more difficult to set up in instruments like the horn and cornet, with flaring conical bores, because the pressure drops rather rapidly with increasing diameter along the bore.

Time-Domain Analysis

When nonlinearity is important or when the initial transient response is of interest, it is more appropriate to consider the dynamics in the time- rather than frequency-domain, just as it was for analysing the transient dynamics of the bowed string. Time-domain analysis in wind and brass instruments was pioneered by *Schumacher* [15.158], *McIntyre* et al. [15.178] and *Ayers* [15.179], and is discussed in *Fletcher* and *Rossing* ([15.5], Sect. 8.14). Time-delayed reflectometry measurements are made by producing short pressure pulses inside the mouthpiece generated by a spark or by a sudden piezoelectric displacement of the mouthpiece end-wall. The pressure in the mouthpiece is then recorded as a function of time after the event.

In the linear response regime, measurement in the time domain gives exactly the same information about an instrument as measurements in the frequency domain, assuming both the magnitude and phase of the frequency response is recorded. This follows because the frequency response $Z(\omega)$ measured in the mouthpiece is simply the Fourier transform of the transient pressure response $p(t)$ for a δ- or impulse-function flow (Av) induced by the spark or wall motion. Knowing $p(t)$ or $Z(\omega)$ one can obtain the other by applying the appropriate Fourier transform

$$Z(\omega) = \int_0^\infty p(t)\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t \quad \text{or}$$

$$p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} Z(\omega)\,\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega\,. \qquad (15.138)$$

Measurements of the impulse response are particularly useful in identifying large discontinuities in the acoustic impedance along the bore of instruments produced by tone holes, valves and bends, which can significantly perturb particular partials. The position of any such discontinuity can be determined by the time it takes for the reflected impulse to return to the mouthpiece.

Time-domain analysis is essential, if one wishes to investigate starting transients, where reflections from the end of the instrument are required to stabilise the pitch of a note. This problem is particularly pronounced for horn players pitching, for example, notes as high as the 12th resonant mode of the instrument. The player must buzz the lips a dozen or so cycles into the mouthpiece before the first reflection from the end of the instrument returns to stabilise the pitch. If the player gets the initial buzzing frequency slightly wrong, the instrument may lock on to the 11th or 13th harmonic rather than the intended 12th,

Amazon
EX1013 Page 641

leading to the familiar *cracked note* of the beginner and sometimes even professional horn players. Furthermore, *false* reflections from discontinuities along the length of the tube, may well confuse the initial feedback, making it more difficult to pitch particular notes. This leads to small pitch-dependent changes in the playing characteristics of instruments made to different designs adopted by different manufacturers.

15.3.5 Air-Jet Excitation

Many ancient and modern musical instruments are excited by blowing a jet of air across a hole in a hollow tube or some other acoustic resonator. Familiar examples include the flute, pan pipes, the ocarina and simple whistle in addition to many organ pipes. Sound excitation in flutes and organ pipes was first considered by *Helmholtz* [15.127] in terms of an interaction between the air jet produced by the lips or a flue channel in the mouthpiece of an instrument and the coupled air column resonator.

In practice, the dynamics of sound production is a very complex aerodynamic flow problem requiring the solution of the Navier–Stokes equations governing fluid flow in often complex geometries. Various simplified solutions have been considered by many authors since the time of *Helmholtz* and *Rayleigh* [15.3]. *Fletcher* and *Rossing* ([15.5], Sect. 16.1) provide references to both historic and more-recent research. *Fabre* and *Hirschberg* [15.180] have also written a recent review of simple models for what are sometimes referred to as flue instruments.

Rayleigh showed that the interface separating two fluids moving with different velocities was intrinsically unstable, resulting in an oscillating sinuous lateral disturbance of the interface that grows exponentially with time Fig. 15.107. This arises because, in the frame of reference in which the two fluids move with the same speed in opposite directions, any disturbance of the interface towards one of the fluids will increase the local surface velocity on that side. This will result in a decrease in Bernoulli pressure on that side of the interface and increase it on the other, creating a net force in the same sense as the disturbance, which will therefore grow exponentially with time. For a layer of air moving at velocity V without friction over a stationary layer, a sinusoidally perturbed deflection of the jet in the laboratory frame of reference at rest increases exponentially with distance as it travel along the interface with velocity $V/2$.

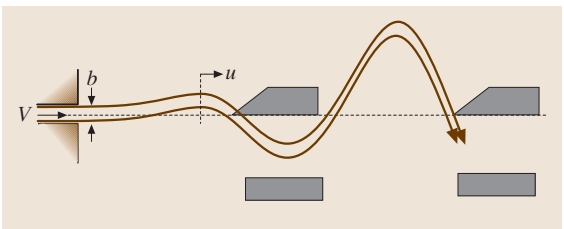


Fig. 15.107 Propagating sinuous instability of an air jet emerging from a flow channel with two possible positions of the angled labium or lip to excite resonances of an attached air column for a given jet velocity V

Similar arguments were used by Rayleigh to describe the instability of an air jet of finite width b and velocity V produced by blowing through an arrow constriction between the pouted lips when playing the flute, or by blowing air through an air channel towards the sharp lip of the recorder or an organ pipe. *Fletcher* [15.181] showed that the lateral displacement $h(x)$ of the jet induced by an acoustic velocity field $v\,\mathrm{e}^{\mathrm{i}\omega t}$ between the jet orifice and the lip varies with position and time as

$$h(x) = -j\left(\frac{v}{\omega}\right)\left[1-\cosh\mu x\,\exp(-\mathrm{i}\omega x/u)\right]\mathrm{e}^{\mathrm{i}\omega t}\,, \tag{15.139}$$

illustrated schematically in Fig. 15.107. The first term simply corresponds to the jet moving with the impressed acoustic field, while the second describes the induced travelling-wave instability moving along the jet with velocity $u \approx V/2$, which dominates the jet displacement at the lip of the instrument. For long-wavelength instabilities on a narrow jet, such that the characteristic wavevector $k \ll 1/b$, Rayleigh showed that the phase velocity $u = \omega/k \sim kbV/2 = (\omega bV/2)^{1/2}$, while the exponential growth factor $\mu = (k/b)^{1/2}$.

In practice, the velocity profile of the jet is never exactly rectangular but in general will be bell-shaped. This results in a slightly different frequency, width b and velocity dependence of the phase velocity, with

$$u = 0.55(\omega b)^{1/3}V^{2/3} \tag{15.140}$$

and corresponding changes in the exponential growth factor (*Savic* [15.182]). Typical propagation velocities of disturbances on the jet of flute and organ pipes measured by *Coltman* [15.183] range from 6.7 m/s to 3.7 m/s with velocity ratios u/V from 0.35 to 0.5 for blowing pressures from 1 to 0.15 inches of water.

Resonances of the attached resonator can be excited when the air-jet instability is in phase with oscillations within the resonator, as shown for the two positions indicated in Fig. 15.107. This corresponds to an odd number of half-wavelengths of the propagating instability, so that $\omega l/u = n\pi$, which is equivalent to $f \sim nV/4$,

Amazon
EX1013 Page 642

where n is an odd integer and we have assumed $u \approx V/2$. For a given length between jet orifice and lip, different frequencies can be excited by varying the jet velocity $V = \sqrt{2P_{\text{mouth}}/\rho}$, where P_{mouth} is the pressure in the mouth creating the jet. When coupled to an acoustic resonator with a number of possible resonances, such as an organ pipe or the pipe of a flute or recorder, the interaction between the jet and oscillating resonator causes the frequency to lock on to a particular resonance, with hysteretic jumps between the resonance excited as the blowing pressure is increased and decreased, as illustrated schematically in Fig. 15.108 (*Coltman* [15.183]). This explains why the pitch of an instrument like the recorder or flute doubles when it is blown more strongly. The line marked "edge tones" indicates the frequency of the excited jet mode instability for the same orifice–lip geometry without an attached pipe and the line marked "f = pipe resonance" indicates when the frequency of the coupled jet coincides with the free vibrations of the attached air column. In practice the instrument is played in close vicinity to the pipe resonances.

The above model is strictly only applicable to small perturbations in jet shape ($\ll b$) and to a nonviscous medium. In practice, measurements of jet displacement by Schlieren photography, hot-wire anemometry (*Nolle* [15.185]), smoke trails *Cremer* and *Ising* [15.186] and *Coltman* [15.187] and particle image velocimetry (PIV) (*Raffel* et al. [15.188]) show that within an acoustic cycle the jet moves to either side of the lip of the instrument with large displacement amplitudes comparable with the physical dimensions of the distance between the orifice and lip. In addition, viscous forces lead to a change in profile of the jet as it moves through the liquid from rectangular to bell-shaped (*Ségoufin* et al. [15.189]). Furthermore, the associated shear forces eventually induce vorticity downstream, with individual vortices shearing away from the central axis of the jet on alternate sides, as observed by *Thwaites* and *Fletcher* [15.190, 191].

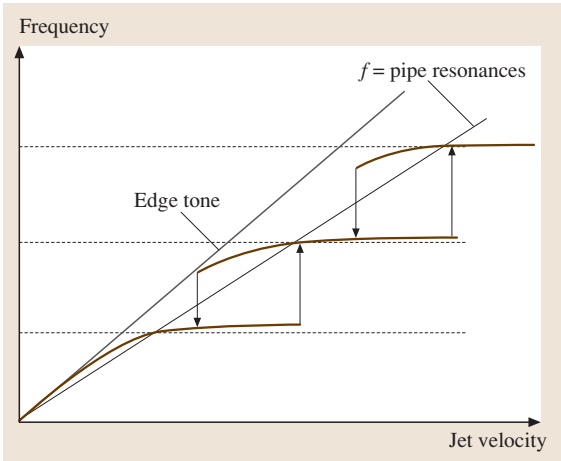


Fig. 15.108 Schematic representation of the frequency of a jet-edge oscillator before and after coupling to a multiresonant acoustic resonator. The *line* marked f = pipe resonator indicates the pressures when the frequencies are the same as the modes of the uncoupled acoustic resonator (after *Coltman* [15.183])

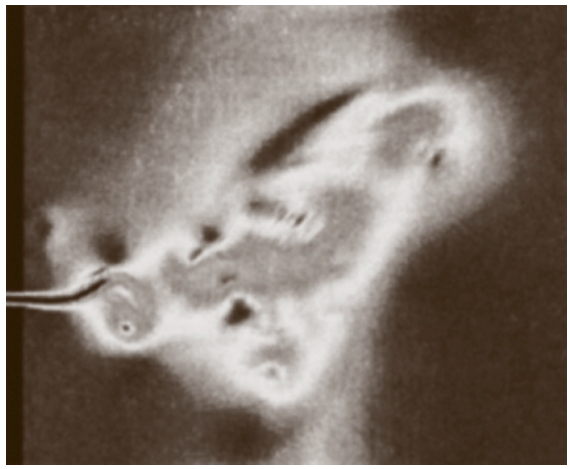


Fig. 15.109 A computed snapshot showing the breakup of a jet and generation of vortices (*Adachi* [15.184])

In recent years, major advances have also been made in studying such problems by computation of solutions of the Navier–Stokes equations describing the nonlinear aerodynamic flow. An example is illustrated by the simulated jet deflections by *Adachi* [15.184] shown in Fig. 15.109. Adachi's computational results are in reasonably good agreement with Nolle's flow measurements [15.185] made with a hot-wire anemometer. A sequence of such computations by Macherey for the jet in a mouthpiece with flute-like geometry is included in a recent review paper on the acoustics of woodwind instruments by *Fabre* and *Hirschberg* [15.192]. The computations show a relatively simple jet structure switching from one side of the lip of a flute to the other during each period of the oscillation.

Air–Jet Resonator Interaction

Despite the obvious limitations of any small-amplitude linear approach to the jet–lip interaction and the excitation of resonator modes, it is instructive to consider the analytic model introduced by *Fletcher* [15.181],

Amazon
EX1013 Page 643

which is discussed in some detail in *Fletcher* and *Rossing* ([15.5], Sect. 16.3), as it includes much of the essential physics in a way that is not always apparent from purely computational models. The assumed geometry is illustrated schematically in Fig. 15.110 and can easily be generalised to model sound excitation in air-jet-driven instruments with different mouthpiece geometries such as the recorder, organ pipe or flute. A uniform jet with a rectangular top-hat velocity profile of width b and velocity V is assumed to impinge on the lip or labium of the instrument, with the jet passing through a fraction A of the attached pipe area S_p. The oscillating travelling-wave jet instability will result in a periodic variation of the fractional area varying as $\alpha e^{i\omega t}$. On entering the pipe channel, the jet will couple to all possible modes of the attached pipe, which in addition to the principle acoustic modes excited, will include many other modes involving radial and azimuthal variations [15.193] that are very strongly attenuated. Much of the energy of the incident jet will therefore be lost by such coupling with typically only a few percent transferred to the important acoustic resonances of the instrument.

To evaluate the transfer of energy to the principle acoustic mode, we consider the pipe impedance across the plane M. In the absence of the jet, the impedances of the attached pipe and mouthpiece section are Z_p and Z_m, where the mouthpiece section has a certain volume and hole area, with the volume in the case of a flute involving an adjustable length of pipe used for fine-tuning. Fletcher simplified an earlier model introduced by *Elder* [15.194], in assuming the principle of linear superposition, with a jet flow superimposed on that of the instrument. The oscillating incident jet then has two principal effects: the oscillating fraction of area α of the jet entering the pipe introduces a flow into the attached pipe

$$U_1 = \alpha S_p Z_m / \left(Z_m + Z_p\right) , \qquad (15.141)$$

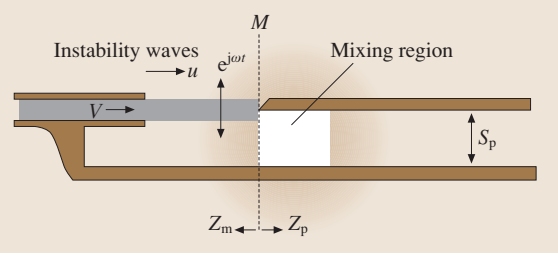


Fig. 15.110 Model to illustrate interaction between air jet and pipe modes (after *Fletcher* [15.193])

while the effective pressure acting on the plane M derived from momentum-balance arguments results in a pipe flow of

$$U_2 = \alpha \rho V^2 / \left(Z_m + Z_p\right) . \qquad (15.142)$$

There is also an additional nonlinear term U_3 mathematically arising from the nonlinear effect of the fractional insertion area of the jet insertion at large amplitudes, which is negligible in comparison with the many other nonlinearities in any realistic model. This model integrates and simplifies earlier models introduced by *Cremer* and *Ising* [15.186], *Coltman* [15.187] and *Elder* [15.194].

It immediately follows from the above arguments that, provided the phase of the jet instabilities are appropriate, instabilities on the jet will excite strong resonances of the instrument when the series impedance $(Z_m + Z_p)$ is a minimum, which is just the impedance of the pipe loaded by the impedance of the mouthpiece assembly. Provided the lengths involved are much less than an acoustic wavelength, $Z_m = i\rho\Delta L / S_p$ where ΔL is the end correction introduced by the mouthpiece assembly. The net flow is then given by

$$U_p = \frac{(V + i\omega\Delta L)\,\rho V \alpha}{S_p \left(Z_m + Z_p\right)} . \qquad (15.143)$$

The resonances are therefore those of a pipe that is open at both ends, but with a small end-correction for the effective volume of the mouthpiece and any additional closed tuning tube and flow in and out of the mouthpiece.

Because the mouthpiece impedance is reactive, the induced vibrations in flow from direct jet flow (U_1) and that induced by the jet pressure (U_2) are in phase quadrature. In addition, the two terms have a different dependence on jet velocity and frequency. In practice, $\omega\Delta L$ is often larger than V, so that the second term usually dominates, though this will not necessarily be true in more realistic models.

For small sinusoidally varying jet perturbations and a top-hat velocity profile, the driving force would only include single frequency components. However, in practice, viscous damping results in a spreading out of the velocity profile in the lateral direction with a bell-shaped profile that increases in width with distance along the jet. If such a profile is offset from the lip, any sinusoidal disturbance of the jet will introduced additional harmonics at frequencies, 2ω, 3ω, etc., with increasing amplitudes for increasing jet oscillations. In practice, the amplitudes of jet oscillation are so large that the jet undergoes a near switching action, alternating its position from one side to the other of the lip. The driving force is therefore

Amazon
EX1013 Page 644

strongly non-sinusoidal and provides a rich spectrum of harmonics to excite the upper partials of any sounded note. A flute player, for example, has considerable control over the quality of the sound produced, by variation of mouth pressure and jet velocity, its velocity profile on leaving the lips, and its direction in relation to the labium or lip of the instrument.

For most musical instruments excited by an air jet, the sinuous instability is the most important, though *Chanaud* [15.195] and *Wilson* et al. [15.196] have highlighted the importance of varicose instabilities (periodic variations in area of the jet), which were also investigated by Rayleigh, as in whistles and whistling, where the air passes through an aperture rather than striking an edge.

Regenerative Feedback

We now consider the effective acoustic impedance of the exciting air column as we did for the vibrating reed. Our emphasis here is to highlight the essential physics rather than provide a rigorous treatment. More details and references are given in *Fletcher* and *Rossing* ([15.5], Sect. 16.4) and *Fletcher* [15.193].

The flow U_m into the mouth of the resonator can be expressed as

$$U_\mathrm{m} = v_\mathrm{m} S_\mathrm{m} = p_\mathrm{m} Y_\mathrm{m} \sim \frac{p_\mathrm{m} S_\mathrm{p}}{\mathrm{i}\omega\rho\Delta L} \,. \qquad (15.144)$$

From (15.139), the lateral displacement of the jet at the lip a distance l from the exit channel of the jet is then given by

$$hl \sim \mathrm{i}\left(\frac{v_\mathrm{m}}{\omega}\right)\cosh\mu l \exp\left(-\omega l/u\right) \,, \qquad (15.145)$$

where u is the phase velocity for disturbances travelling along the jet and the implicit time variation has been omitted. From the ratio of the net flow into the mouthpiece-end and attached resonator to the pressure acting on the air jet, Fletcher and Rossing derive the effective admittance of the air-jet generator,

$$Y_\mathrm{i} \sim \frac{VW}{\rho\omega^2\Delta L}\left(\frac{S_\mathrm{p}}{S_\mathrm{m}}\right)\cosh\mu l \exp\left[-\mathrm{i}\left(\frac{\omega l}{u}\right)\right] \,. \qquad (15.146)$$

Apart from a small phase factor ($\phi \approx V/\omega L$), which we have omitted, the admittance is entirely real and *negative* when $\omega l/u = \pi$, which corresponds to the first half-wavelength of the instability just bridging the distance from channel exit to lip, as shown in Fig. 15.106. This is also true for $\omega l/u = n\pi$, where n is any odd integer, corresponding to any odd number of half-wavelengths of the growing instability between the jet exit and lip of the instrument.

These are just the conditions for positive feedback and the growth of acoustic resonances in the pipe resonances will be excited when $\mathrm{Im}(Y_\mathrm{p} + Y_\mathrm{m}) = 0$, which is equivalent to the condition that Z_s should be a minimum, as expected from (15.129). As *Coltman* [15.183] pointed out, the locus of $\mathrm{Im}(Y_\mathrm{j})$ plotted as a function or $\mathrm{Re}(Y_\mathrm{j})$ as a function of increasing frequency is a spiral about the origin in a clockwise direction (*Fletcher* and *Rossing* [15.5], Fig. 16.10). The jet admittance therefore has a negative real component at all frequencies when the locus point is in the negative half-plane. Resonances can therefore be set up over frequency ranges from $\approx (1/2 \text{ to } 3/2)\omega^*$, $(5/2 \text{ to } 7/2)\omega^*$, etc. where ω^*, $3\omega^*$, $5\omega^*$ are the frequencies when the admittance is purely conductive and negative. By varying the blowing pressure and associated phase velocity of the jet instability, the player can therefore excite instabilities of the jet with the appropriate frequencies to lock on to the resonances of the attached air column, as illustrated schematically in Fig. 15.108.

Measurements by *Thwaites* and *Fletcher* [15.190] are in moderately good agreement with the above model at low blowing pressures and reasonably high frequencies, but deviate somewhat at low frequencies and high blowing pressures. This is scarcely surprising in view of the approximations made in deriving the above result.

Edge Tones and Vortices

Edge tones are set up when jet of air impinges on a lip or thin wire without any coupling to an acoustic resonator. The high flow rates in the vicinity of the lip or wire can generate vortices on the downstream side, which spin off on alternating sides, setting up an alternating force on the lip or wire. If the object is itself part of a mechanical resonating structure, such as a stretched telegraph wire or the strings of an Aeolian harp, wind-blown resonances can be set up, with different resonant modes excited dependent on the strength of the wind. In extreme cases, the excitation of vortices can result in catastrophic build up of mechanical resonances, as in the Tacoma bridge disaster.

Before 1970, many treatments of wind instruments discussed air-jet sound generation in such terms. *Holger* [15.197], for example, proposed a nonlinear theory for edge-tone excitation of sound in wind instruments based on the formation of a vortex sheet, with a succession of vortices already created on alternate sides of the mid-plane of the emerging jet before it hit the lip or labium of the instrument. Indeed, measurements of

Amazon
EX1013 Page 645

the flow instabilities and phase velocity of instabilities in a recorder-like instrument by *Ségoufin* et al. [15.189], as a function of Strouhal number $\omega b/u$, fit the Holger theory rather better than models based on the Rayleigh instability and refined by later authors for both short and long jets, but the experimental errors are rather large. However, the vortex-sheet model does not include the growth of disturbances in the sound field with distance (as measured by *Yoshikawa* [15.198]), which is a crucial parameter for the prediction of the oscillation threshold observed for instruments such as the recorder.

It is also clear that vortex production is important in many wind instruments, especially where the acoustic amplitude is large, as in the vicinity of sharp edges or corners of both open and closed tone holes, as observed in direct measurements of the flow field. For example, *Fabre* et al. [15.199] have recently shown that vortex generation is a significant source of energy dissipation for the fundamental component of a flute note.

In view of the complexity of the fluid dynamics involves, it seems likely that future progress in our understanding of jet-driven wind instruments will largely come from computational simulations, though physical models still provide valuable insight into the basic physics involved.

15.3.6 Woodwind and Brass Instruments

In this last section on wind instruments, we briefly describe a number of woodwind and brass instruments of the modern classical orchestra. All such instruments were developed from much earlier instruments, many of which still exist in folk and ethnic cultures from all around the World. Illustrated guides to a very large number of such instruments are to be found in *Musical Instruments* by *Baines* [15.200] and the encyclopaedia *Musical Instruments of the World* [15.30]. The two text by *Backus* [15.132] and Benade [15.133], both leading researchers in the acoustics of wind instruments, provide many more technical details concerning the construction and acoustics of specific woodwind and brass instruments than space allows here, as does *Fletcher* and *Rossing* ([15.5], Chaps. 13–17) and *Campbell*, *Myers* and *Greated* [15.7].

Woodwind Instruments

The simplest instruments are those based on cylindrical pipes, such as bamboo pan pipes excited by a jet of air blown over one end, or hollow resonators, such as primitive ocarinas, which act as simple Helmholtz resonators with the pitch determined by the openings of the mouthpiece and fingered open holes. Woodwind instruments use approximately cylindrical or conical tubes excited by a reed or a jet of air blown over a hole in the wall of the tube. As we have seen, simple cylindrical and conical tubes retain a harmonic set of resonances independent of their length, which in principle allows a full set of harmonic partials to be sounded when the instrument is artificially shortened by opening the tone holes. In practice, as discussed in the previous section, the harmonicity of the modes is strongly perturbed by a large number of factors including the strongly frequency-dependent end-corrections from tone holes and significant departures from simple cylindrical and conical structures. Such perturbations can, to some extent, be controlled by the skilled instrument maker to preserve the harmonicity of the lowest modes responsible for establishing the playing pitch of an instrument. We will describe the various methods of exciting vibrations by single and double reeds and by air flow in the next section.

Figure 15.111 shows four typical modern orchestral woodwind instruments. All such instruments have distinctive tone qualities and come in various sizes, which cover a wide range of pitches and different tone-colours.

The flute, bass flute and piccolo are based on the resonances of a cylindrical tube, with the open end and mouthpiece hole giving pressure nodes at both ends.



Fig. 15.111 The modern flute, oboe, clarinet and bassoon (not to scale)

Amazon
EX1013 Page 646

Like the recorder, all the chromatic notes of the musical scale can be played by selectively opening and shutting a number of tone holes in the walls of the instrument using the player's fingers (on ancient and baroque flutes) or felted hinged pads operated by a system of keys and levers (on modern instruments). Primitive flutes appear in most ancient cultures.

The clarinet is based on a cylindrical tube excited by a single reed at one end. The reed and mouthpiece close one end of the tube, so that the odd-integer partials are more strongly excited than the even partials, particularly for the lowest notes, when most sideholes are closed. In addition, when overblown, the clarinet sounds a note three times higher (an octave and a fifth). Like all real instruments, perturbations from the tone holes, variations in tube diameter and the nonlinear processes involved in the production of sound vibrations by the reed strongly influence the strength of the excited partials, all of which contribute to the characteristic sound of the instrument.

The single-reed clarinet is a relatively modern instrument developed around 1700 by the German instrument maker Denner. It evolved from the chalumeau, an earlier simple single-reed instrument with a recorder-like body, which still gives its name to the lower register of the clarinet's playing range. In the 1840s, the modern system of keys was introduced based on the Boehm system previously developed for the flute [15.200].

The oboe is based on a conical tube truncated and closed at the playing end by a double reed. As described earlier, a conical tube supports all the integer harmonic partials giving a full and penetrating sound quality that is very rich in upper partials. This is why an oboe note is used to define the playing pitch (usually A4 = 440 Hz) of the modern symphony orchestra. Like all modern instruments, today's oboe developed from much earlier instruments, in this case from the double-reed shawn and bagpipe chanters, which still exist in many ethnic cultures in Europe, Asia and parts of Africa. In addition to the bass oboe, the oboe d'amore and cor anglais, with their bulbous end cavity just before the output bell, have been used for their distinctive plaintive sounds by Bach and many later composers. Like the flute and clarinet, early oboes used mostly open-side holes closed by the fingers, with only one or two holes operated by a key, but developed an increasingly sophisticated key system over time to facilitate the playing of the instrument.

The bassoon is a much larger instrument, producing lower notes of the musical scale. Because of the length of the air column, the spacing of the tone holes would be far too wide to operate by the player's fingers alone. To circumvent this problem, the instrument is folded along its length and relatively long finger holes are cut diagonally through the large wall thickness, so that normally fingered holes can connect to the much wider separation of holes in the resonant air column. The bore of the instrument is based on a largely conical cross section, with the mouthpiece end terminated by a narrow bent tube or *crook* to which a large double reed is attached. Early bassoons included a single key to operate the most distant tone hole on the instrument. Modern instruments have an extended key system to facilitate playing all the notes of the chromatic scale. The contrabassoon includes an additional folded length of tube. Like the oboe, the sound of the bassoon is very rich in upper partials and has a very rich, mellow sound.

Related to the bassoon is the renaissance racket played with a crook and double reed. The instrument looks like a simple cylinder with a set of playing holes cut into its surface. However, in reality, it is a highly convoluted pipe with twelve parallel pipes arranged round the inner diameter of the cylinder and interconnected with short bends at their ends to produce a very long acoustic resonator with easily accessible tone holes. This provides a beautiful example of the centuries-old ingenuity of instrument makers in solving the many acoustic and ergonomic problems involved in the design of musically successful wind instruments.

Brass Instruments

Figure 15.112 illustrates the trumpet, trombone and horn, which like all brass instruments are based on lengths of cylindrical, conical and flared resonant air columns. They are excited by a mouthpiece at one end and a bell at the open end, as described in the previous

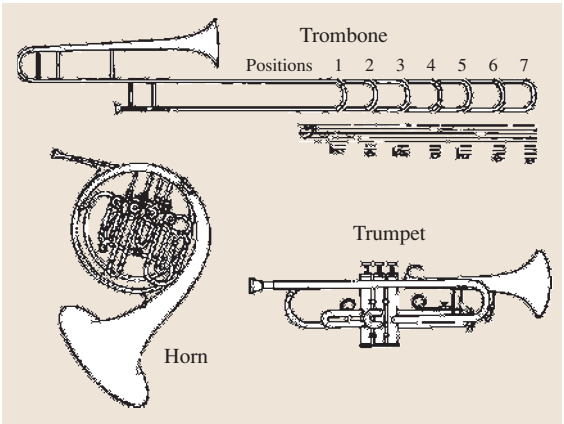


Fig. 15.112 The trombone, French horn and trumpet

Amazon
EX1013 Page 647

section. The player selects the note to be sounded by buzzing the lips, usually at a frequency corresponding to one of the natural resonances of the coupled air column. The essential nonlinearity of this excitation process also excites multiples in frequency of the playing pitch. Ideally, for ease of playing, these harmonics should coincide with the higher modes of the excited air column. As already described, brass instruments are therefore designed to have a full harmonic set of modes. However, because of their shape and outward flare, it is impossible to achieve this for the fundamental mode (Fig. 15.77).

By adjusting the pressure and the tightness of the lips in the mouthpiece, the player can pitch notes based on the $n = 2, 3, 4 \ldots$ modes – the $n = 2$ mode, a fifth above, an octave above, an octave and a fourth above, etc. Trumpet players typically sound notes up to the 8–10 th mode, while skilled horn players can pitch notes up to and sometimes above the 15*th*. In the higher registers, the instruments can therefore play nearly all the notes of a major diatonic scale. A few of the notes can be rather badly out of tune, but a skilled player can usually correct for this by adjusting the lip pressure and flow of air through the mouthpiece. The low notes are based on simple intervals: the perfect fifth, octave, perfect fourth, etc. Trumpets and horns were therefore often used in early classical music to add a sense of military excitement and to emphasise the harmony of the key in which the music is written. However, in later classical music and music of the romantic period, all the notes of the chromatic scale were required. To achieve this, brass instruments such as the trumpet and horn were developed with a set of air valves, which enabled the player to switch in and out various combinations of different lengths of tube, to change the effective resonating length of the vibrating air column and hence playing pitch. Uniquely, the pitch of the trombone is changed by the use of interpenetrating cylindrical sliding tubes, which change the effective length. Modern instruments generally use a folded or coiled tube structure to keep the size of the instrument to manageable proportions.

The trombone can sound all the semitones of the chromatic scale, by the player sliding lengths of closely fitting cylindrical tubing inside each other. In the *first position*, with the shortest length tube (Fig. 15.112), the B-flat tenor trombone sounds the note B-flat at $\approx$115 Hz, corresponding to the $n = 2$ mode, an octave below the lowest note on the B-flat trumpet. To play notes at successive semitones lower, the total length has to be extended sequentially by fractional increases of 1.059. From the shortest to longest lengths there are seven such increasingly spaced *positions*. When fully extended, the trombone then plays a note six semitones lower (E) than the initial note sounded. One can then switch to the $n = 3$ mode to increase the pitch by a perfect fifth, to the note B a semitone higher than the initial note sounded. Using the closer positions enables the next six higher semitones to be played. Higher notes can be excited by suitable combinations of both position and mode excited. The trombone is one of the few musical instruments that can slide continuously over a large range of frequencies, simply by smoothly changing its length. This is widely used in jazz, where it also enables the player to use a very wide, frequency-modulated, vibrato effect and bending of the pitch of a note for expressive effect.

The fully extended length of the vibrating air column in the first position is $\approx$ 2.5 m. Two-thirds of the length is made up of 1.3 cm-diameter cylindrical tubing with the remaining gently flared end-section opening out to a bell diameter of 16–20 cm.

The trumpet achieves the full chromatic range by the use of three piston valves, which enable additional lengths of tubing to be switched in and out of the resonating air column. In the inactive *up* position, the sound travels directly through a hole passing directly through the valve. When the piston is depressed, the valve enables the tube on either side of the valve to be connected to an additional length of tubing, which includes a small, preset, sliding section for fine tuning. The pitch is decreased by a tone on depressing the first piston and a semitone by the second. Pressing them down together therefore lowers the pitch by three semitones (a minor third). Depressing the third valve also lowers the pitch by three semitones, so that when all three valves are depressed the pitch is lowered by six semitones. However, the tuning is not exact, because whenever any single valve is depressed the effective tube length is lengthened. Therefore, when a second (or third) valve is depressed, the fractional increase in effective length is less when the second valve alone is used. This is related to the need to *increase* the spacing of the semitone positions on the trombone as it is extended. Similar mistuning problems arise for all combinations of valves used.

To circumvent these difficulties, compromises have to be made, if the instrument is to play in tune (*Backus* [15.132], pp. 270–271). The added length of tubing to produce the semitone and tone intervals are therefore purposely made slightly too long, giving semitone and tone intervals that are slightly flat, but which in combination produce a three-semitone interval which is slightly sharp. Similarly, the third valve is tuned to give a pitch change of slightly more than three semitones. This allows the full of range of semitones to be played

Amazon
EX1013 Page 648

with only slight fluctuations above and below the correct tuning. The error is greatest when all three pistons are depressed. In the modern trumpet, the mistuning can be compensated by a small length of tubing operated by an additional small valve operated by the little finger of the playing hand. To regulate the overall tuning of the instrument, the playing length of the instrument can be varied by a sliding U-tube section at the first bend away from the mouthpiece. As we will show later, the skilled player can adjust the muscles controlling the dynamics of the lip-reed excitation to correct for any slight mistuning inherent in the design of the instrument.

The bends and valves incorporated into the structure of brass instruments will clearly result in sudden changes in acoustic impedance of the resonating air column, which will produce reflections and perturbations in the frequency of the frequency of the excited modes. Surprising, as shown earlier, such effects are acoustically relatively unimportant, though for the player they may affect the feel of the instrument and ease with which it can be played. In particular, when a brass player is pitching one of the higher modes of an instrument such as the horn several oscillations have to be produced before any feedback returns from the end of the instrument to stabilise the playing frequency. For example, when pitching the 12th mode on a horn, the player has to buzz the lips for about 12 cycles before the first reflection returns from the end of the instrument, which demonstrated the difficulty of exact pitching of notes in the higher registers. Note that, because of the dispersion of sound waves in a flared tube, the group velocity determining the transit time will not be the same as the phase velocity determining modal frequencies. Any strong reflections from sharp bends and discontinuities in acoustic impedance introduced by the valve structures can potentially confuse the pitching of notes and the playability of an instrument. Such transients can be investigated directly by time-domain acoustic reflectometry (see, for example, *Ayers* [15.179]). Different manufacturers choose different methods to deal with the various tuning and other acoustic problems involved in the design of brass instruments and players develop strong preferences for a particular type of instrument based on both the sound they produce and their ease of playing.

The trumpet bore is ≈ 137 cm long with largely cylindrical tubing with a diameter of ≈ 1.1, which tapers down to ≈ 0.9 cm at the mouthpiece end over a distance of ≈ 12–24 cm. It opens up over about the last third of its length with an end bell of diameter ≈ 11 cm. To reduce the overall length, its length is coiled with a single complete turn, as illustrated in Fig. 15.112.

The cornet is closely related to the trumpet but has a largely conical rather than cylindrical bore and is further shortened by having two coiled sections. This results in a somewhat lower cut-off frequency giving a slightly warmer but less brilliant sound quality. The bugle is an even simpler double-coiled valveless instrument of fixed length, so that it can only sound the notes of the natural harmonics. It was widely used by the military to send simple messages to armies and is still used today in sounding the *last post*, accompanying the burial of the military dead.

The modern French horn developed from long straight pipes with flared ends played with a mouthpiece. Technology then enabled horns to be made with coiled tubes, like hunting horns, greatly reducing their size. In early classical music, horns were only expected to play a few simple intervals in the key in which the music was written. For music written in different keys, the player had to add an additional section of coiled tube called a *crook*, to extend the length of the instrument accordingly.

To extend the range of notes that a horn could play, the player can place his hand into the end of the instrument. Depending how the hand is inserted, the pitch of individual harmonics can be lowered or raised by around a semitone, producing what is referred to as

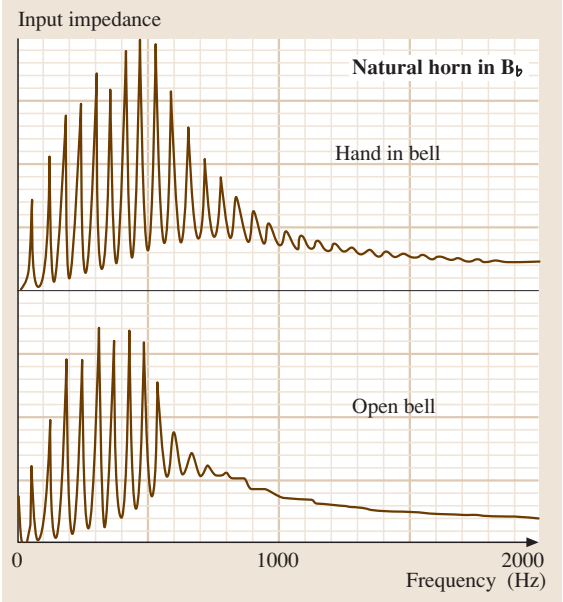


Fig. 15.113 Input impedance measurements of a natural horn, with and without a hand in the bell (after *Benade* [15.133])

Amazon
EX1013 Page 649

a *hand-stopped* note. If the hand is partially inserted into the bell, it dramatically increases the cut-off frequency giving the player access to a much larger number of higher modes, as illustrated in measurements by *Benade* ([15.133], Fig. 20.17) reproduced in Fig. 15.113. Although the associated changes in pitch may be relatively small, inserting the hand into the end of a horn significantly changes the spectrum of the radiated sound, which allows the horn player some additional freedom in the quality of the sound produced.

The modern orchestral horn produces all the notes of the chromatic scale using rotary valves to switch in combinations of different length tube, rather like the trumpet. The modern instrument is a combination of a horn in F and a horn in B-flat, which can be interchanged by a rotary valve operated by the thumb of the left hand, while the first three fingers operate the three main valves, which are common to both horns.

The total length of the F horn is about 375 cm, a third longer than the B-flat trombone, enabling it to play down to the note F2. The F-horn is used for the lowest notes on the instrument, while the B-flat horn is used to give the player a much higher degree of security in pitching the higher notes. Like the primitive hunting horn, the modern horn is coiled to accommodate its great length and has a bore that opens up gently over its whole length from a diameter of ≈ 0.9 mm at the mouthpiece end to a rapidly flared output bell with a diameter of ≈ 30 cm.

There are many other instruments played with a mouthpiece in both ancient and ethnic cultures around the world. Many primitive instruments are simply hollowed-out tubes of wood, bamboo or animal bones. The notes that such instruments can produce are limited to the principal, quasi-harmonic, resonances of the instrument. There are also many hybrid instruments played with a mouthpiece, which use finger-stopped holes along their length, just like woodwind instruments. Important examples are the renaissance *cornett* and the now almost obsolete *serpent*, a spectacularly large, multiple bend, s-shaped, instrument. Many modern players of baroque-period *natural* trumpets have also added finger holes to the sides of their instruments, to facilitate the pitching of the highest notes.

15.4 Percussion Instruments

Compared with the extensive literature on stringed, woodwind and brass instruments, the number of publications on percussion instruments is somewhat smaller. This is largely because the acoustics of percussion instruments is almost entirely dominated by the well-understood, free vibrations of relatively simple structures, without complications from the highly nonlinear excitation processes involved in exciting string and wind instruments. However, as this section will highlight, the physics of percussion instruments involves a number of fascinating and often unexpected features of considerable acoustic interest.

The two most important references on the acoustics of percussion instruments are *The Physics of Musical Instruments*, *Fletcher* and *Rossing* ([15.5], Chaps. 18–21) and the *Science of Percussion Instruments* [15.201] by *Rossing*, the general editor of this Handbook, who has pioneered research on a very wide range of classical and ethnic percussion instruments. *James Blade*'s *Percussion Instruments and their History*[15.202] provides an authoritative survey of the development of percussion instruments from their primitive origins to their modern use.

Percussion instruments are amongst the oldest and most numerous of all musical instruments. Primitive instruments played by hitting sticks against each other or against hollowed-out tree stumps or gourds are likely to have evolved very soon after man discovered tools to make weapons and other simple artefacts essential for survival. The rhythmic beating of drums by Japanese Kodo drummers, the marching bands of soldiers down the centuries, and the massed percussion section of a classical orchestra still instil the same primeval feelings of power and excitement used to frighten away the beasts of the forest and to raise the fighting spirits of early groups of hunters. The beating of drums would also have provided a simple way of communicating messages over large distances, the rhythmic patterns providing the foundation of the earliest musical language – the organisation of sound to convey information or emotion. Martial music, relying heavily on the beating of drums continues to be used, and as often misused, to instil a sense of belonging to a particular group or nation and to instil fear in the enemy.

In China, bells made of clay and copper were already in use well before 2000 BC. The discovery of bronze quickly led to the development of some of the most sophisticated and remarkable sets of cast bells ever made, reaching its peak in the western Zhou (1122–771 BC) and eastern Zhou (770–249 BC) dynasties (*Fletcher*

Amazon
EX1013 Page 650

and *Rossing* ([15.5], Sect. 21.15). Inscriptions on the set of 65 tuned chime bells in the tomb of Zeng Hou Yi (433 BC), show that the Chinese had already established a 12-note scale, closely related to our present, but much later, western scale. The ceremonial use of bells and gongs is widespread in religious cultures all over the world and in western countries has the traditional use of summoning worshippers to church and accompanying the dead to their graves.

In the 18th century classical symphony orchestra of Haydn and Mozart's time, the timpani helped emphasise the beat and pitch of the music being played, particularly in loud sections, with the occasional use of cymbals and triangle to emphasise exotic and often Turkish influences. The percussion section of today's symphony orchestra may well be required to play up to 100 different instruments for a single piece, as in *Notations I–IV* by *Boulez* [15.203]. This typical modern score includes timpani, gongs, bells, metals and glass chimes, claves, wooden blocks, cowbells, tom-toms, marimbas, glockenspiels, xylophones, vibraphones, sizzle and suspended cymbals, tablas, timbales, metal blocks, log drums, boobams, bell plates in C and B flat, side drums, Chinese cymbals, triangles, maracas, a bell tree, etc. Modern composers can include almost anything that makes a noise – everything from typewriters to vacuum cleaners. The percussion section of the orchestra is required to play them all – often simultaneously.

We will necessarily have to be selective in the instruments that we choose to consider and will concentrate largely on the more familiar instruments of the modern classical symphony orchestra. We will also constrict our attention to instruments that are struck and will ignore instruments like whistles, rattles, scrapers, whips and other similar instruments that percussion players are also often required to play.

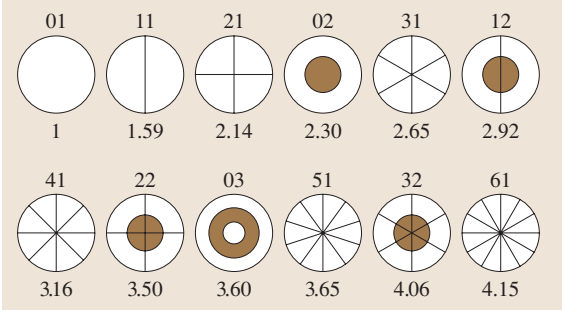


Fig. 15.114 The first 12 modes of a circular membrane illustrating the mode nomenclature, nodal lines and frequencies relative to the fundamental 01 mode

15.4.1 Membranes

Circular Membrane

A uniform membrane with areal density σ, stretched with uniform tension T over a rigid circular supporting frame, supports acoustically important transverse displacements z perpendicular to the surface described by the wave equation

$$T\left(\frac{\partial^2 z}{\partial r^2}+\frac{1}{r}\frac{\partial z}{\partial r}+\frac{1}{r^2}\frac{\partial^2 z}{\partial \phi^2}\right)=\sigma\frac{\partial^2 z}{\partial t^2}\,, \qquad (15.147)$$

which has solutions of the form

$$z\,(r,\phi,t)=J_m(k_{mn}r)\left\{\begin{array}{l}A\cos m\phi\\ +B\sin m\phi\end{array}\right\}\mathrm{e}^{\mathrm{i}\omega t}\,. \qquad (15.148)$$

$J_m(k_{mn}r)$ are Bessel functions of order m, where n denotes the number of radial nodes and m the number of nodal diameters. The eigenfrequencies $\omega_{mn}=k_{mn}\sqrt{T/\sigma}$ are determined by the requirement that $J_m(k_{mn}a)=0$ on the boundary at $r=a$. The frequency of the fundamental (01) mode is $(2.405/2\pi a)\sqrt{T/\sigma}$, where $J_0(k_{01}a)=0$. The relative frequencies of the first 12 modes and associated nodal lines and circles are shown in Fig. 15.114.

For ideal circular symmetry, the independent azimuthal cosine and sine solutions result in degenerate modes having the same resonant frequencies. The degeneracy of such modes can be lifted by a nonuniform tension, variations in thickness when calfskin is used, and a lack of ideal circularity of the supporting rim. Any resulting splitting of the frequencies of such modes can result in beats between the sound of given partials, which the player can eliminate by selectively adjusting the tension around the perimeter of the membrane or by hitting the membrane at a nodal position of one of the contributing modes.

Unlike transverse waves on a stretched string, the modes of a circular membrane are inharmonic. As a consequence, the waveforms formed by the combination of such partials excited when the drumhead is struck are non-repetitive. Audio CD-ROM illustrates the frequencies of the first 12 modes played in sequence. CD-ROM illustrates their sound when played together as a damped chord, which already produces the realistic sound of a typical drum, having a reasonably well-defined sense of pitch, despite the inharmonicity of the modes.

Although percussion instruments may often lack a particularly well-defined sense of pitch, one can nev-

Amazon
EX1013 Page 651

ertheless describe the overall pitch as being high or low. For example, a side drum has a higher overall pitch than a bass drum and a triangle higher than a large gong. From an acoustical point of view, we will be particularly interested in the lower-frequency quasi-harmonic modes. However, one must never forget the importance of the higher-frequency inharmonic modes in defining the initial transient, which is very important in characterising the sound of an instrument.

Nonlinear effects arise in drums in much the same way as in strings (Sect. 15.2.2). The increase in tension, proportional to the square of the vibrational amplitude, leads to an increase in modal frequencies. In addition, nonlinearity can result in mode conversion (Sect. 15.2.2) and the transfer of energy from initially excited lower-frequency modes with large amplitudes to higher partials. Although the pitch of a drum is raised when strongly hit, this may to some extent be compensated by the psychoacoustic effect of a low-frequency note sounding flatter as its intensity increased. Changes in perceived pitch of a drum with time can often be emphasised by the use of digital processing, to increase the frequency of the recorded playback without changing the overall envelope in time (audio CD-ROM).

Air Loading and Radiation

The above description of the vibrational states of a membrane neglects the induced motion of the air on either side of the drum skin. At low frequencies, this adds a mass $\approx \frac{8}{3}\rho a^3$ to the membrane (*Fletcher* and *Rossing* [15.5], Sect. 18.1.2), approximating to a cylinder of air with the same thickness as the radius a of the drum head. The added mass lowers the vibrational frequencies relative to those of an ideal membrane vibrating in vacuum. The effect is largest at low frequencies, when the wavelength in air is larger than or comparable with the size of the drumhead. For higher-frequency modes, with a number of wavelengths λ across the width of the drumhead, the induced air motion only extends a distance $\approx \lambda 2\pi (\ll a)$ from the membrane. Air loading therefore leaves the higher-frequency modes relatively unperturbed.

Drums can have a single drum skin stretched over a hollow body, such as the kettle drum of the timpani, or two drum heads on either side of a supporting cylinder or hollowed out block of wood, like the side drum and southern Indian mrdanga (*Fletcher* and *Rossing* [15.5], Sect. 18.5). By stretching the drum head over a hollow body, the sound radiated from the back surface is eliminated, just like mounting a loudspeaker cone in an enclosure. At low frequencies, the drum then acts as a monopole source with isotropic and significantly enhanced radiation efficiency. This is illustrated by the much reduced 60 dB decay time of the (11) dipole mode of a stretched timpani skin, when the drum skin was attached to the kettle –

Table 15.7 Ideal and measured frequencies of the modal frequencies of a timpani drum head normalised to the acoustically important (11) mode before and after mounting on the kettle, and the internal air resonances of the kettle. The *arrows* indicate the sense of the most significant perturbations of drum head frequencies and the *asterisks* indicate the resulting quasi-harmonic set of acoustically important modes (adapted from Fletcher and Rossing)

Mode	Ideal membrane	Drumhead in air		Coupled internal air resonances		Drumhead on kettle	
01	0.63	82 Hz	0.53	(0,1,0) (0,1,1)	385 Hz	127 Hz	0.85 ↑
11	1.00	160	1.0	(1,1,0) (1,1,1)	337 Hz 566 Hz	150	1.00 ↓ ***
21	1.34	229	1.48	(2,1,0) (2,1,1)	537 Hz 747 Hz	227	1.51 ↓ ***
02	1.44	241	1.55	(0,1,0)	(0,2,0)	252	1.68 ↑
31	1.66	297	1.92	(3,1,0)	(3,1,1)	298	1.99 ***
12	1.83	323	2.08	(1,2,0)	(1,2,1)	314	2.09 ↓
41	1.98	366	2.36			366	2.44 ***
22	2.20	402	2.59			401	2.67
03	2.26	407	2.63	(0,1,0)		418	2.79 ↑
51	2.29	431	2.78			434	2.89
32	2.55	479	3.09			448	2.99 ***
61	2.61	484	3.12			462	3.08

Amazon
EX1013 Page 652

from 2.5 s to 0.5 s (*Fletcher* and *Rossing* [15.5], Table 18.4).

In addition, any net change in the volume of the enclosed air resulting from vibrations of the drum skin will increase the restoring forces acting on it and hence raise the modal frequencies, see Table 15.7. In this example, the coupling raises the frequency of the (01) drumhead mode from 82 to 127 Hz, the (02) mode from 241 to 252 Hz, and the (03) mode from 407 to 418 Hz. In contrast, the asymmetric, volume-conserving, (11) mode is lowered in frequency from 160 to 150 Hz, which probably results from coupling to the higher frequency 337 Hz internal air mode having the same aerial symmetry.

As in any coupled system in the absence of significant damping, the separation in frequency of the normal modes resulting from coupling will tend to be greater than the separation of the uncoupled resonators (Fig. 15.46b). In addition, any enclosed air will provide a coupling between the opposing heads of a double-sided drum. For example, the 227 and 299 Hz (01) uncoupled (01) modes of the opposing heads of a snare drum become the 182 and 330 Hz normal modes of the double-sided drum (*Fletcher* and *Rossing* [15.5], Fig. 18.7).

Excitation by Player

The quality of the sound produced by any percussion instrument depends as much on the player's skills as on the inherent qualities of the instrument itself. Drums are not simply hit. A player uses considerable manual dexterity in controlling the way the drumstick strikes and is allowed to bounce away from the stretched drum skin, xylophone bar or tubular bell. It is important that contact of the stick with the instrument is kept to a minimum; otherwise the stick will strongly inhibit the vibrational modes that the player intends to excite. The *lift off* is just as important as *the strike*, and it takes years of practice to perfect, for example, a continuous side drum or triangle roll.

It is also important to strike an instrument in the right place using the right kind of stick or beater to produce the required tone, resonance or loudness required for a specific musical performance. As discussed earlier in relation to the excitation of modes on a stretched string and modal analysis measurements, one can selectively emphasise or eliminate particular partials by striking an instrument at antinodes or nodes of particular modes. A skilled timpani player can therefore produce a large number of different sounds by hitting the drumhead at different positions from the rim. Striking timpani close to their outer rim preferentially excites the higher-frequency modes, while striking close to the centre results in a rather dull *thud*. This is due to the preferential excitation of the inefficient (00) mode and elimination of the acoustically important (0*n*) modes. The most sonorous sounding notes are generally struck about a quarter of the way in from the edge of the drum.

The sound is also strongly affected by the dynamics of the beater–drumhead interaction, which is rather soft and *soggy* near the centre and much harder and *springy* near the outer rim. The sound is also strongly affected by the type of drumstick used. Light, hard wooden sticks will make a more immediate impact with a bar or drum skin than heavily felted beaters. Such difference are similar to the effect of using heavy or light force hammers to preferentially excite lower- or higher-frequency modes in modal analysis investigations. A professional timpanist or percussion player will use completely different sticks for different styles of music and obtain effects on the same instrument ranging from the muffled drum beats of the *Death March* from Handel's *Saul* to the sound of cannons firing in Tchaikovsky's 1812 overture.

There has been much less research on the mechanics of the drumstick–skin interaction than on the excitation of piano strings by the piano hammer (Sect. 15.2.2), though much of the physics involved is very similar. In particular, the shortest time that the drumstick can remain in contact with the skin will be determined by the time taken for the first reflected wave to return from the rim of the drum to bounce it off. This is illustrated schematically in Fig. 15.115, which illustrates qualitatively the force applied to the drumhead as a function of time for a hard and a soft drumstick struck near the centre (*solid line*) and then played nearer the edge (*dashed*). The overall spectrum of sound of the drum will be controlled by the frequency content of such impulses. Short contact times will emphasise the higher partials and give

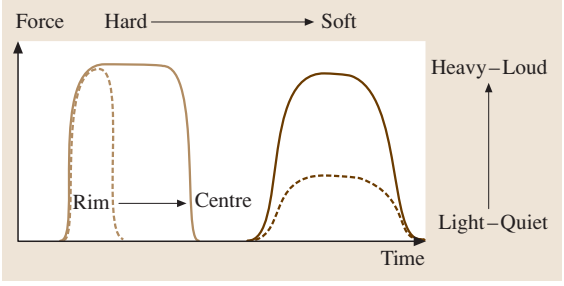


Fig. 15.115 Schematic impulses from a striking drumstick, illustrating the effect of exciting the drum head at different positions, with different strengths, and with a hard and soft drumstick

Amazon
EX1013 Page 653

rise to a more percussive and brighter sound. Higher partials will also be emphasised by the use of metal beaters or drumsticks with hard-wooden striking heads rather than leather or soft felt-covered stick heads. This is illustrated by the second pulse, which would produce a softer, mellower sound, without such a strong initial attack. Clearly, the loudness of the drum note will be proportional to the mass m of the striking drumstick head and its impact velocity v, delivering an impulse of $\approx mv$.

Audio CD-ROM illustrates the change in sound of a timpani note, as the player progressively strikes the drum with a hard felt stick, starting from the outside edge and moving towards the centre, in equal intervals of ≈ one eighth of the radius. Audio CD-ROM illustrates the sound of a timpani when struck at one quarter of the radius from the edge, using a succession of drumsticks of increasing hardness, from a large softly felted beater to a wooden beater.

In modern performances of baroque and early classical music, the timpanist will use relatively light sticks, with leather-coated striking heads, while for music of the romantic period larger and softer felt-covered drumsticks will often be used.

Many drums of ethnic origin are played with the hands, hitting the drum head with fingers, clenched fists or open palms to create quite different kinds of sounds. In some cases, the player can also press down on the drum head to increase the tension and hence change pitch of the note. For a double-headed drum, the coupling of the air between the drum heads can even enable the player to change the pitch and sound of a given note by applying pressure to the drum head not being struck.

We now consider a number of well-known percussion instruments based on stretched membranes, which illustrate the above principles. These will include drums with a well defined pitch, such as kettle drums (timpani) and the Indian tabla and mrdanga, and drums with no defined pitch, such as the side and bass drum.

Kettle Drums (Timpani)

The kettle drum or timpani traditionally used a specially prepared calfskin stretched over a hollow, approximately hemispherical, copper kettle generally beaten out of copper sheet. Nowadays, thin (0.19 mm) mylar sheet is often used in preference to calfskin for the drum skin, because of its uniformity and reduced susceptibility to changes in tension from variations in temperature and humidity. The drum skin is stretched over a supporting ring attached to the kettle, with the tension of the skin typically adjusted using 6–8 tuning screws equally spaced around the circumference. The player adjusts these screws to tune the instrument and to optimise the quality of tone produced. In modern instruments, a mechanical pedal arrangement can be used to quickly change the tension and thereby the tuning, by pushing the supporting ring up against the drumhead. Typically, such an arrangement can increase the tension by up to a factor of two, raising the pitch by a perfect fifth. In the modern classical symphony orchestra, the timpanist will use two or three timpani of different sizes to cover the range of pitched notes required.

Figure 15.116 shows the waveform and spectrum of the immediate and delayed sound of a timpani note (the first drum note in audio CD-ROM). The initial sound includes contributions from all the modes excited. This includes not only the vibrational modes of the drum head, but also the air inside the kettle, the kettle itself and even the supporting legs and vibrations induced in the floor. Many of these vibrations die away rather quickly, leaving a number of prominent, slowly decaying, drumskin modes. Note in particular, the strongly damped (01) mode at ≈ 140 Hz and the less strongly damped modes (02) and (03) modes at 210 Hz and 284 Hz, tuned ap-

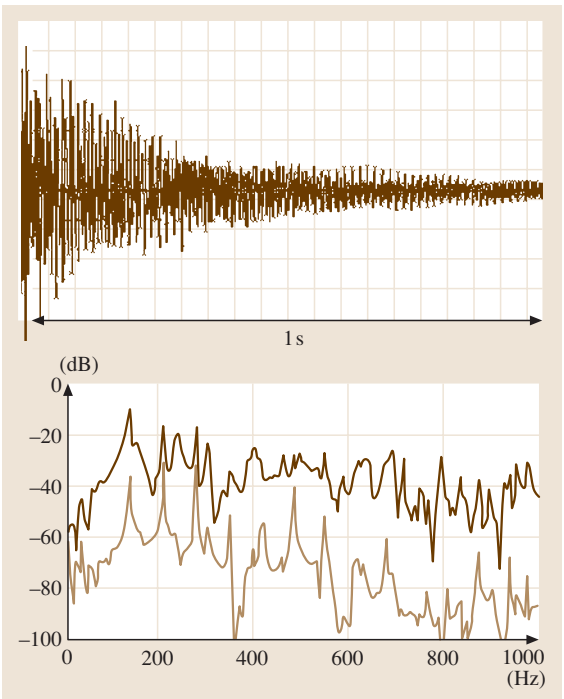


Fig. 15.116 Decaying waveform of a timpani note and FFT spectra at the start of a note (upper trace) and after 1 s (lower trace)

Amazon
EX1013 Page 654

proximately to a perfect fifth and an octave above the fundamental. As noted by Rayleigh in relation to church bells ([15.3] Vol. 1, Sect. 394), the pitch of a note is often determined by the higher quasi-harmonically related partials rather than the lowest partial present. This is demonstrated by the second drum beat in CD-ROM, which has all frequency components below 250 Hz removed. The perceived pitch at long times is unchanged, though there is a considerable loss in *depth* or *body* of the resulting sound.

Modal frequencies for a typical kettle drum have already been listed in Table 15.7, which includes a set of nearly harmonic modes indicated by asterisks achieved, in part, by empirical design of the coupled membrane and kettle air vibrations. To a first approximation, the modal frequencies are determined by the volume of the kettle rather than its shape. The smaller the enclosed volume, the larger its effect on the lowest-order drumhead modes. Nevertheless, there are distinct differences in the sounds of timpani used by orchestras in Vienna and those used elsewhere in Europe (*Bertsch* [15.204]). Such differences can be attributed to the Viennese preference for calfskin rather than mylar drum heads, a small shape dependence affecting the coupling to the internal air resonances, and a different tuning mechanism. The modal frequencies of the Viennese timpani measured by Bertsch were similar to those listed in Table 15.7, with the (11), (21), (31) and (41) modes again forming a quasi-harmonic set of partials, in the approximate ratios 1:1.5:2.0:2.4:2.9. Rather surprisingly, the relative frequencies of the lower two modes could be interchanged with tuning.

Indian Tabla and Mrdanga

Another way of achieving a near harmonic set of resonances of a vibrating drumhead is to add mass to the drum head and hence change the frequencies of its normal modes of vibration. For the single- and double-headed Indian tabla and mrdanga drums, this is achieved by selectively loading the drum skin with several coatings of a paste of starch, gum, iron oxide and other materials – see *Fletcher* and *Rossing* ([15.5], Sect. 18.5). The acoustics of the tabla was first investigated by *Raman* [15.205], who obtained Chladni patterns for many of the lower-frequency modes of the drum head. *Rossing* and *Sykes* [15.206] measured the incremental changes in frequency of the loaded membrane as each additional layer was added. A 100 layers lowered the fundamental mode by about an octave. The resulting five lowest modes were harmonically related and including several degenerate modes derived from the smoothly transformed modes of the original unloaded membrane. The results were very similar to those obtained earlier by Raman. Investigations by *Ramakrishna* and *Sondhi* [15.207] and by *De* [15.208] showed that, to achieve a quasi-harmonic set of low-frequency modes, the areal density at the centre of such drums should be approximately 10 times that of the unloaded sections.

Figure 15.117 illustrates the decaying waveform and spectra of a well-tuned tabla drum (audio CD-ROM) from 200 ms FFTs of the initial sound and after 0.5 s. The spectra show three prominent partials at 549, 826 and 1107 Hz, in the near-harmonic ratios 1:1.51:2.02, which results in a well-defined sense of pitch. In contrast to the timpani, these partials dominate the sound and determine the pitch from the very beginning of the note. Note too the very wide spectrum of the rapidly decaying initial transient.

Side and Snare Drum

The side or snare drum is the classical two-headed drum of the modern symphony orchestra. It is usually played with either very short percussive beats or as a roll, with rapidly alternating notes from two alternating drumsticks. This results in a quasi-continuous

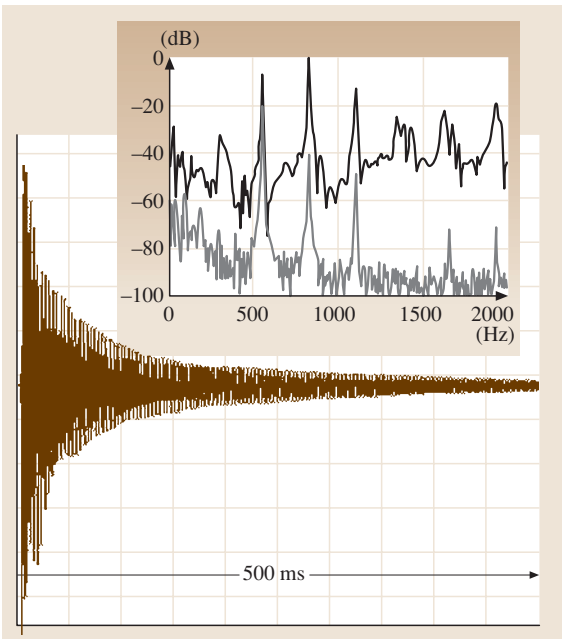


Fig. 15.117 Decaying waveform of a tabla drum with initial FFT spectrum (upper trace) and after 0.5 s (lower) illustrating the weakly damped, near-harmonic, resonances of the loaded drumhead

Amazon
EX1013 Page 655

noise source, which can be played over a very wide range of intensities, from very soft to very loud, to support, for example, a Rossini crescendo. Because the side drum is designed to produce short percussive sounds or a wide-band source of noise, little effort is made to tune the partial frequencies of the two drum heads.

Like the timpani and Indian drums, the vibrational modes of the drumheads can be strongly perturbed in frequency by the air coupling. When used as a snare drum, the induced vibrations of the non-striking head can be sufficient for it to rattle against a number of metal cables tightly stretched a few mm above the surface of the non-striking head. The resulting interruption of the vibrations, on impact with the snares, leads to the generation of additional high-frequency noise and the *sizzle* effect of the sound excited. A not dissimilar effect is used on the Indian tambura, an Indian stringed instrument investigated by *Raman* [15.210], which has a bridge purposely designed to cause the strings to rattle *Fletcher* and *Rossing* ([15.5], Fig. 9.30).

Figure 15.118 shows the waveform and time-averaged FFT of a side-drum roll (audio CD-ROM). The spectrum is lacking in spectral features other than a modest peak in noise at around 100–200 Hz, associated with the vibration of the lower head against the snares.

Although the exact placing of the vibrational modes of the strike and snare heads are of little acoustic importance, their coupling via the enclosed air illustrates the general properties of double-headed drums of all types. The first four coupled normal modes are shown in Fig. 15.119, which is based on data from *Rossing* ([15.211] Sect. 4.4). For a freely supported drum, momentum has to be conserved, so that normal modes with the two heads vibrating in the same phase will also involve motion of the supporting shell of the drum, as indicated by the arrows in Fig. 15.119.

As anticipated, the air coupling increases the separation of the (01) modes from 227 and 299 Hz to 182 and 330 Hz, and the (11) modes from 284 and 331 Hz to 278 and 341 Hz. The perturbations in modal frequencies will always be largest when the coupled modes have similar frequencies. Such perturbations becomes progressively weaker at higher frequencies, partly because the coupling from the enclosed air modes becomes weaker and partly because the frequencies of the two drum-head modes having the same symmetry become more widely separated. The higher modal frequencies are therefore little different from those of the individual drum heads in isolation.

Figure 15.119 also illustrates the anticipated polar radiation patterns for the normal modes measured by *Zhao* [15.209] and reproduced in *Rossing* ([15.201] Figs. 4.7 and 4.8). The coupled (10) normal modes act as a monopole radiation source, when the heads move in opposite directions, and a dipole source, when vibrating in anti-phase. In contrast, the (11) modes with heads vibrating in phase act as a quadrupole source, and a dipole source, when vibrating in anti-phase.

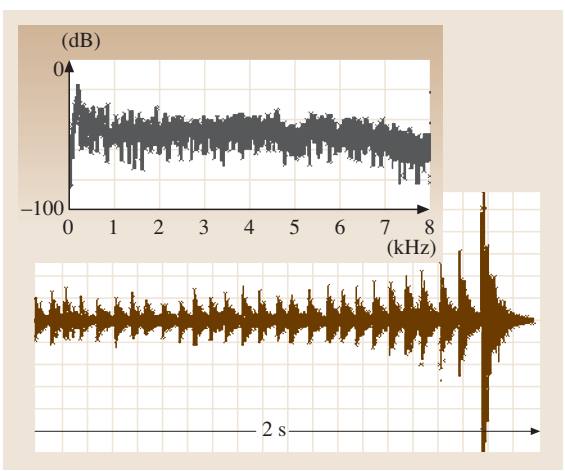


Fig. 15.118 Time-averaged FFT spectrum and waveform of the sound of a snared side-drum roll of increasing intensity

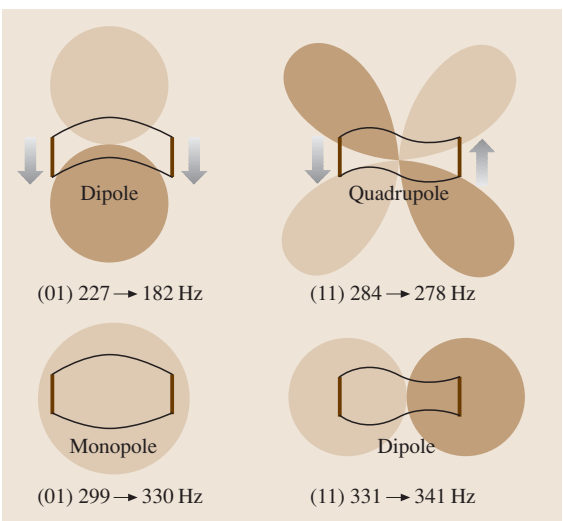


Fig. 15.119 Coupled motions of the two drumheads of a side drum, indicating the change in frequency of the drumhead modes from air coupling within the drum and the associated polar radiation patterns (after *Zhao* et al. [15.209])

Amazon
EX1013 Page 656

Although any induced motion of the relatively heavy supporting structure will not significantly affect the frequencies of the normal modes, it can result in appreciable additional damping. As a consequence, the sound of a side drum can sound very different depending on how it is supported – freely suspended on rubber bands or rigidly supported by a heavy stand (*Rossing* [15.201] Sect. 4.4). Rossing has also made detailed vibrational and holographic studies of the free drum shell ([15.201] Fig. 4.5). As the induced motions are only a fraction of a percent of those of the drumhead, such vibrations will not radiate a very large amount of sound. Nevertheless, they may play an important role in defining the initial sound.

Bass Drum

The bass drum is large with a typical diameter of 80–100 cm. It can produce a peak sound output of 20 W, the largest of any orchestral instrument. Single-headed drums are used when a well-defined sense of pitch is required, but double-headed drums sound louder because they act as monopole rather than dipole sources. Modern bass-drum heads generally use 0.25 mm-thick mylar, though calfskin is also used.

The *batter* or beating drum head is normally tuned to about a fourth above the *carry* or resonating head (*Fletcher* and *Rossing* [15.5], Sect. 18.2). The change in modal frequencies induced by the enclosed air is illustrated in Table 15.8 (*Fletcher* and *Rossing* [15.5], Table 18.5). Note the strong splitting of the lowest frequency (01) and (11) normal modes, when the two heads are tuned to the same tension. In this example, the frequencies of the first five modes are almost harmonic, giving a sense of pitch to the sound (audio CD-ROM illustrates the rather realistic synthesised sound of the first six modes of the batter head tuned to the carry head with equal amplitude and decay times). Drums with heads tuned to the same pitch have a distinctive timbre.

Table 15.8 Modal frequencies in Hz of the batter head of a 82 cm bass drum (after *Fletcher* and *Rossing* [15.5])

Mode	Batter head with carry head at lower tension	Batter head with heads at same tension (Hz)
(01)	39	44 , 104 split normal modes
(11)	80	76, 82 split normal modes
(21)	121	120
(31)	162	160
(41)	204	198
(51)	248	240

15.4.2 Bars

This section is devoted to percussion instruments based on the vibration of wooden and metallic bars, both in isolation and in combination with resonating air columns. Such instruments are referred to as idiophones – bars, plates and other structures that vibrate and produce sound without having to be tensioned, unlike the skins of a drum (membranophones). Representative instruments considered in this section include the glockenspiel, celeste, xylophone, marimbas, vibraphone and triangle.

The vibrations of thick and thin plates have already been considered in the context of the vibrational modes of the wooden plates of stringed instruments (Sect. 15.2.6). The most important acoustic modes of a rectangular plate are the torsional (twisting) and flexural (bending) modes, both of which involve acoustically radiating displacements perpendicular the surface of the plate.

The torsional vibrations of a bar are discussed by *Fletcher* and *Rossing* ([15.5], Sect. 2.20). For a bar of length L, the frequency of the twisting modes is given by $f_n = nc_\theta/2L$, where c_θ is the dispersionless velocity of torsional waves. For a rectangular bar with width w significantly larger than its thickness h, by $c_\theta \sim 2t/w\sqrt{E/2\rho(1+\nu)}$, where E is the Young's modulus and ν the Poisson ratio. For a bar with circular, cross-section, like the sides of a triangle, $c_\theta = \sqrt{E/2\rho(1+\nu)}$.

Musically, the most important modes of a thin bar are the flexural modes involving a displacement z perpendicular to their length, which for a rectangular bar satisfies the fourth-order wave equation

$$\frac{E}{12(1-\nu^2)}h^2\frac{\partial^4 z}{\partial x^4} + \rho\frac{\partial^2 z}{\partial t^2} = 0\,, \qquad (15.149)$$

with standing-wave solutions of the general form

$$z(x,t) = (A\sin kx + B\cos kx + C\sinh kx + D\cosh kx)\,\mathrm{e}^{\mathrm{i}\omega t}\,, \qquad (15.150)$$

where

$$\omega = \sqrt{\frac{E}{12\rho(1-\nu^2)}}hk^2\,. \qquad (15.151)$$

As discussed in the earlier section on the vibrational modes of soundboards and the plates of a violin or guitar, the sinh and cosh functions decay away from the ends of the bar or from any perturbation in the geometry, such

Amazon
EX1013 Page 657

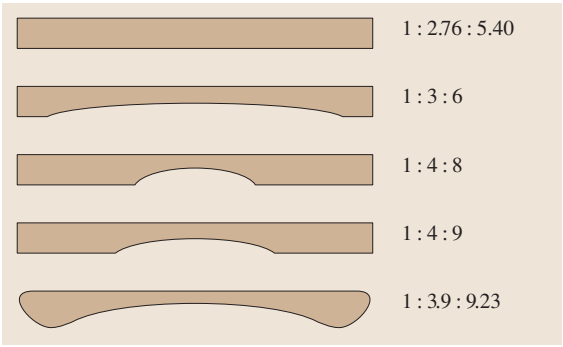


Fig. 15.120 Ratio of the frequencies of the first three partials of a simple rectangular bar for three selectively thinned xylophone bars and a typical marimba bar (after *Fletcher* and *Rossing* [15.5])

as local thinning or added mass, over a distance

$$k^{-1} \sim \left(\frac{E}{12\rho\left(1-\nu^2\right)}\right)^{1/4} \sqrt{\frac{h}{\omega}}\,.$$

Well away from the ends of a bar, the standing-wave solutions at high frequencies are therefore dominated by the sinusoidal wave components.

The lowest flexural modes of a freely supported thin rectangular bar are inharmonic, with frequencies in the ratios 1:2.76:5.40:8.93. However, by selectively thinning the central section, the frequency of the lowest mode can be lowered, to bring the first four harmonics more closely into a harmonic ratio, as illustrated schematically for a number of longitudinal cross-sections in Fig. 15.120, which also includes the measured frequencies of a more-complex-shaped marimba bar (*Fletcher* and *Rossing* [15.5], Figs. 19.2 and 19.7).

Pitch Perception

The audio CD-ROM contrasts the synthesised sounds of the first four modes of a simple rectangular bar, followed by a note having the same fundamental but with partials in the ratio 1:3:6, while the final note has the inharmonic (1:8.96) fourth partial of the rectangular bar added. Despite the inharmonicity of the partials, the synthesised sound of a rectangular bar has a surprisingly well-defined sense of pitch. The main effect of replacing the second and third partials with partials in the ratio 1:3:6 is to raise the perceived pitch by around a tone, even though the first partial is unchanged at 400 Hz. This again emphasises that the perceived pitch is determined by a weighted average of the partials present and not by the fundamental tone alone. Adding the fourth inharmonic partial gives an increased edge or *metallic* characteristic to the perceived sound, without changing the perceived pitch.

The metal or wooden bars of tuned percussion instruments are usually suspended on light threads or rest on soft pads at the nodal positions of their fundamental mode, which reduces the damping to a minimum. The resulting 60 dB decay time for an aluminium vibraphone bar can be as long as 40 s (*Rossing* [15.201] Sect. 7.3) compared with a few seconds for the lower-frequency notes on the wooden bars of a marimba (*Rossing* [15.201], Sect. 6.4). The damping of vibrating bars is therefore highly material dependent and is largely determined by internal damping losses rather than radiation. This accounts for the very different sounds of wooden and metal bars on instruments like the glockenspiel and xylophone.

Glockenspiel and Celeste

The simplest of all idiophones are those instruments based on the vibrations of freely supported thin rectangular plates. Such instruments include the glockenspiel played with a variety of hard and soft round-headed hammers and the celeste played with strikers operated from a keyboard, with a sustaining pedal to control the damping. The playing range of the glockenspiel is typically

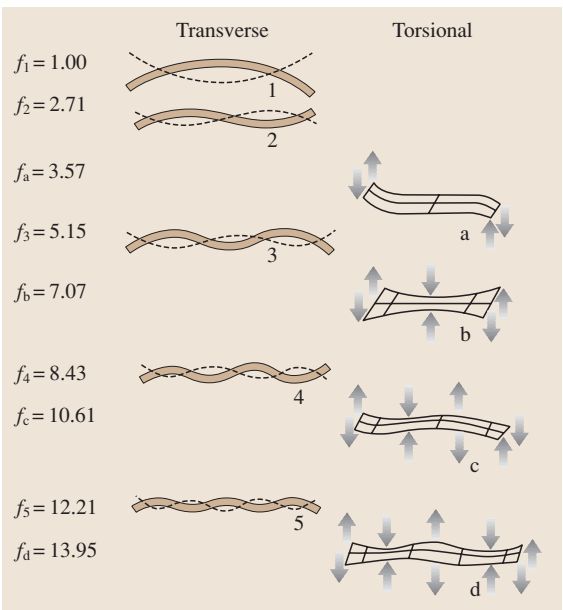


Fig. 15.121 Measured flexural and torsional modes of a glockenspiel bar (after *Fletcher* and *Rossing* [15.5])

Amazon
EX1013 Page 658

two and a half octaves from G5 to C8, while the celeste has a range of 4–5 octaves, with a separate box-resonator used for each note.

Both instruments produce a bright, high-pitched, bell-like, sparkling sound, as in the *Dance of the Sugar Plum Fairy* in Tchaikovsky's *Nutcracker Suite*. No attempt is made to adjust the thickness of the plates to achieve a more nearly harmonic set of modes.

Figure 15.121 illustrates the lowest order flexural and torsional modes and measured ratios of frequencies for a C6 glockenspiel bar (*Fletcher* and *Rossing* [15.5], Fig. 19.1). A typical wave-envelope and spectrum of a glockenspiel note is shown in Fig. 15.122. FFT spectra are shown for 200 ms sections from the initial transient and after 200 ms. There are two strongly contributing weakly damped partials at 1045 Hz and 2840 Hz (in the ratio 1:2.72), which can be identified as the first two flexural modes of the bar. The lower of the two long sounding partials gives the sense of pitch, while the strong inharmonic upper partial gives the note its "glockenspiel" character. Audio CD-ROM compares the recorded glockenspiel note with synthesised tones at 1045 and 2840 Hz, first played separately then together. In this case, the inharmonicity of the strongly sounded partials plays a distinctive role in defining the character of the sound. The spectrum is typical of all the notes on a glockenspiel, which demonstrates that only a few of the modes shown in Fig. 15.121 contribute significantly to the perceived sound.

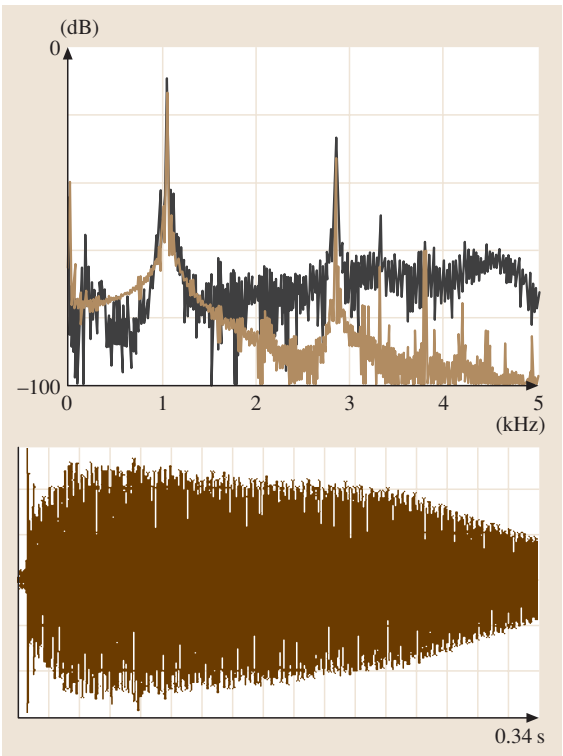


Fig. 15.122 The waveform envelope and FFT spectra of the prompt sound (upper) and the sound after ≈ 0.2 s (lower) of a typical glockenspiel note illustrating the long-time dominance of a few slowly decaying, inharmonic partials

Xylophone, Marimba and Vibraphone

We now consider a number of acoustically related idiophones, with bars that are selectively thinned to produce a more nearly harmonic set of resonances and well-defined sense of pitch.

The modern xylophone has a playing range of typically 3 to 4 octaves and uses wooden bars, which are

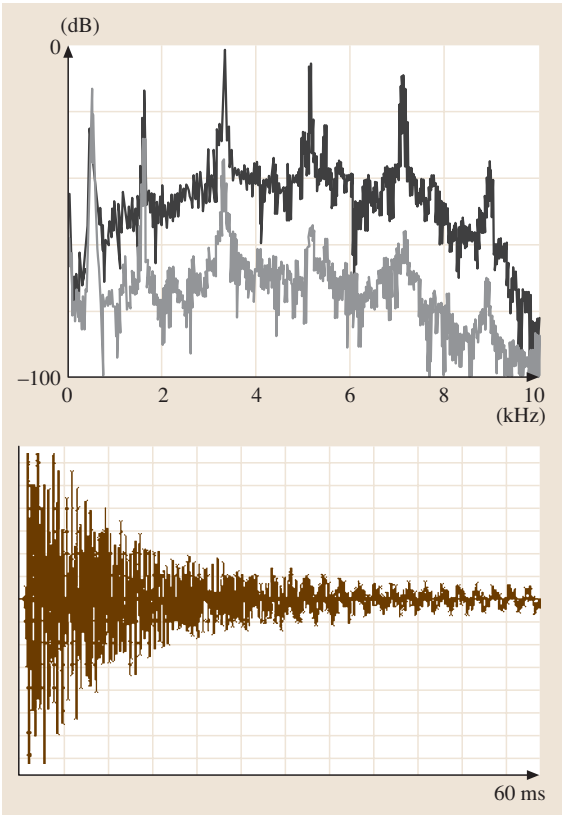


Fig. 15.123 The initial 60 ms of a xylophone waveform showing the rapid decay of high-frequency components and FFT spectra at the start of the note (upper trace) and after 0.2 s (lower trace), highlighting the persistence of the strong low-frequency air resonances excited

Amazon
EX1013 Page 655

undercut on their back surface to improve the harmonicity of the lower frequency modes (Fig. 15.120). Each bar has an acoustic resonator immediately below it, consisting of a cylindrical tube, which is closed at the far end. Any of the flexural and torsional modes can contribute to the initial sound, when the bar is struck by a hammer; however, most modes die away rather quickly so that at longer times the sound is dominated by the resonances of the coupled pipe resonator.

Figure 15.123 shows the initial part of the waveform and spectrum of a typical xylophone note (CD-ROM), illustrating the initial large amplitudes and rapid decay of the higher frequency bar modes excited and the strongly excited but slowly decaying resonances of the first two modes of the air resonator. All modes contribute to the initial sound but the sound at longer times is dominated by the lowest-frequency bar modes and resonantly tuned air resonators.

The marimba is closely related to the xylophone, but differs largely in its playing range of typically two to four and a half octaves from A2 (110 Hz) to C7 (2093), though some instruments play down to C2 (65 Hz). In contrast to xylophone bars, which are undercut near their centre to raise the frequency of their second partial from 2.71 to 3.0 above the fundamental, marimba bars are often thinned still further to raise the frequency of the second partial to four times the fundamental frequency (Fig. 15.119).

Marimbas produce a rather mellow sound and are usually played with much softer sticks than traditionally used for the xylophone. Although the marimba is nowadays used mostly as a solo percussion instrument, in the 1930s ensembles with as many as 100 marimbas were played together. In many ways, such ensembles were the forerunners of today's Caribbean steelbands, to be described later in this section.

The vibraphone is similar to the marimba, but uses longer-sounding aluminium rather than wooden bars and typically plays over a range of three octaves from F3 to F6. Like the marimba, the bar thickness is varied to give a second partial two octaves above the fundamental. They are usually played with soft yarn-covered mallets, which produce a soft, mellow tone. In addition, the vibraphone incorporates electrically driven rotating discs at the top of each tuned air resonator, which periodically changes the coupling. This results in a strong amplitude-modulated vibrato effect. The wave envelope of audio CD-ROM is shown in Fig. 15.124 for a succession of notes played on the vibraphone with vibrato, which are then allowed to decay freely. The vibrato rate can be adjusted by changing the speed of the electric motor.

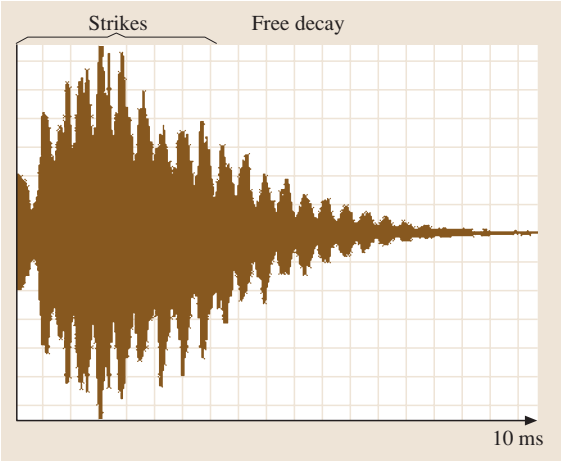


Fig. 15.124 Envelope of a succession of notes on the vibraphone, which freely decay with modulated coupling to tuned air resonators to produce an amplitude modulated vibrato effect

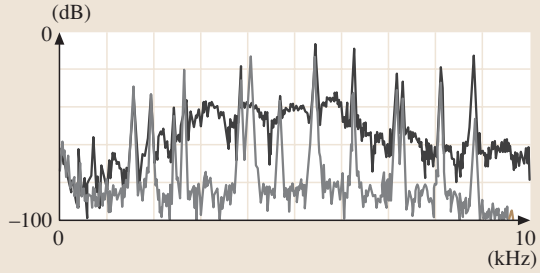
Note the very long decay of the sound, which can be controlled by a pedal-operated damper.

Triangle

The triangle is a very ancient musical instrument formed from a cylindrical metal bar bent into the shape of a triangle, with typical straight arm lengths of 15–25 cm. They are usually struck with a similar-diameter metal rod. Although the instrument is small and therefore a very inefficient acoustic radiator, it produces a characteristic high-frequency *ping* or repetitive high-pitched rattle, which is easily heard over the sound of a large symphony orchestra (audio CD-ROM). The quality of the sound can be varied by beating at different positions along the straight arms. The triangle is usually supported be a thread around the top bend of the hanging instrument.

The flexural modes of a freely suspended bar of circular cross section are $f_n \sim (a/2)\sqrt{E/\rho}(2n+1)^2\pi/8L^2$, with frequencies in the ratios $3^2{:}5^2{:}7^2{:}9^2{:}11^2{:}13^2$ (see Sect. 15.2.4). Transverse flexural vibrations can be excited perpendicular or parallel to the plane of the instrument. For vibrations perpendicular to the plane, the bends are only small perturbations. Transverse modes in this polarisation are therefore almost identical to those of the straight bar from which the triangle are bent (*Rossing* [15.212] Sect. 7.6). However, for flexural vibrations in the plane, there is a major discontinuity in impedance at each bend, because the transverse vibrations within one arm couple strongly to longitudinal vibrations in

Amazon
EX1013 Page 660



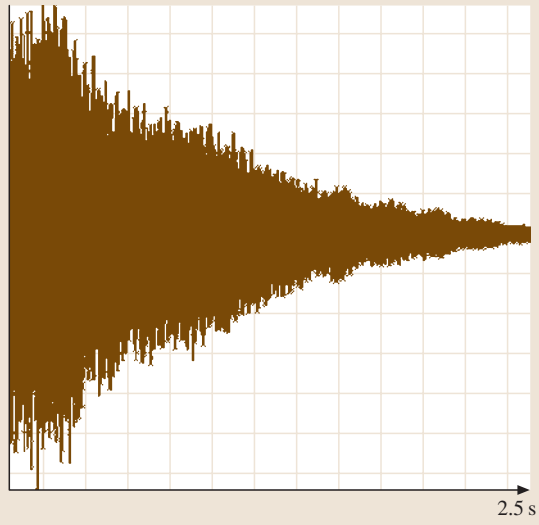


Fig. 15.125 The decaying waveform and FFT spectra at the start (upper trace) and after 1 s (lower trace) of a struck triangle note

the adjacent arms. Hence each arm will support its own vibrational modes, which will be coupled to form sets of normal-mode triplets, since each arm is of similar length.

Figure 15.125 illustrates the envelope and 50 ms FFTs of the initial waveform and after 1 s, illustrating the very well-defined and only weakly attenuated high-frequency modes of the triangle. Note the wide-band spectrum at the start of the note from the initial impact with the metal beater.

Chimes and Tubular Bells

We include orchestral chimes and bells in this section because their acoustically important vibrations are flexural modes, just like those of a xylophone or triangle. The radius of gyration of a thin-walled cylindrical tube of radius a is $\approx a/\sqrt{2}$. The frequency of the lowest flexural modes is then be given by $f_n \sim a\sqrt{E/2\rho}(2n+1)^2\pi/8L^2$.

Orchestral chimes are generally fabricated from lengths of 32–38 mm-diameter thin-walled tubing, with the striking end often plugged by a solid mass of brass with an overhanging lip, which provides a convenient striking point.

Fletcher and *Rossing* ([15.5], Sect. 19.8) note that the perceived pitch of tubular bells is determined by the frequencies of the higher modes, excited with frequencies proportional to 9^2, 11^2, and 13^2, in the approximate ratios 2:3:4. The pitch should therefore sound an octave below the lowest of these. Readers can make there own judgement from audio CD-ROM, which compares the rather realistic sound of a tubular bell synthesised from the first six equal amplitude modes of an ideal bar sounded together, followed by the 9^2, 11^2, and 13^2 modes in combination, and then by a pure tone an octave below the 9^2 partial. Such comparisons highlight the problem of subjective pitch perception in any sound involving a combination of inharmonic partials.

15.4.3 Plates

Flexural Vibrations

This section describes the acoustics of plates, cymbals and gongs, which involve the two-dimensional flexural vibrations of thin plates described by the two-dimensional version of (15.149). Unlike stringed instruments, we can usually assume that the plates of percussion instruments are isotropic. Well away from any edges or other perturbing effects such as slots or added masses, the standing-wave solutions at high frequencies will be simple sinusoids. However, close to the free edges, and across the whole plate at low frequencies, contributions from the exponentially decaying solutions will be equally important over a distance $\sim (E/12\rho(1-\nu^2))^{1/4}(h/\omega)^{1/2}$. The nodes of the sinusoidal wave contributions will be displaced a distance $\sim 1/4\lambda$ from the edges. Hence, the higher frequency modes of a freely supported rectangular plate of length a, width b and thickness h will be given, to a first approximation, by

$$\omega_{mn} \sim h\left(\frac{E}{12\rho\left(1-\nu^2\right)}\right)^{1/4}\pi^2 \left[\left(\frac{m+1/2}{a}\right)^2+\left(\frac{n+1/2}{b}\right)^2\right]. \quad (15.152)$$

A musical instrument based on the free vibrations of a thin rectangular metal plate is the thunder plate, which when shaken excites a very wide range of closely spaced

Amazon
EX1013 Page 661

modes, which can mimic the sharp clap followed by the rolling sound of thunder in the clouds.

Before the age of digital sound processing, such plates were often used in radio and recording companies to add artificial reverberation to the recorded sound. The plate was suspended in a chamber along with a loudspeaker and pick-up microphone. The sound to which reverberation was to be added was played through the loudspeaker, which excited the weakly damped vibrational modes of the plate, which were then re-recorded using the microphone to give the added reverberant sound. As described earlier (Sect. 15.2.4), the density of vibrational modes of a flat plate of area A and thickness h is given by $1.75A/c_{\mathrm{L}}h$. This can be very high for a large-area thin metal sheet, giving a reverberant response with a fairly uniform frequency response.

Most percussion instruments which involve the flexural vibrations of thin sheets are axially symmetric, such as cymbals, gongs of many forms, and the vibrating plate regions of steeldrums or pans used in Caribbean steelbands. Such instruments have many interesting acoustical properties, many derived from highly nonlinear effects when the instrument is instrument is struck strongly.

The displacements of the flexural modes of an axially symmetric thin plate in polar coordinates are given by

$$z(r,\phi,t) = [AJ_m(k_{mn}r) + BI_m(k_{mn}r)] \\ [C\cos(m\phi) + D\sin(m\phi)]\,\mathrm{e}^{\mathrm{i}\omega_{mn}t}\,, \qquad (15.153)$$

where $J_m(kr)$ and $I_m(kr)$ are ordinary and hyperbolic Bessel functions, the equivalent of the sinusoidally varying and exponentially damped sinh and cosh functions used to describe flexural waves in a rectangular geometry. The hyperbolic Bessel functions are important near the edges of a plate or at any perturbation, but decay over a length scale of $\sim k_{mn}^{-1}$. The values of k_{mn} are determined from the boundary conditions, in just the same way as considered earlier for rectangular plates.

The first six vibrational modes for circular plates with free, hinged and clamped outside edges are shown in Fig. 15.126, with frequencies expressed as a ratio relative to that of the fundamental mode (*Fletcher* and *Rossing* [15.5], Sect. 3.6). In each case, for large values of m and n, the frequency is given by the empirical Chladni's Law (1802),

$$\omega_{mn} \sim \sqrt{\frac{E}{12\rho\left(1-\sigma^2\right)}}\,\frac{\pi^2 h}{4a^2}(m+2n)^2\,, \qquad (15.154)$$

justified much later by *Rayleigh* ([15.3] Vol. 1, Chap. 10). For arched plates, *Rossing* [15.213] showed

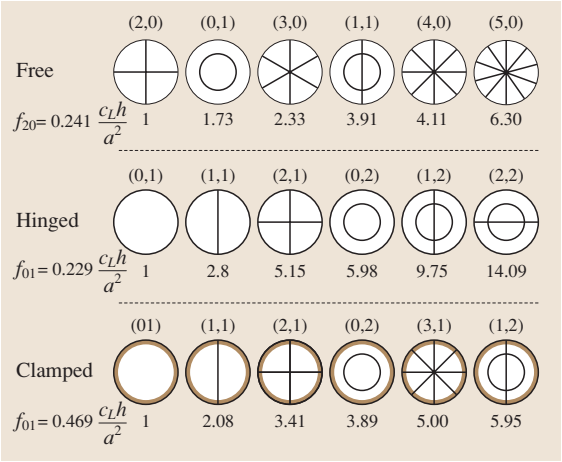


Fig. 15.126 The vibrational modes of circular plates with free, hinged and clamped outer edges, with the lowest frequencies and ratio of frequencies of higher modes indicated

that the frequencies are more closely proportional to $(m+bn)^p$, where p is somewhat less than 2 and b is in the the range 2–4.

All the axially symmetric modes involving nodal diameters are doubly degenerate, with a complementary solution with nodal diameters bisecting those drawn in Fig. 15.126. Any perturbation of the structure from cylindrical symmetry will split the degeneracy of such modes. Modes with a nodal diameter passing through the point at which the plate is struck will not be excited.

Instruments like the cymbal are slightly curved over their major surfaces but with a sudden break to a more strongly arched central section, to which the leather holding straps or support stand are attached. The outer edges can therefore be treated as free surfaces with the transition to the central cupped region providing an additional internal boundary condition, which will be intermediate between clamped and hinged. In contrast, gongs tend to have a relatively flat surface but with their edges turned though a right angle to form a cylindrical outer rim. The rim will add mass to the modes involving a significant radial displacement at the edge, but will also increase the rigidity of any mode having an azimuthal dependence. Thus, although the modes of an ideal circular plate provide a guide to modal frequencies and waveforms, we would expect significant deviations in modal frequencies for real percussion instruments. Any sudden change in plate profile on a length scale smaller than an acoustic wavelength will involve a strong coupling between the transverse flexural and longitudinal waves resulting in reflections from the discontinuity in

Amazon
EX1013 Page 662

acoustic impedance. This is why, for example, the indented region on the surface of a steeldrum pan can support localised standing waves on the indented regions. Indented areas of different sizes can then be used to produce a wide range of different notes on a single drum head with relatively little leakage in vibrations between them.

Nonlinear Effects

Figure 15.127 illustrates the cross section of some typical axially symmetric cymbals, gongs and a steelpan. Nonlinear effects in such instruments can be important when excited at large amplitudes. Such effects are particularly marked for gongs with relatively thin plates. For Chinese opera gongs, the nonlinearity can result in dramatic upward or downward changes in the pitch of the sound after striking. In addition, nonlinearity results in mode conversion, with the high-frequency content of cymbal and large gong sounds increasing with time, giving a characteristic *shimmer* to the sound quality.

The shape dependence of the nonlinearity arises from the arching of the plate. For the lowest mode of a flat plate, the potential energy initially increases quadratically with displacement, though the energy increases more rapidly at large-amplitude excursions from stretching, as indicated schematically in Fig. 15.128a. Although the energy of an arched plate initially also increases quadratically with distance about its displaced equilibrium position, the energy will first increase then decrease when the plate is pushed through the central plane, (Fig. 15.128b). If the plate were to be pushed downwards with increasing force, it would suddenly spring to a new equilibrium position displaced about the central plane by the same initial displacement, but in the opposite direction. In combination with a Helmholtz radiator, this is indeed how some insects such as cicadas generate such strong acoustic signals – as high as 1 mW at 3 kHz (see Chap. 19 by Neville Fletcher). The nonlinear bistable flexing of a belled-out plate can be disastrous, turning a cheap pair of thin cymbals inside out, when crashed together too strongly.

The central peak in potential energy of an arched plate, like the gently domed Chinese gong illustrated in Fig. 15.127, therefore leads to a reduced restoring force for large-amplitude vibrations and a lower pitch. In contrast, the restoring force of a flat plate, like the central playing region of the upper of the two Chinese gongs, increases with vibration amplitude. This arises from the additional potential energy involved in stretching the plate, in just the same way as we considered earlier for the large-amplitude vibrations of a stretched string.

For a flat plate of thickness h, *Fletcher* [15.214] has shown that the increase in frequency of the lowest axisymmetric mode increases with the amplitude of vibration a approximately as

$$\omega \sim \omega_0 \left[1 + 0.16(a/h)^2\right] . \qquad (15.155)$$

The nonlinearity also generates additional components at three times the frequency of any initial modes present and at multiples of any new modes excited. In addition it produces cross-modulation products when more than one mode is present. For example, for modes with frequencies f_i and f_2 initially present, the nonlinearity will generate inter-modulation products at $2f_1 \pm f_2$ and $2f_2 \pm f_1$.

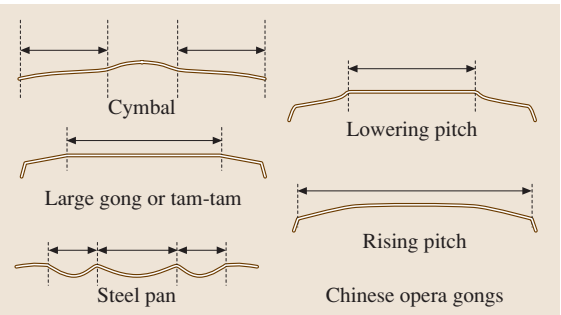


Fig. 15.127 Schematic cross sections of axially symmetric plate instruments and a steelpan, with *arrows* indicating the principal areas producing sound

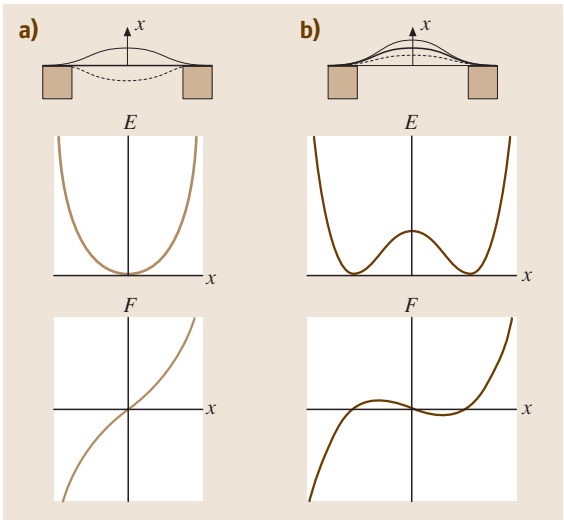


Fig. 15.128a,b Potential energy and restoring force of **(a)** a clamped flat and **(b)** arched circular plate as a function of displacement from the central plane

Amazon
EX1013 Page 663

Grossmann et al. [15.215] and *Fletcher* [15.214] have considered the vibrations of a spherical-cap shell of height H and thickness h. Interestingly, the change in frequency of the asymmetric vibrations about the equilibrium point depends only on the ratio of the amplitude to the arching height, a/H, as illustrated in Fig. 15.129. When the height of the dome is much less than the thickness, the frequency increases approximately as a^2, as expected from the induced increase in tension with amplitude. However, when the arching becomes comparable with and greater than the thickness, the asymmetry of the potential energy dominates the dynamics and results in an initial decrease in frequency, which increases strongly with the ratio of arching height to thickness. At very large amplitude, $a \gg H$, the frequency is dominated by the increase in tension and therefore again increases with amplitude like a flat plate. At large amplitudes, *Legge* and *Fletcher* [15.216] have shown that changes in the curvature of the plate profile result in a large transfer of energy to the higher-frequency plate modes.

We now show how many of the above properties relate to the sounds of cymbals, gongs of various types and steelpans.

Cymbals

Many types of cymbals are used in the classical symphony orchestra, marching bands and jazz groups. They are normally made of bronze and have a diameter of 20–70 cm. The low-frequency modes of a cymbal are very similar to those of a flat circular plate and can be described using the same (mn) mode nomenclature (*Fletcher* and *Rossing* [15.5], Fig. 20.2). However, small changes in curvature across a cymbal will results in modes that are linear combinations of ideal circular-plate modes.

Cymbals are usually played by striking with a wooden stick or soft beater or a pair can be crashed together, each method producing a distinctive sound. They can even be played by bowing with a heavy rosined bow against the outer rim. Rossing and Shepherd showed that the characteristic 60 dB decay time of the excited modes of a large cymbal varies approximately inversly proportional with frequency, with a typical decay time for the lowest (20) mode as long as 300 s (*Fletcher* and *Rossing* [15.5], Fig. 20.5)

Figure 15.130 shows the waveform envelope and spectra of the initial sound of a cymbal crash and after 1 s. Audio CD-ROM illustrates a recorded cymbal crash followed by the same sound played first through a 0–1 kHz and then a 1–10 kHz band-pass filter, illustrating the decay of the low- and high-frequency wide-band noise.

When a cymbal is excited with a drumstick, waves travel out from the excitation point with a dispersive group velocity proportional to k, inversely proportional to the dimensions of the initial flexural indention of the surface made by the drumstick. The dispersive pulse strikes and is reflected from the edges of the cymbal

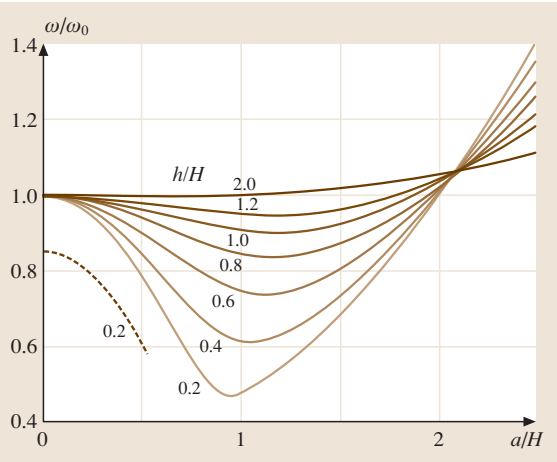


Fig. 15.129 Vibration amplitude a dependence of frequency of lowest axisymmetric mode of a spherical cap of dome height H as a function of thickness h to H ratio

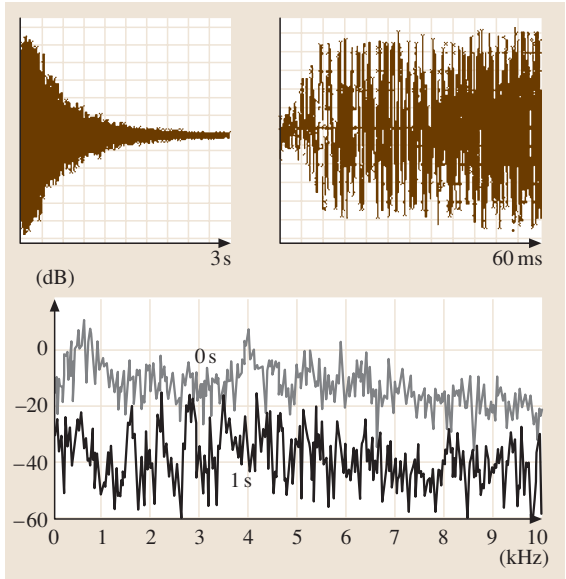


Fig. 15.130 Wave envelope and spectrum at start and after 1 s of a cymbal clash illustrating wide-band noise at all times

Amazon
EX1013 Page 664

and the transitional region to the central curved cup, so that eventually the energy will be dispersed across the whole vibrating surface. This has been investigated using pulsed video holography by *Schedin* et al. [15.217]. On reflection there will also be considerable mixing of modes. In addition, because the plates are rather thin and are often hit extremely strongly, nonlinear effects are important. On large cylindrical plates this results in the continuous transfer of energy from strongly excited low-frequency modes to higher modes. Many of the nonlinear effects can be investigated in the laboratory using sinusoidal excitation. Measurements by *Legge* and *Fletcher* [15.216] have revealed a wide range of nonlinear effects including the generation of harmonic, bifurcations and even chaotic behaviours at large intensities.

When a plate is struck by a beater, the acoustic energy is distributed across a very wide spectrum of closely spaced resonances of the plate. To distinguish individual partials requires the sound to be sampled over a time of at least $\approx 1/\Delta f$, where Δf is the separation of the modes at the frequencies of interest. However, because the modes of a large cymbal are so closely spaced, the times involved can be rather long. Unlike the sound of pitched instruments such as the glockenspiel and xylophone, there are no particular resonances of the cymbal that dominate the sound, which is characterised instead by the *sizzle* produced by the very wide spectrum of very closely spaced resonances almost indistinguishable from wide-band high-frequency noise.

Large Gongs and Tam-Tams

Gongs are also very ancient instruments, which have a very characteristic sound when strongly struck by a soft beater, notably as a prelude to classic films by the Rank organisation. A typical tam-tam gong used in a symphony is a metre or even larger in diameter. It is made of bronze and, like cymbals, is sufficiently ductile not to shatter when strongly hit. The damping is very low, so the sound of large gongs can persist for very long times.

When initially struck strongly by a soft beater, the initial sound is largely associated with the lower-frequency partials that are strongly excited. However, on a time scale of a second or so, the sound can appear to grow in intensity, as nonlinear effects transfer energy from lower- to higher-frequency modes (audio CD-ROM). This is illustrated in Fig. 15.131 (*Fletcher* and *Rossing* [15.5], Fig 20.8), which shows the build up and subsequent decay of acoustic energy in the higher-frequency bands at considerable times after the initial impact. The fluctuations in intensity within these bands were taken as evidence for chaotic behaviour. However, even in a linear system, interference between the large number of very closely spaced inharmonic partials would also result in apparently random fluctuations in amplitude.

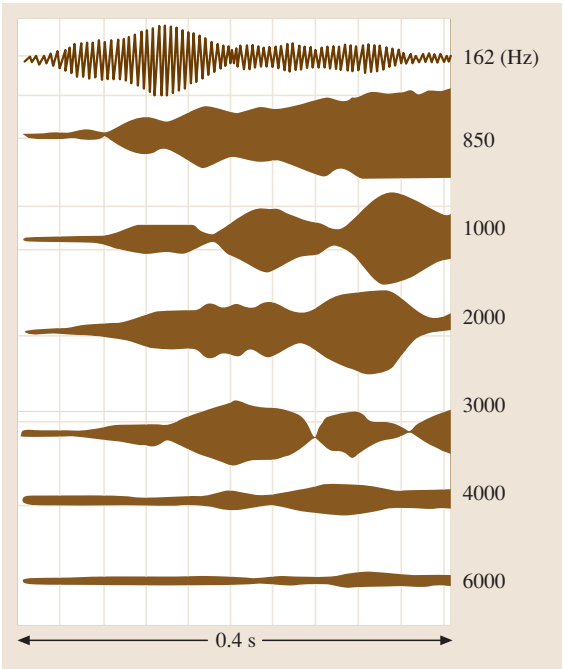


Fig. 15.131 Buildup and decay in intensity of a struck tam-tam sound in frequency bands centred on the indicated frequencies (after *Fletcher* and *Rossing* [15.5])

Chinese Opera Gongs

Chinese gongs provide the most dramatic illustration of nonlinearity in percussion instruments, with upwards or downwards pitch glides of several semitones over a sizeable fraction of a second after being strongly struck. The direction of the pitch *glide* depends on the profile of the vibrating surface as previously described.

Figure 15.132 shows the decaying waveforms of the sound of three Chinese gongs with a downward pitch glide (audio CD-ROM) played in succession. The initial spectrum of the third note played is followed by spectra at 0.75 and 1.5 s after striking, illustrating the transfer of energy to lower-frequency modes. The much broader width of the initial spectrum reflects the decrease in lifetime of the initial modes struck resulting from the nonlinear loss of energy to higher-frequency modes. The two well-defined peaks between the two major peaks are

Amazon
EX1013 Page 665

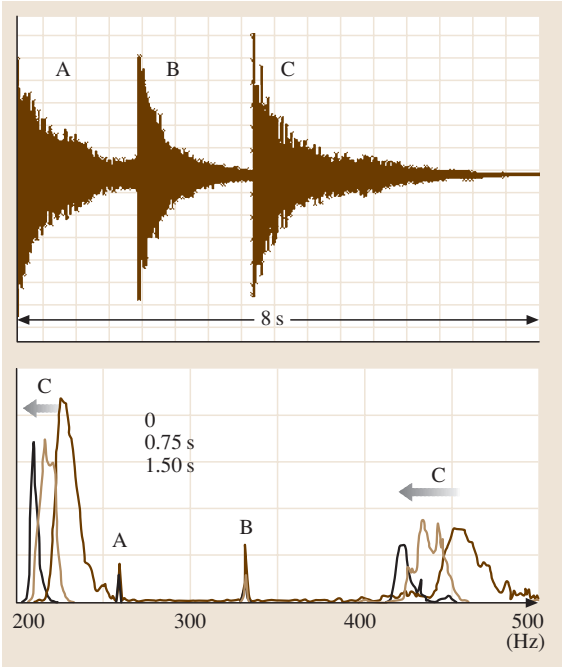


Fig. 15.132 The wave envelope of sounds from three downward-sliding Chinese gongs followed by the spectrum of the third gong at the start, after 0.75 and 1.5 s illustrating the nonlinear frequency shifts

from the long-ringing principal partials of the first two gongs.

Steelpans

Finally, in this section on percussion instruments based on vibrating plates, we consider steelpans originating from the Caribbean, which were initially fabricated by indenting the top of oil cans left on the beaches by the British navy after World War II. They have become a immensely popular instrument in that part of the world and are just as interesting from a musical acoustics viewpoint (see *Fletcher* and *Rossing* [15.5], Sect. 20.7 for further details).

Pitched notes on a given drum are produced by hammered indentations of different sizes on the top face of the drum. Different ranges of notes are produced by drums or pans of different sizes (e.g. lead tenor, double tenor, alto, cello and bass). Typical indented regions on a double-tenor steelpan are shown in Fig. 15.133, adapted from drawings for a full set of pans in *Fletcher* and *Rossing* ([15.5], Fig. 20.17). The various indented areas on the drum head can be considered as an array of relatively weakly coupled resonators. An individual indented area on an infinite sheet would have very similar acoustic modes as those of a hemispherical cap indented in an infinite plate. The frequency of the modes would be determined by the size and arching of the indented areas. The effective cap size would be defined by the rate of change of the curvature of the plate and the associated change in the acoustic impedance at the edge of the indented region. However, all such regions on a steelpan will be coupled together by the relatively weak transfer of acoustic energy between them, to form a set of coupled modes. Hitting one particular region

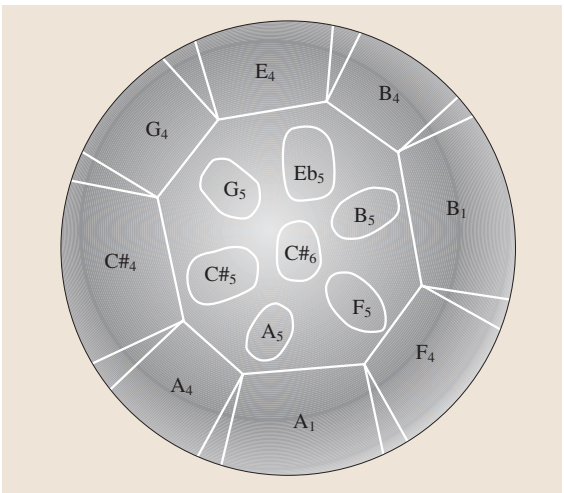


Fig. 15.133 Typical indentation areas in a tenor steelpan (after *Fletcher* and *Rossing* [15.5])

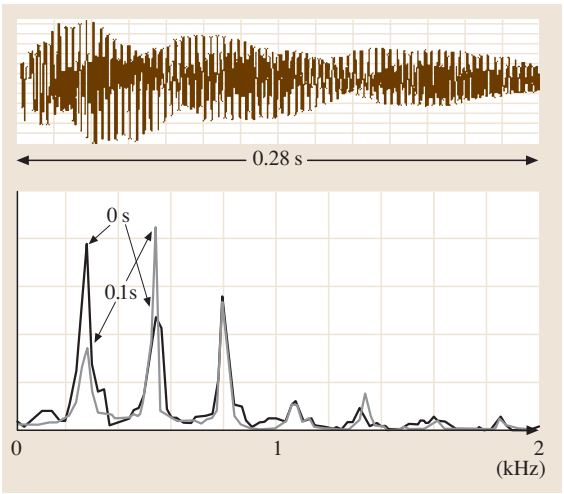


Fig. 15.134 Waveform and initial and time-delayed spectra of the note C#4 on a steeldrum

Amazon
EX1013 Page 666

will therefore excite other regions, especially those that have closely matching partials. The coupling between such regions has been investigated holographically by *Rossing* ([15.201] Fig. 20.20).

Audio CD-ROM illustrates a succession of notes played on a steeldrum. Figure 15.134 shows a typical decaying waveform with initial and time-delayed spectra of a single note. A relatively large number of well-defined modes can easily be identified. However, the subjective absolute pitch of the note is not particularly well defined and there is a strong sense of pitch circularity in the sound of an ascending scale (Sect. 15.1.3 and audio CD-ROM), sometimes making it difficult to identify the octave to which a particular note should be attributed. Note the *increase* in the amplitude of the second partial with time, which could result from nonlinear effects in such thin-walled structures, or possibly from interference beats between degenerate modes split in frequency by the lack of axial symmetry of the hand-beaten indentations.

15.4.4 Shells

Blocks and Bells

Finally, we consider the acoustics of three-dimensional shell-like structures. This could include percussion instruments such as the wooden-block and hollow gourd instruments like the maracas. However, the physics of such instruments is essentially the same as that of the soundbox of stringed instruments and involves little of additional acoustic interest. In this section, we will therefore concentrate on the acoustics of bell-like structures, which are usually axially symmetric structures, closed at one end, and of variable thickness and radius along their length. We will also consider non-axisymmetric bells with a quasi-elliptical cross section, which produce two different pitched notes, depending on where the bell is struck. All such structures have a rich spectrum of modes, which are generally tuned to give long-ringing notes with a well-defined sense of pitch. The bronze used in their construction is typically an alloy of 80% copper and 20% tin and has to be sufficiently ductile not to crack under the impact of the beater or clapper. Metallurgical treatment is required to produce a grain structure producing little damping at acoustic frequencies.

Some of the oldest bells are to be found in China. Such bells are supreme examples of the art of bronze casting dating to the fifth century BC. Bells in church towers have traditionally marked the passage of time, while peals of bells with internal swinging clappers continue to summon the faithful to worship. In more recent times, carillons with up to 77 tuned bells have been developed to play keyboard music from the top of specially constructed bell towers, notably in the centres of Dutch and American towns and college campuses. Bells come in all sorts of shapes and sizes ranging from small hand bells to the giant church bells on display inside the Kremlin walls. However, the acoustics of all bells is essentially the same, so no attempt will be made to provide a comprehensive coverage of every type, [for such information, see *Fletcher* and *Rossing* ([15.5], Chap. 21), and *Rossing* [15.201]].

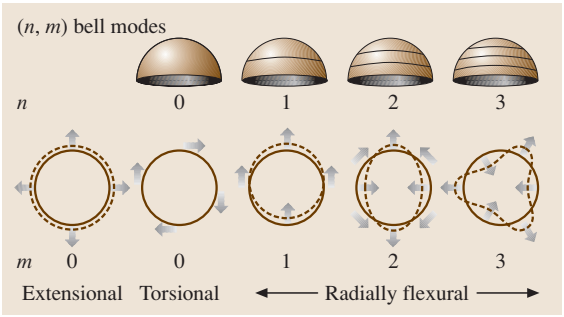


Fig. 15.135 Nomenclature of the (m, n) modes of a bell illustrating displacements of rim for given m-values

Bell modes can be related to the longitudinal, torsional and flexural modes of a cylindrical disc that is axially deformed into a bell-shaped domed structure. Although the modal frequencies will clearly be strongly perturbed by such a transformation, the modal shapes will remain unchanged, as illustrated in Fig. 15.135, where m represents the number of radial nodal lines and n the number of nodal circles between the (fixed) centre and free edge.

The first example in Fig. 15.135 illustrates the rim displacements of the $(0, n)$ extensional modes. Although such "breathing" modes would be efficient sound radiators the energy involved in stretching the surfaces of the bell leads to very high modal frequencies, so that such modes are not strongly excited. Likewise, the torsional $(0, n)$–modes involve no motion perpendicular to the bell surfaces, and therefore generate a negligible amount of sound. With the bell rigidly supported at its top, the $m = 1$ swinging modes again involve large elastic strains and cannot be strongly excited. The first modes to produce a significant amount of sound are therefore the $m = 2$ and above flexural modes, which involve the transverse motions of the outer edges with negligible extension in circumferential length for small amplitude vibrations. When the wavelength of the flex-

Amazon
EX1013 Page 667

ural waves is much smaller than the overall curvature, the vibrational modes will be closely related to the flexural waves of a circular disc, with frequencies satisfying Chladni's generalised empirical law, $f_{mn} \sim c(m+2n)^p$, as confirmed by *Perrin* et al. [15.218].

The flexural modes involve radial displacements proportional to $\cos(m\phi)$. Continuity requires that there must also be a tangential displacement, such that $u + \partial v/\partial\phi = 0$, where u and v are the radial and tangential velocities respectively. Coupling to the tangential motion explains why it is possible to feed energy into a vibrating wine glass or the individual glass resonators of a glass harmonica (*Rossing* [15.201] Chap. 14), by rubbing a wetted finger around the rim. The excitation is very similar to the slip–stick mechanism used to excite the bowed string (Sect. 15.2.4).

Figure 15.136 illustrates a set of holographic measurements by *Rossing* ([15.201], Fig. 12.4), which is typical of most bell shapes. The $(m, 1)$ and $(m, 2)$ modes can immediatly be related to the (m, n) modes of a cupped disc. However, there is a distinct change in character for $n > 0$, with an additional node appearing close to the rim – referred to as a $(m, 1\#)$ mode. Some insight is provided by the finite element solutions for a typical large English church bell illustrated in Fig. 15.137 [15.221]. The three modes illustrated are very similar to the $(3, 0)$, $(3, 1)$ and $(3, 2)$ modes expected from the simple cupped disc model, except for the lowest frequency $(3, 0)$ mode,

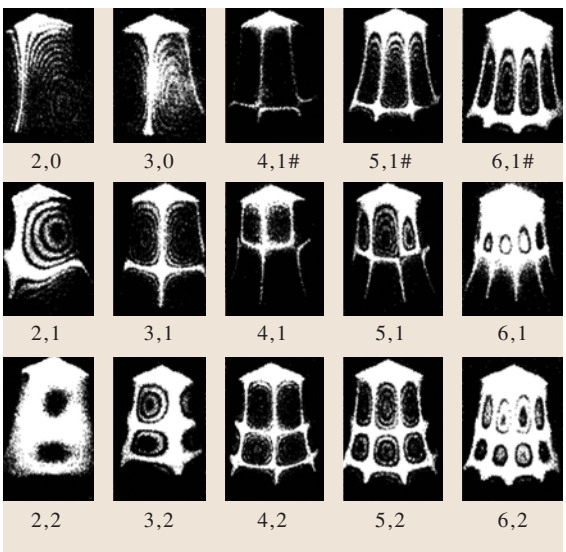


Fig. 15.136 Holographic interferograms and nomenclature for vibrational modes of a hand bell (after *Rossing* [15.201])

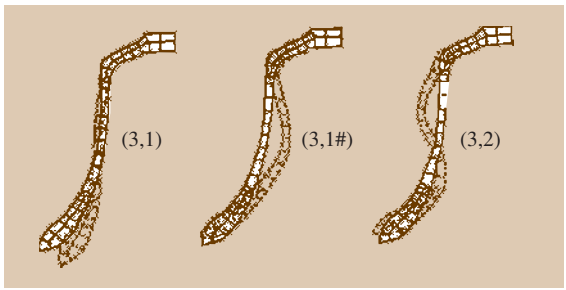


Fig. 15.137 Finite-element solutions for the lowest-order m = 3 modes of an English church bell (after *Perrin* et al. [15.219]) illustrating the (3,1), (3,1#) and (3,2) modes

in which the top surfaces of the bell move in antiphase with the rim, to give a nodal line about half-way along the length. Similarly, the anticipated $(4, 0)$, $(5, 0)$ and $(6, 0)$ modes of the handbell investigated by *Rossing* [15.201] acquire an additional nodal line close to the end rim denoted as $(4, 1\#)$, $(5, 1\#)$ and $(6, \#)$ modes. Such features can generally only be accounted for by detailed computational analysis.

Bell Tuning

Figure 15.138 shows the frequencies and associated modes of a well-tuned traditional church bell (*Rossing* and *Perrin* [15.220]), ordered into groups based on mode

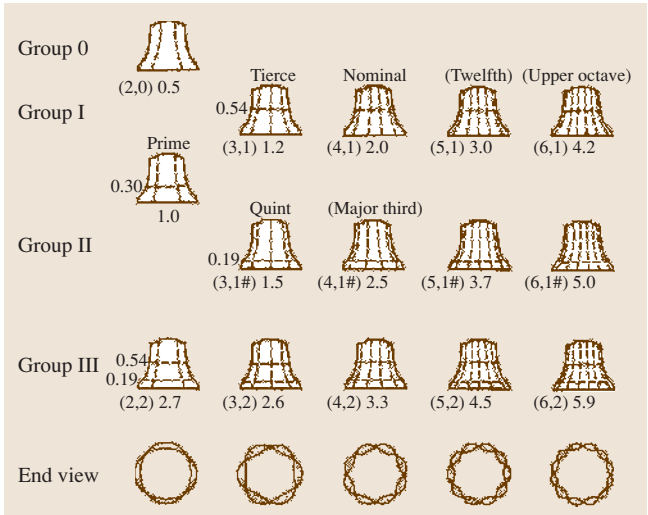


Fig. 15.138 Measured frequencies for a typical D5 church bell, indicating the relative frequencies of the observed mode, with the traditional names associated with such modes indicated (after *Rossing* and *Perrin* [15.220])

Amazon
EX1013 Page 668

shapes, which enable correlations to be made between bells of different shapes and sizes.

Fine-quality bells are carefully tuned by the bell maker, so that the lowest modes have harmonically related frequencies, thereby achieving a well-defined sense of pitch. This is achieved by selective thinning of the thickness of the bell on a very large lathe after its initial casting. In an ideally tuned bell, the principal modes are designated as the hum (2,0) mode, an octave below the prime or fundamental (2,1#) mode. A minor third above the prime (ratio 5:4) is the tierce (3,1) mode, a perfect fifth above that (ratio 3:2) is the quint or fifth (3,1#) mode, and an octave above is the nominal (4,1) mode. *Fletcher* and *Rossing* ([15.5], Table 21.1) compare these and higher modes with measured values for a particular bell.

The art of tuning the partials of bells to achieve a well-sounding note was initially developed in the seventeenth century in the low countries (what is now Holland and the northern parts of Germany and France). Many fine bells from this period by François and Pieter Hemony are still in use in carillons today (*Rossing* [15.201]). However, the art of bell tuning then appears to have been lost until the end of the 19th century, when it was rediscovered by Canon Arthur Simpson in England, following Lord Rayleigh's pioneering research on the sound of bells ([15.3] Vol. 1, Sect. 235).

Acoustic Radiation

The strongest-sounding partials of most bells are the group I $(m,1)$ modes, with a nodal circle approximately halfway up the bell (*Fletcher* and *Rossing* [15.5], Sect. 21.11). Such modes have $2m$ antinodal areas providing spatially alternating sound sources in antiphase. The radiation efficiency of such modes increases rapidly with size of bell and frequency. If the acoustic wavelength is much larger than the separation of such antinodes, the sound from such sources will tend to cancel out. However, above a crossover frequency, such that the velocity of sound is equal to that of the flexural waves on the surface of the bell, $v_{\text{flex}} = \sqrt{1.8 c_{\text{L}} h f}$, the spacing between the antinodes will exceed the wavelength in air and the modes will radiate more efficiently. For large church bells, this condition is satisfied for almost all but the very lowest partials, so almost all partials radiate sound rather efficiently. In contrast, hand bells with rather thin walls are significantly less efficient. There is also a small intensity of sound radiated axially at double the modal frequencies, from the induced fluctuations in the volume of air enclosed within the bell.

When a bell is struck, usually by a cast or wrought-iron, ball-shaped, clapper, the first sound heard, the strike note, contains contributions from the very large range of largely inharmonic partials of the bell. Nevertheless the listener can usually identify a pitch to the initial note, determined by the prominent partials excited. However, it is not always easy to attribute the pitch of a note to a particular octave, which is a common feature of many struck percussion instruments (e.g. notes on a xylophone, steeldrum and even timpani). The pitch of the strike note appears to be determined principally by the excited partials with frequencies in the ratios 2:3:4. The ear attributes the pitch to be that of the missing fundamental an octave below the lowest of the partials principally excited, which does not necessarily correspond to the pitch of the lowest partial excited, unless the bell is well tuned.

The majority of the higher partials decay rather quickly, with the long-term sound dominated by the hum note. For a 70 cm church bell, *Perrin* et al. [15.218] measured T_{60} decay times of 52 s for the (2,0) hum mode, 16 s for the (2.1#) and (3,1) prime and minor third modes, 6 s for the (4,1) octave and 3 s for the (4,1#) upper major third, with progressively shorter decay times

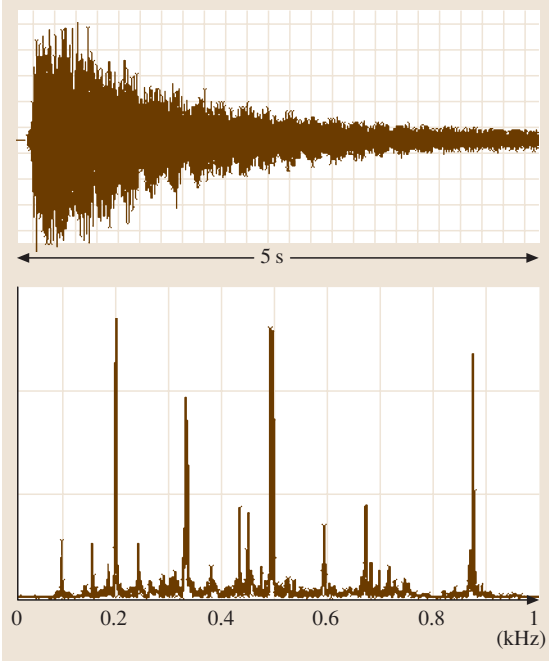


Fig. 15.139 Decaying vibrations and spectrum of Big Ben, London

Amazon
EX1013 Page 669

for the higher modes. Audio CD-ROM illustrates the synthesised sound of the above harmonic modes excited with equal amplitudes. This is followed by a synthesised bell with a major rather than a minor third partial. Such a bell has recently been realised (1999) based on FEA studies at the Technical University in Eindhoven in collaboration with the Eijsbouts Bell foundry. It produces a less mournful sound when played in the major scale music in a carillon. In both cases the synthesised sound reproduces the gentle sound of a bell rather well, though it lacks the initial clang that results from the higher-frequency inharmonic modes of a real bell.

As an example of the waveform and spectrum of a large bell, in Fig. 15.139 and audio signal CD-ROM, we illustrate the sound and rich spectrum of partials of one of the most famous large bells in the world, Big Ben, hung high above the Houses of Parliament in London. This bell is broadcast each day following the Westminster Chimes, marking the hours and end of broadcasting on the BBC each night for UK listeners.

Non-Axial Symmetry

Bell modes with $m > 1$ are doubly degenerate with orthogonal modes varying as $\cos m\phi$ and $\sin m\phi$ and nodal lines in the azimuthal directions that bisect each other. Any departures from axial symmetry will lift the degeneracy and give rise to a split pair of orthogonal modes. If both modes are excited together, the two frequencies will beat against each other, as already evident in the sound of Big Ben and in the sound (CD-ROM) of a slightly asymmetrical glass beaker with a pouring spout, which lifts the degeneracy of the otherwise axially symmetric modes.

Bells with strongly distorted or elliptical cross sections can have two completely different pitches, which can be sounded independently by beating at the antinodal positions of one set of modes mode and nodes of the other. Drums with quasi-elliptical cross sections will therefore sound a single note when struck at the narrowest and widest cross-sectional radii, and a second note, when struck at appropriate positions in between.

A dramatic example of axially asymmetric bells is provided by the 65 ancient bells from the tomb of Zeng Hou Yi found at Sui Xiang from around 433 BC. These bells from the second millennium BC are masterpieces of Chinese art and bell casting. They have oval cross sections and range from small hand bells to well over 1.5 m in height. When struck at different positions along the flattened surfaces, two quite distinct tones can be excited, which were designed to be about a major or minor third apart. For further details of these and other Chinese and other eastern bells see *Rossing* ([15.201] Chap. 13) and [15.222].

References

15.1 T. Levenson: *Measure for Measure: How Music and Science together have explored the Universe* (Oxford Univ. Press, Oxford 1997)

15.2 S. Hawkins (Ed.): *On the Shoulders of Giants* (Running, Philadelphia 2002)

15.3 Lord Rayleigh: *The Theory of Sound: I and II (1896)*, 2nd edn. (Dover, New York 1945), Reprint

15.4 S. Hawkins: *The Universe in a Nutshell* (Bantam, London 2001)

15.5 N.H. Fletcher, T.D. Rossing: *The Physics of Musical Instruments*, 2nd edn. (Springer, New York, Berlin 1998)

15.6 T.D. Rossing, F.R. Moore, P.A. Wheeler: *The Science of Sound*, 3rd edn. (Addison Wesley, San Francisco 2002)

15.7 D.M. Campbell, A. Myers, C.A. Greated: *Musical Instruments: History, Technology and Performance of Instruments of Western Music* (Oxford Univ. Press, Oxford 2006)

15.8 G. Bissinger: Contemporary generalised normal mode violin acoustics, Acustica **90**, 590–599 (2004)

15.9 A. Hirschberg, J. Gilbert, R. Msallam, A.P.J. Wijnands: Shock waves in trombones, J. Acoust. Soc. Am. **99**, 1754–1758 (1996)

15.10 T.J.W. Hill, B.E. Richardson, S.J. Richardson: *Modal Radiation from Classical Guitars: Experimental Measurements and Theoretical Predictions. Proc SMAC 03* (KTH, Stockholm 2003) pp. 129–132

15.11 L. Cremer: *The Physics of the Violin, section 11.2* (MIT Press, Mass 1984)

15.12 B. Britten: *Serenade for Tenor, Horn and Strings* (Hawkes & Son, London 1944)

15.13 J.M. Barbour: *Tuning and Temperament* (Michigan State College Press, East Lansing 1953)

15.14 National Instruments: *Windowing: Optimizing FFTs Using Windowing Functions* (National Instruments, Austin 2006), http://zone.ni.com/devzone/conceptd.nsf/

15.15 M.-P. Verge, A. Hirschberg: Turbulence noise in flue instruments. In: *Proc. Int. Symp. Mus. Acoust.*, SMAC 95 (IRCAM, Paris 1995) pp. 94–99

15.16 M.E. McIntyre, R.T. Schumacher, J. Woodhouse: Aperiodicity in bowed-string motion: On the dif-

Amazon
EX1013 Page 670

ferential slipping mechanism, Acustica **49**, 13–32 (1981)
15.17 J. Meyer: Zur klanglichen wirkung des streicher-vibratos, Acustica **76**, 283–291 (1992)
15.18 C.E. Gough: Measurement, modelling and synthesis of violin vibrato sounds, Acta Acustica **91**, 229–240 (2005)
15.19 E.I. Prame: Measurement of the vibrato rate of ten singers, J. Acoust. Soc. Am. **96**, 1979–1984 (1994)
15.20 J. Gilbert, L. Simon, J. Terroir: *Vibrato in single reed instruments, SMAC 03* (Royal Swedish Academy of Music, Stockholm 2003) pp. 271–273
15.21 D.W. Robinson, R.S. Dadson: A re-determination of the equal-loudness relations for pure tones, Brit. J. Appl. Phys. **7**, 166–181 (1956)
15.22 H.A. Fletcher, W.A. Munson: Loudness, its definition, measurement and calculation, J. Acoust. Soc. Am. **9**, 82–108 (1933)
15.23 B. Patterson: Musical dynamics, Sci. Am. **231**(5), 78–95 (1974)
15.24 B.C.J. Moore: *Psychology of Hearing* (Academic, London 1997)
15.25 A.J.M. Houtsma, T.D. Rossing, W.M. Wagenaars: *Auditory Demonstrations, Phillips compact disc 1126-061* (Acoust. Soc. Am., Sewickley 1989)
15.26 C.M. Hutchins (Ed.): *Musical Acoustics, Part I (Violin Family Components) and II (Violin Family Functions) Benchmark papers in Acoustics, Nos. 5 and 6* (Dowden Hutchinson, Ross, Stroudsburg 1975,1976)
15.27 C. Hutchins, V. Benade (Eds.): *Research Papers in Violin Acoustics 1975-1993, Vols. 1 and 2*, (Acoust. Soc. Am., Melville 1997)
15.28 CAS Newsletters (1994-1987), CAS Journals (1988-2004), www.catgutacoustical.org
15.29 L. Cremer: *The Physics of the Violin* (MIT Press, Cambridge 1984)
15.30 R. Midgley: *Musical Instruments of the World* (Paddington, New York 1978)
15.31 A. Baines: *European and American Musical Instruments* (Chancellor, London 1983)
15.32 G.G. Stokes: *On the communication of vibrations from a vibrating body to a surrounding gas* (Philos. Trans., London 1868)
15.33 C.E. Gough: The resonant response of a violin G-string and the excitation of the wolf-note, Acustica **44**(2), 673–684 (1980)
15.34 H.F. Helmholtz: *The Sensation of Tone (1877), translated by A.J. Ellis* (Dover, New York 1954)
15.35 C.V. Raman: On the mechanical theory of the vibrations of bowed strings and of musical instruments of then violin family, with experimental verification of the results: Part 1, Indian Assoc. Cultivation Sci. Bull. **15**, 1–158 (1918)
15.36 J.C. Schelling: The bowed string and player, J. Acoust. Soc. Am. **9**, 91–98 (1973)
15.37 S. Thwaites, N.H. Fletcher: Some notes on the clavichord, J. Acoust. Soc. Am. **69**, 1476–1483 (1981)
15.38 D.E. Hall: Piano string excitation in the case of a small hammer mass, J. Acoust. Soc. Am. **79**, 141–147 (1986)
15.39 D.E. Hall: Piano string excitation II: General solution for a hard narrow hammer, J. Acoust. Soc. Am. **81**, 535–546 (1987)
15.40 D.E. Hall: Piano string excitation III: General solution for a soft narrow hammer, J. Acoust. Soc. Am. **81**, 547–555 (1987)
15.41 D.E. Hall: Piano string excitation IV: Non-linear modelling, J. Acoust. Soc. Am. **92**, 95–105 (1992)
15.42 P.M. Morse, K.U. Ingard: *Theoretical Acoustics* (McGraw-Hill, New York 1968), Reprinted by Princeton Univ. Press, Princeton 1986
15.43 H. Fletcher: Normal vibration frequencies of a stiff piano string, J. Acoust. Soc. Am. **36**, 203–209 (1964)
15.44 E.L. Kent: Influence of irregular patterns in the inharmonicity of piano-tone partials upon tuning practice, Das Musikinstrument **31**, 1008–1013 (1982),
15.45 N.C. Pickering: *The Bowed String* (Bowed Instruments, Southampton 1991)
15.46 N.C. Pickering: Physical Properties of Violin Strings, J. Catgut Soc. **44**, 6–8 (1985), paper 21 in ref [15.42]
15.47 C. Vallette: The mechanics of vibrating strings. In: *Mechanics of musical instruments*, ed. by A. Hirschberg, J. Kergomard, G. Weinreich (Springer, Berlin, New York 1995) pp. 115–183
15.48 K. Legge, N.H. Fletcher: Nonlinear generation of missing modes on a vibrating string, J. Acoust. Soc. Am. **76**, 5–12 (1984)
15.49 C.E. Gough: The nonlinear free vibration of a damped elastic string, J. Acoust. Soc. Am. **75**, 1770–1776 (1984)
15.50 J. Miles: Resonant non-planar motion of a stretched string, J. Acoust. Soc. Am. **75**, 1505–1510 (1984)
15.51 R.J. Hanson, J.M. Anderson, H.K. Macomber: Measurement of nonlinear effects in a driven vibrating wire, J. Acoust. Soc. Am. **96**, 1549–1556 (1994)
15.52 R.J. Hanson, H.K. Macomber, A.C. Morrison, M.A. Boucher: Primarily nonlinear effects observed in a driven asymmetrical vibrating wire, J. Acoust. Soc. Am. **117**, 400–412 (2005)
15.53 J.A. Elliot: Intrinsic nonlinear effects in vibrating strings, Am. J. Phys. **48**, 478–480 (1980)
15.54 J. Woodhouse, P.M. Galluzzo: The bowed string as we know it today, Acustica **90**, 579–590 (2004)
15.55 F.A. Saunders: Recent work on violins, J. Acoust. Soc. Am. **25**, 491–498 (1953)
15.56 J.C. Schelling: The bowed string and the player, J. Acoust. Soc. Am. **53**, 26–41 (1973)
15.57 J.C. Schelling: The physics of the bowed string. In: *The Physics of Music*, ed. by C. Hutchins, Scientific American (W.H. Freeman & Co., San Fransisco 1978)
15.58 R.T. Schumacher: Measurements of some parameters of bowing, J. Acoust. Soc. Am. **96**, 1985–1998 (1994)

Amazon
EX1013 Page 671

15.59 F.G. Friedlander: On the oscillation of the bowed string, Proc. Cambridge Phil. Soc. **49**, 516–530 (1953)

15.60 M.E. McIntyre, J. Woodhouse: On the fundamentals of bowed string dynamics, Acustica **43**, 93–108 (1979)

15.61 J. Woodhouse: On the playability of violins, Part II: Minimum bow force and transients, Acustica **78**, 137–153 (1993)

15.62 M.E. McIntyre, R.T. Schumacher, J. Woodhouse: Aperiodicity in bowed-string motion: On the differential slipping mechanism, Acustica **49**, 13–32 (1981)

15.63 R.T. Schumacher: Oscillations of bowed strings, J. Acoust. Soc. Am. **43**, 109–120 (1979)

15.64 K. Guettler: On the creation of the Helmholtz motion in bowed strings, Acustica **88**, 2002 (2002)

15.65 P.M. Galluzzo: *On the Playability of stringed instruments, PhD Thesis* (Cambridge University, Cambridge 2003)

15.66 J.H. Smith, J. Woodhouse: The tribology of rosin, J. Mech. Phys. Solids **48**, 1633–1681 (2000)

15.67 W. Reinicke: *Dissertation, Institute for Technical Acoustics* (Technical University of Berlin, Berlin 1973)

15.68 J. Woodhouse, R.T. Schumacher, S. Garoff: Reconstruction of bowing point friction force in a bowed string, J. Acoust. Soc. Am. **108**, 357–368 (2000)

15.69 J.H. Smith: *Stick-slip vibration and its constitutive laws, PhD thesis* (Cambridge University, Cambridge 1990)

15.70 J. Woodhouse: Bowed string simulation using a thermal friction model, Acustica **89**, 355–368 (2003)

15.71 W. Reinicke, L. Cremer: Application of holographic interferometry to the bodies of sting instruments, J. Acoust. Soc. Am. **48**, 988–992 (1970)

15.72 J. Woodhouse: data kindly provided for this chapter

15.73 H. Dünnewald: Ein erweitertes Verfahren zur objektiven Bestimmung der Klangqualität von Violinen, Acustica **71**, 269–276 (1990), reprinted in English as Deduction of objective quality parameters on old and new violins, J. Catgut Acoust. Soc. 2nd Ser. 1(7) 1-5 (1991), included in Hutchins [15.27, Vol 1]

15.74 E.V. Jansson: Admittance measurements of 25 high quality violins, Acta Acustica **83**, 337–341 (1997)

15.75 E.V. Jansson: Violin frequency response – bridge mobility and bridge feet distance, Appl. Acoust. **65**, 1197–1205 (2004)

15.76 J. Woodhouse: On the "bridge hill" of the violin, Acustica **91**, 155–165 (2005)

15.77 J.A. Moral, E.V. Jansson: Eigenmodes, input admittance and the function of the violin, Acustica **50**, 329–337 (1982)

15.78 M. Hacklinger: Violin timbre and bridge frequency response, Acustica **39**, 324–330 (1978)

15.79 D. Gill: *The Book of the Violin*, ed. by D. Gill (Phaedon, Oxford 1984) p. 11, Prologue

15.80 C.E. Gough: The theory of string resonances on musical instruments, Acustica **49**, 124–141 (1981)

15.81 J. Woodhouse: On the synthesis of guitar plucks, Acustica **89**, 928–944 (2003)

15.82 J. Woodhouse: Plucked guitar transients: comparison of measurements and synthesis, Acustica **89**, 945–965 (2003)

15.83 G. Weinreich: Coupled piano strings, J. Acoust. Soc. Am. **62**, 1474–1484 (1977)

15.84 C.G.B. Baker, C.M. Thair, C.E. Gough: A photodetector for measuring resonances of violin strings, Acustica **44**, 70 (1980)

15.85 F. Savart: *Mémoires sur le construction des Instruments Ã Cordes et Ã Archet* (Deterville, Paris 1819), See also the brief summary and references to modern translations of Savart's research in Hutchins 15.26 Part 1, page 8

15.86 M.E. McIntyre, J. Woodhouse: On measuring the elastic and damping of orthotropic sheet materials, Acta. Metall. **36**, 1397–1416 (1988)

15.87 M.D. Waller: Vibrations of free rectangular plates, Proc. Phys. Soc. London B **62**, 277–285 (1949)

15.88 A.W. Leissa: *Vibration of Plates* (NASA SP-160, Washington 1993), Reprinted by Acoust. Soc. Am., Woodbury 1993

15.89 C. Hutchins: The Acoustics of Violin Plates, Sci. Am. **1**, 71–186 (1981)

15.90 B.E. Richardson, G.W. Roberts: *The adjustment of mode frequencies in guitars: a study by means of holographic interferometry and finite element analysis. SMAC 83* (Royal Swedish Academy of Music, Stockholm 1985) pp. 285–302

15.91 C. Hutchins: A rationale for BI-TRI octave plate tuning, Catgut Acoust. Soc. J. **1**(8), 36–39 (1991), Series II

15.92 E. Reissner: On axi-symmetrical vibrations of shallow spherical shells, Q. Appl. Math. **13**, 279–290 (1955)

15.93 K.D. Marshall: Modal analysis of a violin, J. Acoust. Soc. Am. **77**(2), 695–709 (1985)

15.94 G. Bissinger: Contemporary generalised normal mode violin acoustics, Acustica **90**, 590–599 (2004)

15.95 N.J.-J. Fang, O.E. Rodgers: Violin soundpost elastic vibration, Catgut Acoust. Soc. J. **2**(1), 39–40 (1992)

15.96 G.A. Knott: *A modal Analysis of the Violin using MSC.NASTRAM and PATRAN. MSc Thesis* (Naval Postgraduate School, Monterey 1987), Reproduced in Hutchins/Benade 15.27

15.97 J.P. Beldie: *Dissertation* (Technical University of Berlin, Berlin 1975), as described in Cremer 15.29 Sect. 10.4-5

15.98 J. Meyer: Die Abstimmung der Grundresonanzen von Guitarren, Das Musikinstrument **23**, 179–186

Amazon
EX1013 Page 672

(1974), English translation in J. Guitar Acoustics No.5, 19 (1982)
15.99 Q. Christensen: Qualitative models for low frequency guitar function, J. Guitar Acoust. **6**, 10–25 (1982)
15.100 T.D. Rossing, J. Popp, D. Polstein: *Acoustical Response of Guitars. SMAC 83* (Royal Swedish Academy of Music, Stockholm 1985) pp. 311–332
15.101 E.V. Jansson: On higher air modes in the violin, Catgut Acoust. Soc. Newsletter **19**, 13–15 (1973), Reprinted in Musical Acoustics, Part 2, ed. C.M.Hutchins (Dowden, Hutchinson & Ross, Stroudsberg 1976)
15.102 G. Derveaux, A. Chaigne, P. Joly, J. Bécache: Time-domain simulation of guitar: Model and method, J. Acoust. Soc. Am. **114**, 3368–3383 (2003)
15.103 G. Roberts, G. Finite: *Element analysis of the violin, PhD Thesis* (Cardiff University, Cardiff 1986), extract reproduced in Hutchins and Benade 15.27, pp.575–590
15.104 B.E. Richardson, G.W. Roberts: *The adjustment of mode frequencies in guitars: A study by means of holographic and finite element analysis. Proc. SMAC. 83* (Royal Swedish Academy of Music, Stockholm 1985) pp. 285–302
15.105 G. Derveaux, A. Chaigne, P. Joly, J. Bécache: *Numerical simulation of the acoustic guitar* (INRIA, Rocquencourt 2003), from http://www.inria.fr/multimedia/Videotheque-fra.html
15.106 G. Bissinger: A unified materials-normal mode approach to violin acoustics, Acustica **91**, 214–228 (2005)
15.107 M. Schleske: On making "tonal copies" of a violin, J. Catgut Acoust. Soc. **3**(2), 18–28 (1996)
15.108 L.S. Morset: *A low-cost pc-based tool for violin acoustics measurements, ISMA 2001* (Fondazione Scuola Di San Giorgio, Venice 2001) pp. 627–630
15.109 H.O. Saldner, N.-E. Molin, E.V. Jansson: Vibrational modes of the violin forced via the bridge and action of the soundpost, J. Acoust. Soc. Am. **100**, 1168–1177 (1996)
15.110 E.V. Jansson, N.-E. Molin, H. Sundin: Resonances of a violin body studied by hologram interferometry and acoustical methods, Phys. Scipta **2**, 243–256 (1970)
15.111 N.-E. Molin, A.O. Wählin, E.V. Jansson: Transient response of the violin, J. Acoust. Soc. Am. **88**, 2479–2481 (1990)
15.112 N.-E. Molin, A.O. Wählin, E.V. Jansson: Transient response of the violin revisited, J. Acoust. Soc. Am. **90**, 2192–2195 (1991)
15.113 J. Curtin: Innovation in violin making, Proc. Int. Symp. Musical Acoustics, CSA-CAS **1**, 11–16 (1998)
15.114 J. Meyer: Directivity of the bowed string instruments and its effect on orchestral sound in concert halls, J. Acoust. Soc. Am. **51**, 1994–2009 (1972)
15.115 G. Weinreich, E.B. Arnold: Method for measuring acoustic fields, J. Acoust. Soc. Am. **68**, 404–411 (1982)
15.116 G. Weinreich: Sound hole sum rule and dipole moment of the violin, J. Acoust. Soc. Am. **77**, 710–718 (1985)
15.117 T.J.W. Hill, B.E. Richardson, S.J. Richardson: Acoustical parameters for the characterisation of the classical guitar, Acustica **89**, 335–348 (2003)
15.118 G. Bissinger, A. Gregorian: Relating normal mode properties of violins to overall quality signature modes, Catgut Acoust. Soc. J. **4**(8), 37–45 (2003)
15.119 E.V. Jansson: Admittance measurements of 25 high quality violins, Acustica **83**, 337–341 (1997)
15.120 G. Weinreich: Directional tone colour, J. Acoust. Soc. Am. **101**, 2338–2346 (1997)
15.121 J. Meyer: Zur klangichen Wirkung des Streicher-Vibratos, Acustica **76**, 283–291 (1992)
15.122 H. Fletcher, L.C. Sanders: Quality of violin vibrato tones, J. Acoust. Soc. Am. **41**, 1534–1544 (1967)
15.123 M.V. Matthews, K. Kohut: Electronic simulation of violin resonances, J. Acoust. Soc. Am. **53**, 1620–1626 (1973)
15.124 J. Nagyvary: Modern science and the classical violin, Chem. Intel. **2**, 24–31 (1996)
15.125 L. Burckle, H. Grissino-Mayer: Stradivari, violins, tree rings and the Maunder minimum, Dendrichronologia **21**, 41–45 (2003)
15.126 C.M. Hutchins: A 30-year experiment in the acoustical and musical development of violin family instruments, J. Acoust. Soc. Am. **92**, 639–650 (1992),
15.127 H.L.F. Helmholtz: *On the Sensations of Tone*, 4th edn. (Dover, New York 1954), trans. by A.J. Ellis
15.128 H. Bouasse: *Instruments á Vent* (Delagrave, Paris 1929)
15.129 M. Campbell, C. Greated: *The Musicians Guide to Acoustics* (Oxford Univ. Press, Oxford 1987)
15.130 C.J. Nederveen: *Acoustical Aspects of Woodwind Instruments* (Northern Illinois Univ. Press, DeKalb 1998)
15.131 A. Hirschberg, J. Kergomard, G. Weinreich: *Mechanics of Musical Instruments* (Springer, Berlin, New York 1995)
15.132 J. Backus: *The Acoustical Foundations of Music*, 2nd edn. (Norton, New York 1977)
15.133 A.H. Benade: *Fundamentals of Musical Acoustics* (Oxford Univ. Press, Oxford 1975)
15.134 D.M. Campbell (Ed.): Special Issue on Musical Wind Instruments Acoustics, Acustica **86**(4), 599–755 (2000)
15.135 P.M. Moorse, K.U. Ingard: *Theoretical Acoustics* (McGraw-Hill, New York 1968), reprinted by Princeton Univ. Press, Princeton (1986)
15.136 L.L. Beranek: *Acoustics* (McGraw Hill, New York 1954) pp. 91–115, reprinted by Acoust. Soc. Am. (1986)

Amazon
EX1013 Page 673

15.137 H. Levine, L. Schwinger: On the radiation of sound from an unflanged pipe, Phys. Rev. **73**, 383–406 (1948)

15.138 R.D. Ayers, L.J. Eliason, D. Mahrgereth: The conical bore in musical acoustics, Am. J. Phys. **53**, 528–527 (1985)

15.139 H.F. Olson: *Acoustic Engineering* (Van Nostrand-Reinhold, Princeton 1957) pp. 88–123

15.140 L.E. Kinsler, A.R. Frey, A.B. Coppens, J.V. Sanders: *Fundamentals of Acoustics* (Wiley, New York 1982), Fig. 14.19

15.141 A.G. Webster: Acoustical impedance, and the theory of horns and the phonograph, Proc. Nat. Acad. Sci. (US) **5**, 275–282 (1919)

15.142 C.J. Nederveen: *Acoustical Aspects of Woodwind Instruments* (Knuf, Amsterdam 1969) p. 60

15.143 D.H. Keefe, A.H. Benade: Wave propagation in strongly curved ducts, J. Acoust. Soc. Am. **74**, 320–332 (1983)

15.144 R. Caussé, J. Kergomard, X. Luxton: Input impedance of brass instruments – Comparison between experiment and numerical models, J. Acoust. Soc. Am. **75**, 241–254 (1984)

15.145 N.H. Fletcher, R.K. Silk, L.M. Douglas: Acoustic admittance of air-driven reed generators, Acustica **50**, 155–159 (1982)

15.146 A. Almeida, C. Verges, R. Caussé: Quasistatic nonlinear characteristics of double-reed instruments, J. Acoust. Soc. Am. **121**, 536–546 (2007)

15.147 S. Ollivier, J.-P. Dalmont: *Experimental investigation of clarinet reed operation*, *SMAC 03* (Royal Swedish Academy of Music, Stockholm 2003) pp. 283–289

15.148 J.-P. Dalmont, J. Gilbert, S. Ollivier: Nonlinear characteristics of single-reed instruments: Quasistatic flow and reed opening measurements, J. Acoust. Soc. Am. **114**, 2253–2262 (2003)

15.149 J. Backus, C.J. Nederveen: *Acoustical Aspects of Woodwind Instruments* (Knuf, Amsterdam 1969) pp. 28–37

15.150 S. Ollivier, J.-P. Dalmont: *Experimental investigation of clarinet reed operation*, *Proc. SMAC 03* (Royal Swedish Academy of Music, Stockholm 2003) pp. 283–289

15.151 A.P.J. Wijnands, A. Hirschberg: Effect of pipeneck downstream of a double reed. In: *Proc. Int. Symp. Musical Acoustics* (IRCAM, Paris 1995) pp. 148–151

15.152 N.H. Fletcher: Excitation mechanisms in woodwind and brass instruments, Acustica **43**, 63–72 (1979), erratum Acustica 50, 155–159 (1982)

15.153 P.D. Koopman, C.D. Hanzelka, J.P. Cottingham: Frequency and amplitude of vibration of reeds from American reed organs as a function of pressure, J. Acoust. Soc. Am. **99**, 2506 (1996)

15.154 J.-P. Dalmont, J. Gilbert, J. Kergomard: Reed instruments, from small to large amplitude periodic oscillations and the Helmholtz motion analogy, Acustica **86**, 671–684 (2000)

15.155 J. Backus: Vibrations of the reed and air column of the clarinet, J. Acoust. Soc. Am. **33**, 806–809 (1961)

15.156 N.H. Fletcher: Nonlinear theory of musical wind instruments, Appl. Acoust. **30**, 85–115 (1990)

15.157 R.T. Schumacher: Self-sustained oscillations of a clarinet; an integral equation approach, Acustica **40**, 298–309 (1978)

15.158 R.T. Schumacher: Ab initio calculations of the oscillations of a clarinet, Acustica **48**, 71–85 (1981)

15.159 J. Gilbert, J. Kergomard, E. Ngoya: Calculation of steady state oscillations of a clarinet using the harmonic balance technique, J. Acoust. Soc. Am. **86**, 35–41 (1989)

15.160 D.M. Campbell: Nonlinear dynamics of musical reed and brass wind instruments, Contemp. Phys. **40**, 415–431 (1999)

15.161 N. Gand, J. Gilbert, F. Lalöe: Oscillation threshold of wood-wind instruments, Acta Acustica **1**, 137–151 (1997)

15.162 J.-P. Dalmont, J. Kergomard: Elementary model and experiments for the Helmholtz motion of single reed wind instruments. In: *Proc. Int. Symp. Mus. Acoust.* (IRCAM, Paris 1995) pp. 115–120

15.163 A.H. Benade: Air column reed and player's windway interaction in musical instruments. In: *Vocal Fold Physiology, Biomechanics, Acoustics and Phonatory Control*, ed. by I.R. Titze, R.C. Scherer (Denver Centre for the Performing Arts, Denver 1985) pp. 425–452

15.164 G.P. Scavone: *Modelling vocal tract influence in reed wind instruments*, *Proc. SMAC 03* (Royal Swedish Academy of Music, Stockholm 2003) pp. 291–294

15.165 N.H. Fletcher: Mode locking in non-linearly excited inharmonic musical oscillators, J. Acoust. Soc. Am. **64**, 1566–1569 (1978)

15.166 D. Ayers: *Basic tests for models of the lip reed*, *ISMA 2001* (Fondazione Scuola Di San Giorgio, Venice 2001) pp. 83–86

15.167 S.J. Elliot, J.M. Bowsher: Regeneration in brass wind instruments, J. Sound Vibrat. **83**, 181–207 (1982)

15.168 F.C. Chen, G. Weinreich: Nature of the lip-reed, J. Acoust. Soc. Am. **99**, 1227–1233 (1996)

15.169 S. Adachi, M. Sato: Trumpet sound simulation using a two-dimensional lip vibration model, J. Acoust. Soc. Am. **99**, 1200–1209 (1996)

15.170 I.R. Titze: The physics of small-amplitude oscillation of the vocal folds, J. Acoust. Soc. Am. **83**, 1536–1552 (1988)

15.171 D.C. Copley, W.J. Strong: A stroboscopic study of lip vibrations in a trombone, J. Acoust. Soc. Am. **99**, 1219–1226 (1996)

15.172 S. Yoshikawa, Y. Muto: *Brass player's skill and the associated li-p wave propagation*, ISMA 2001 (Fondazione Scuola Di San Giorgio, Venice 2002) pp. 91–949

Amazon
EX1013 Page 674

15.173 J. Gilbert, S. Ponthus, J.F. Petoit: Artificial buzzing lips and brass instruments: experimental results, J. Acoust. Soc. Am. **104**, 1627–1632m (1998)

15.174 J. Cullen, J.A. Gilbert, D.M. Campbell: Brass instruments: linear stability analysis and experiments with an artificial mouth, Acustica **86**, 704–724 (2000)

15.175 N.H. Fletcher: Nonlinear theory of musical wind instruments, Appl. Acoust. **30**, 85–115 (1990)

15.176 T.H. Long: The performance of cup-mouthpiece instruments, J. Acoust. Soc. Am. **19**, 892–901 (1947)

15.177 A. Hirschberg, J. Gilbert, R. Msallam, A.P.J. Wijnands: Shock waves in trombones, J. Acoust. Soc. Am. **99**, 1754–1758 (1996)

15.178 M.E. McIntyre, R.T. Schumacher, J. Woodhouse: On the oscillations of musical instruments, J. Acoust. Soc. Am. **74**, 1325–1345 (1983)

15.179 R.D. Ayers: Impulse responses for feedback to the driver of a musical wind instrument, J. Acoust. Soc. Am. **100**, 1190–1198 (1996)

15.180 B. Fabre, A. Hirschberg: Physical modelling of flue instruments: a review of lumped models, Acustica **86**, 599–610 (2000)

15.181 N.H. Fletcher: Sound production in organ pipes, J. Acoust. Soc. Am. **60**, 926–936 (1976)

15.182 P. Savic: On acoustically effective vortex motions in gaseous jets, Philos. Mag. **32**, 287–252 (1941)

15.183 J.W. Coltman: Jet drive mechanism in edge tones and organ pipes, J. Acoust. Soc. Am. **60**, 723–733 (1976)

15.184 S. Adachi: *CFD analysis of air jet deflection-comparison with Nolle's measurements, SMAC 03* (Royal Swedish Academy of Music, Stockholm 2003) pp. 313–319

15.185 A.W. Nolle, Sinuous instability of a planar air jet: Propagation parameters and acoustic excitation, J. Acoust. Soc. Am. **103**, 3690–3705 (1998)

15.186 L. Cremer, H. Ising: Die selbsterregten Schwingungen von Orgelpfeifen, Acustica **19**, 143–153 (1967)

15.187 J.W. Coltman: Sounding mechanism of the flute and organ pipe, J. Acoust. Soc. Am. **44**, 983–992 (1968)

15.188 M. Raffel, C. Willert, J. Kompenhans: *Particle Image Velocimetry – A Practical Guide* (Springer, Berlin, Heidelberg 1998)

15.189 C. Ségoufin, B. Fabre, M.P. Verge, A. Hirschberg, A.P.J. Wijnands: Experimental study of the influence of mouth geometry on sound production in recorder-like instruments: windway length and chamfers, Acta Acustica **86**, 649–661 (2000)

15.190 S. Thwaites, N.H. Fletcher: Wave propagation on turbulent jets, Acustica **45**, 175–179 (1980)

15.191 S. Thwaites, N.H. Fletcher: Wave propagation of turbulent jets: II, Acustica **51**, 44–49 (1982)

15.192 B. Fabre, A. Hirschberg: *From sound synthesis to instrument making: an overview of recent researches on woodwinds, SMAC 03* (Royal Swedish Academy of Music, Stockholm 2003) pp. 239–242

15.193 N.H. Fletcher: Jet-drive mechanism in organ pipes, J. Acoust. Soc. Am. **60**, 481–483 (1976)

15.194 S.A. Elder: On the mechanism of sound production in organ pipes, J. Acoust. Soc. Am. **54**, 1554–1564 (1973)

15.195 R.C. Chanaud: Aerodynamic whistles, Sci. Am. **222**(1), 40–46 (1970)

15.196 T.A. Wilson, G.S. Beavers, M.A. DeCoster, D.K. Holger, M.D. Regenfuss: Experiments on the fluid mechanics of whistling, J. Acoust. Soc. Am. **50**, 366–372 (1971)

15.197 D.K. Holger, T. Wilson, G. Beavers: Fluid mechanics of the edge-tone, J. Acoust. Soc. Am. **62**, 1116–1128 (1977)

15.198 S. Yoshikawa: Jet wave amplification in organ pipes, J. Acoust. Soc. Am. **103**, 2706–2717 (1998)

15.199 B. Fabre, A. Hirschberg, A.P.J. Wijnands: Vortex shedding in steady oscillations of a flue organ pipe, Acustica – Acta Acustica **82**, 877–883 (1996)

15.200 A. Baines: *European and American Musical Instruments* (Chancellor, London 1983)

15.201 T.D. Rossing: *Science of Percussion Instruments* (World Scientific, Singapore 2000)

15.202 J. Blades: *Percussion Instruments and Their History* (Faber, Faber, London 1974)

15.203 P. Boulez: *Notations I–IV* (Universal Edition, Vienna 1945, 78, 84)

15.204 M. Bertsch: *Vibration patterns and sound analysis of the Viennese timpani. ISMA 2001* (Fondazione Scuola Di San Giorgio, Venice 2001)

15.205 C.V. Raman: The Indian Musical Drum, Proc. Indian Acad. Sci. A1 179 (1934). In: *Reprinted in Musical Acoustics Selected Reprints*, ed. by T.D. Rossing (Am. Assoc. Phys. Teach., College Park 1988)

15.206 T.D. Rossing, W.A. Sykes: Acoustics of indian drums, Percuss. Notes **19**(3), 58 (1982)

15.207 B.S. Ramakrishna, M.M. Sondhi: Vibrations of indian musical drums regarded as composite membranes, J. Acoust. Soc. Am. **26**, 523 (1954)

15.208 S. De: Experimental study of the vibrational characteristics of a loaded kettledrum, Acustica **40**, 206 (1978)

15.209 H. Zhao: *Acoustics of Snare Drums: An Experimental Study of the Modes of Vibration, Mode Coupling and Sound Radiation Pattern, M.S. thesis* (Northern Illinois Univ., DeKalb 1990)

15.210 C.V. Raman: On some Indian stringed instruments, Proc. Indian Assoc. Adv. Sci. **7**, 29–33 (1922)

15.211 T.D. Rossing, I. Bork, H. Zhao, D. Fystrom: Acoustics of snare drums, J. Acoust. Soc. Am. **92**, 84–94 (1992)

15.212 T.D. Rossing: Acoustics of percussion instruments: Part I, Phys. Teacher **14**, 546–556 (1976)

15.213 T.D. Rossing: Chladni's law for vibrating plates, Am. J. Phys. **50**, 271–274 (1982)

Amazon
EX1013 Page 675

15.214 N.H. Fletcher: Nonlinear frequency shifts in quasi-spherical-cap shells: Pitch glide in Chinese Gongs, J. Acoust. Soc. Am. **78**, 2069 (1985)
15.215 P.L. Grossman, B. Koplik, Y.-Y. Yu: Nonlinear vibration of shallow spherical shells, J. Appl. Mech. **36**, 451–458 (1969)
15.216 K.A. Legge, N.H. Fletcher: Non-linear mode coupling in symmetrically kinked bars, J. Sound Vib. **118**, 23–34 (1987)
15.217 S. Schedin, P.O. Gren, T.D. Rossing: Transient wave response of a cymbal using double pulsed TV holography, J. Acoust. Soc. Am. **103**, 1217–1220 (1998)
15.218 R. Perrin, T. Charnley, H. Banu, T.D. Rossing: Chladni's law and the modern English church bell, J. Sound Vib. **102**, 11–19 (1985)
15.219 R. Perrin, T. Charnley, J. de Pont: Normal modes of the modern English church bell, J. Sound Vib. **102**, 28048 (1983)
15.220 T.D. Rossing, R. Perrin: Vibration of bells, Appl. Acoust. **20**, 41–70 (1988)
15.221 R. Perrin, T. Charnley, J. de Pont: Normal modes of the modern English church bell, J. Sound Vibr. **102**, 28048 (1983)
15.222 T.D. Rossing, D.S. Hampton, B.E. Richardson, H.J. Sathoff: Vibrational modes of Chinese two-tone bells, J. Acoust. Soc. Am. **83**, 369–373 (1988)

Amazon
EX1013 Page 676

16. The Human Voice in Speech and Singing

This chapter describes various aspects of the human voice as a means of communication in speech and singing. From the point of view of function, vocal sounds can be regarded as the end result of a three stage process: (1) the compression of air in the respiratory system, which produces an exhalatory airstream, (2) the vibrating vocal folds' transformation of this air stream to an intermittent or pulsating air stream, which is a complex tone, referred to as the voice source, and (3) the filtering of this complex tone in the vocal tract resonator. The main function of the respiratory system is to generate an overpressure of air under the glottis, or a subglottal pressure. Section 16.1 describes different aspects of the respiratory system of significance to speech and singing, including lung volume ranges, subglottal pressures, and how this pressure is affected by the ever-varying recoil forces. The complex tone generated when the air stream from the lungs passes the vibrating vocal folds can be varied in at least three dimensions: fundamental frequency, amplitude and spectrum. Section 16.2 describes how these properties of the voice source are affected by the subglottal pressure, the length and stiffness of the vocal folds and how firmly the vocal folds are adducted. Section 16.3 gives an account of the vocal tract filter, how its form determines the frequencies of its resonances, and Sect. 16.4 gives an account for how these resonance frequencies or formants shape the vocal sounds by imposing spectrum peaks separated by spectrum valleys, and how the frequencies of these peaks determine vowel and voice qualities. The remaining sections of the chapter describe various aspects of the acoustic signals used for vocal communication in speech and singing. The syllable structure is discussed in Sect. 16.5, the closely related aspects of rhythmicity and timing in speech and singing is described in Sect. 16.6, and pitch and rhythm aspects in Sect. 16.7. The impressive control of all these acoustic characteristics of vocal signals is discussed in Sect. 16.8, while Sect. 16.9 considers expressive aspects of vocal communication.

16.1 **Breathing** 669
16.2 **The Glottal Sound Source** 676
16.3 **The Vocal Tract Filter** 682
16.4 **Articulatory Processes, Vowels and Consonants** 687
16.5 **The Syllable** 695
16.6 **Rhythm and Timing** 699
16.7 **Prosody and Speech Dynamics** 701
16.8 **Control of Sound in Speech and Singing** 703
16.9 **The Expressive Power of the Human Voice** 706
References 706

16.1 Breathing

The process of breathing depends both on mechanical and muscular forces (Fig. 16.1).

During *inspiration* the volume of the chest cavity is expanded and air rushes into the lungs. This happens mainly because of the contraction of the *external intercostals* and the *diaphragm*. The external intercostal muscles raise the ribs. The diaphragm which is the dome-shaped muscle located below the lungs, flattens on contraction and thus lowers the floor of the thoracic cavity.

The respiratory structures form an elastic mechanical system that produces expiratory or inspiratory subglottal pressures, depending on the size of the lung volume, Fig. 16.2. Thus, exhalation and inhalation will always produce forces whose effect is to move the ribs and the lungs back to their resting state, often referred

Amazon
EX1013 Page 677

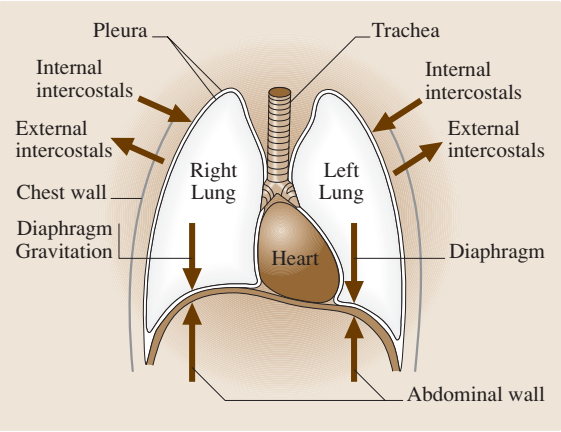


Fig. 16.1 Directions of the muscular and elastic forces. The direction of the gravitation force is applicable to the upright position

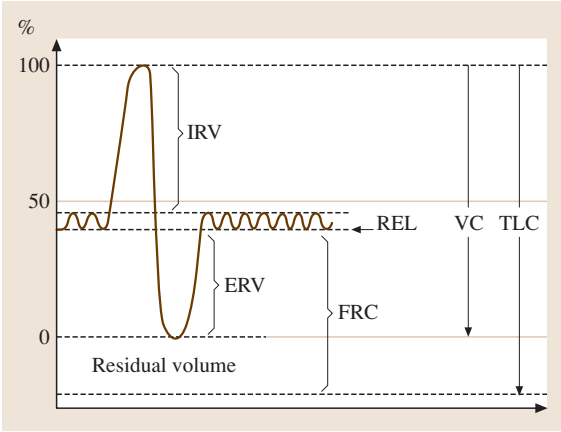


Fig. 16.3 Definition of the various terms for lung volumes. The graph illustrates the lung volume changes during quiet breathing interrupted by a maximal inhalation followed by a maximal exhalation. VC is the vital capacity, TLC is the total lung capacity, IRV and ERV are the inspiratory and expiratory reserve volume, REL is the resting expiratory level, FRC is the functional residual capacity

to as the resting expiratory level (REL). The deeper the breath, the greater this force of *elastic recoil*. This component plays a significant part in pushing air out of the lungs, both in speech and singing, and especially at large lung volumes. The elasticity forces originate both from the rib cage and the lungs. As illustrated in Fig. 16.2, the rib cage produces an exhalatory force at high lung volumes and an inhalatory force at low lung volumes, and the lungs always exert an exhalatory force. As a consequence, activation of inspiratory muscles is needed for producing a low subglottal pressure, e.g. for singing a soft (pianissimo) tone, at high lung volume. Conversely, activation of expiratory muscles is needed for producing a high subglottal pressure, e.g. for singing a loud (fortissimo) tone, at low lung volume.

In addition to mechanical factors, exhaling may involve the activity of the *internal intercostals* and the *abdominal muscles*. Contraction of the former has the effect of lowering the ribs and thus compressing the chest cavity. Activating the abdominal muscles generates upward forces that also contribute towards reducing the volume of the rib cage and the lungs. The function of these muscles is thus expiratory.

Another significant factor is gravity whose role depends on body posture. In an upright position, the diaphragm and adjacent structures tend to be pulled down, increasing the volume of the thoracic cavity. In this sit-

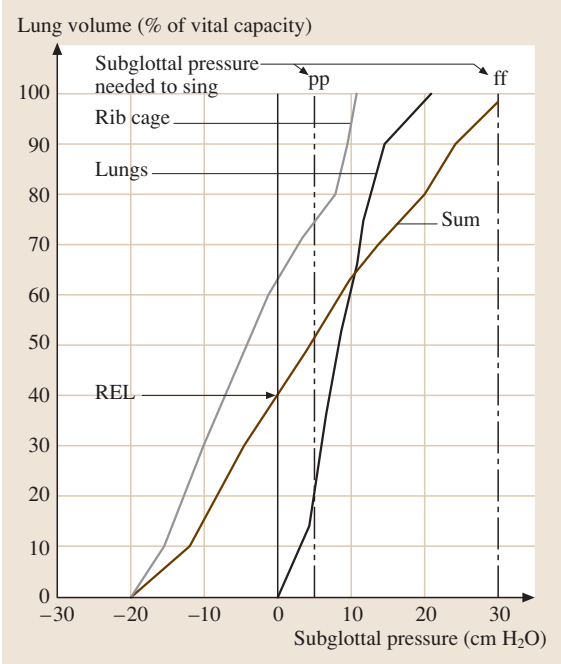


Fig. 16.2 Subglottal pressures produced at different lung volumes in a subject by the recoil forces of rib cage and lungs. The resting expiratory level (REL) is the lung volume at which the inhalatory and the exhalatory recoil forces are equal. The *thin* and *heavy chain-dashed lines* represent subglottal pressures typically needed for soft and very loud phonation

Amazon
EX1013 Page 678

uation, the effect of gravity is inspiratory. In contrast, in the supine position the diaphragm tends to get pushed up into the rib cage, and expiration is promoted [16.1].

The total air volume that is contained in a maximally expanded rib cage is called the *total lung capacity* (TLC in Fig. 16.3). After maximum exhalation a small air volume, the *residual volume*, is still left in the airways. The greatest air volume that can be exhaled after a maximum inhalation is called the *vital capacity* (VC) and thus equals the difference between the TLC and the residual volume. The lung volume at which the exhalatory and inhalatory recoil forces are equal, or REL, is reached after a relaxed sigh, see Figs. 16.2 and 16.3. During tidal breathing inhalation is initiated from REL, so that inhalation is active resulting from an activation of inspiratory muscles, and exhalation is passive, produced by the recoil forces. In tidal breathing only some 10% of VC is inhaled, such that a great portion of the VC, the *inspiratory reserve volume*, is left. The air volume between REL and the residual volume is called the *expiratory reserve volume*.

VC varies depending on age, body height and gender. At the age of about 20 years, an adult female has a vital capacity of 3–3.6 l depending on body height, and for males the corresponding values are about 4–5.5 l.

Experimental data [16.4] show that, during tidal breathing, lung volume variations are characterized by a regular quasi-sinusoidal pattern with alternating segments of inspiration and expiration of roughly equal duration (Fig. 16.4). In speech and singing, the pattern is transformed. Inspirations become more rapid and expiration occurs at a slow and relatively steady rate. Increasing vocal loudness raises the amplitude of the lung volume records but leaves its shape relatively unchanged.

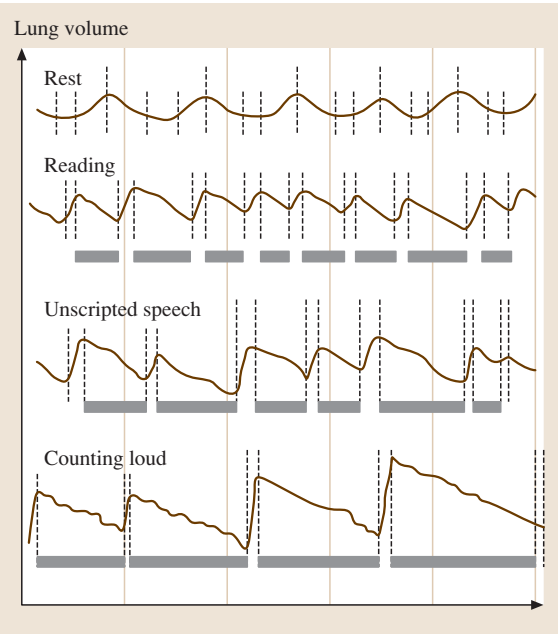


Fig. 16.4 Examples of speech breathing

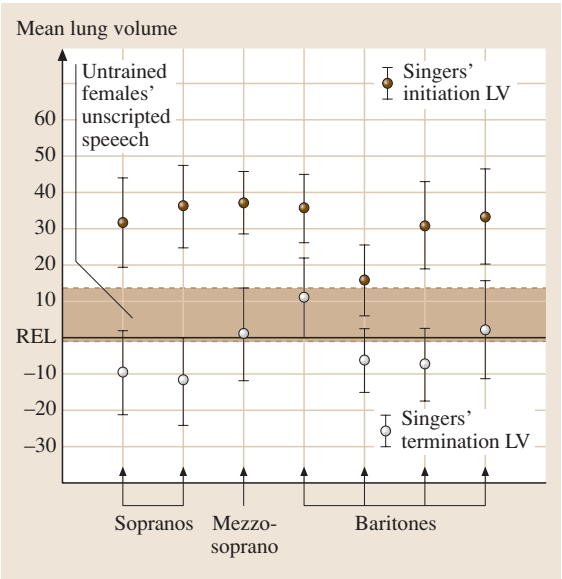


Fig. 16.5 Lung volume averages used in speech and operatic singing expressed as percentages of vital capacity relative to the resting expiratory level REL. The shaded band represents the mean lung volume range observed in untrained female voices' unscripted speech (after [16.2]). The filled and open symbols show the corresponding measures for professional opera singers of the indicated classifications when performing opera arias according to *Thomasson* [16.3]. The bars represent ± one SD

Figure 16.5 shows mean lung volumes used by professional singers when singing well-rehearsed songs. The darker band represents the mean lung volume range observed in spontaneous speech [16.2]. The lung volumes of conversational speech are similar to those in tidal breathing. Loud speech shows greater air consumption and thus higher volumes (Fig. 16.4). Breath groups in speech typically last for 3–5 seconds and are terminated when lung volumes approach the relaxation expiratory level REL, as illustrated in Fig. 16.5. Thus, in phonation, lung volumes below REL are mostly avoided.

Amazon
EX1013 Page 679

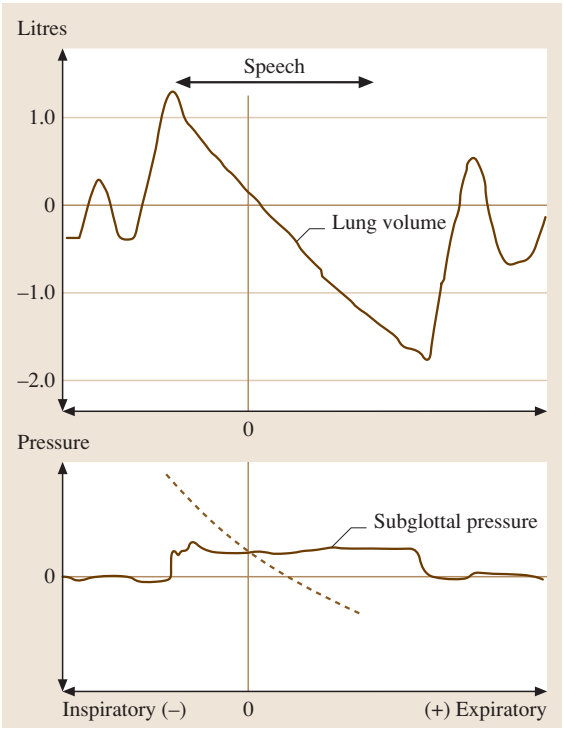


Fig. 16.6 The records show lung volume (relative to the mid-respiratory level), subglottal (esophageal) pressure and stylized muscular activity for speaker counting from one to 32 at a conversational vocal effort. (After Draper et al. [16.5]) . To the *left* of the vertical line recoil forces are strongly expiratory, to the *right* they are inspiratory. *Arrows* have been added to the *x*-axis to summarize the original EMG measurements which indicate that the recoil forces are balanced by muscular activity (EMG = electromyography, measurement of the electrical activity of muscles). To the *left of the vertical line* (as indicating by *left-pointing arrow*) the net muscular force is inspiratory, to the *right* it is expiratory (*right-pointing arrow*). To keep loudness constant the talker maintains a stable subglottal pressure and recruits muscles according to the current value of the lung volume. This behavior exemplifies the phenomenon known as *motor equivalence*

In singing, breath groups tend to be about twice as long or more, and air consumption is typically much greater than in conversational speech. Mostly they are terminated close to the relaxation expiratory level, as in speech, but sometimes extend into the *expiratory reserve volume* as illustrated in Fig. 16.5. This implies that, in singing, breath groups typically start at much higher lung volumes than in speech. This use of high lung volumes implies that singers have to deal with much greater recoil forces than in speech.

Figure 16.6 replots, in stylized form, a diagram published by Ladefoged et al. [16.7]. It summarizes measurements of lung volume and subglottal pressure, henceforth referred to as Ps, recorded from a speaker asked to take a deep breath and then to start counting. The dashed line intersecting the Ps record represents the relaxation pressure. This line tells us that elastic recoil forces are strongly expiratory to the left of the vertical line. To the right they produce a pressure lower than the target pressure and eventually an inspiratory pressure. The Ps curve remains reasonably flat throughout the entire utterance.

The question arises: how can this relative constancy be achieved despite the continually changing contribution of the relaxation forces? In a recent replication of this classical work [16.8], various criticisms of the original study are addressed but basically the original answer to this question is restated: the motor system adapts to the external goal of keeping the Ps fairly constant (Fig. 16.6). Initially, when recoil forces are strong, muscle activity is predominantly in the *inspiratory* muscles such as the *diaphragm* and the *external intercostals*. Gradually, as recoil forces decline, the *ex-*

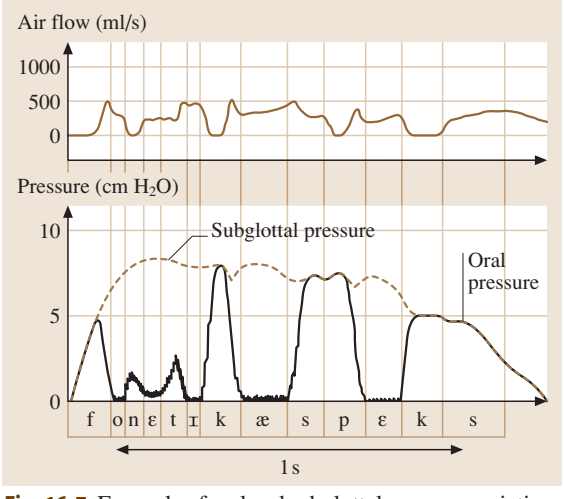


Fig. 16.7 Example of oral and subglottal pressure variations for the phrase "*phonetic aspects*" (after *Netsell* [16.6]). As the glottis opens and closes for voiceless and voiced sounds and the vocal tract opens or closes for vowels and consonants, the expired air is opposed by varying degrees of glottal and supraglottal impedance. Both the oral and the subglottal pressure records reflect the combined effect of these resistance variations

Amazon
EX1013 Page 680

piratory muscles (the *internal intercostals*, the *rectus abdominis* among others) take over increasingly as the other group relaxes (cf. arrows, Fig. 16.6). According to our present understanding [16.8, 9], this adaptation of breathing for speech is achieved by constantly updating the balance between agonist and antagonist muscles in accordance with current conditions (lung volume, body posture, etc.) and in order to meet the goal of constant Ps. Muscle recruitment depends on lung volume.

Figure 16.7 presents a representative example of oral and Ps records. The phrase is "*phonetic aspects*" (After *Netsell* [16.6]). The top panel shows a record of oral air flow. Vertical lines indicate acoustic segment boundaries. The bottom diagram superimposes the curves for oral and subglottal pressure.

The Ps shows a falling pattern which becomes more pronounced towards the end of the phrase and which is reminiscent of the declination pattern of the fundamental frequency contour typical of declarative sentences [16.10, p. 127]. The highest values occur at the midpoints of [ɛ] and [æ], the stressed vowels of the utterance. For vowels, oral pressure is close to atmospheric pressure (near zero on the *y*-axis cm H_2O scale). The [k] of *phonetic* and the [p] of *aspects* show highly similar traces. As the tongue makes the closure for [k], the air flow is blocked and the trace is brought down to zero. This is paralleled by the oral pressure rising until it equals the Ps. As the [k] closure progresses, a slight increase builds up in both curves. The release of the [k] is signaled by a peak in the air flow and a rapid decrease in Ps. An almost identical pattern is seen for [p].

In analyzing Ps records, phoneticians aim at identifying variations based on an *active control* of the respiratory system and phenomena that can be attributed to the system's *passive response* to ongoing activity elsewhere, e.g., in the vocal tract and or at the level of the vocal folds [16.11].

To exemplify passive effects let us consider the events associated with [k] and [p] just discussed. As suggested by the data of Figs 16.4 and 16.6, speech breathing proceeds at a relatively steady lung volume decrement. However, the open or closed state of the glottis, or the presence of a vocal tract constriction/closure, is capable of creating varying degrees of impedance to the expired air. The oral pressure record reflects the combined effect of glottal and articulatory resistance variations. Ps is also affected by such changing conditions. As is evident from Fig. 16.7, the Ps traces during the [k] and [p] segments first increase during the stop closures. Then they decrease rapidly during the release and the aspiration phases. These effects are passive responses to the segment-based changes in supraglottal resistance and are unlikely to be actively programmed [16.12, 13].

Does respiration play an active role in the production of stressed syllables? In "*phonetic aspects*" main stress occurs on [ɛ] and [æ]. In terms of Ps, these vowels exhibit the highest values. Are they due to an active participation of the respiratory system in signaling stress, or are they fortuitous by-products of other factors?

An early contribution to research on breathing and speech is the work by *Stetson* [16.14]. On the basis of aerodynamic, electromyographic and chest movement measurements, Stetson proposed the notion of the *chest pulse*, a chunk of expiratory activity corresponding to the production of an individual syllable.

In the late fifties, Ladefoged and colleagues published an electromyographic study [16.7] which cast doubt on Stetson's interpretations. It reported increased activity in expiratory muscles (*internal intercostals*) for English syllables. However, it failed to support Stetson's chest pulse idea in that the increases were found only on stressed syllables. *Ladefoged* [16.8] reports findings from a replication of the 1958 study in which greater activity in the internal intercostals for stressed syllables was confirmed. Moreover, reduced activity in inspiratory muscles (*external intercostals*) was observed to occur immediately before each stressed syllable.

Ps measurements provide further evidence for a positive role for respiration in the implementation of stress. *Ladefoged* [16.15, p. 143] states: "Accompanying every stressed syllable there is always an increase in the Ps". This claim is based on data on disyllabic English noun–verb pairs differing in the position of the main stress: *TORment* (noun)–*torMENT* (verb), *INsult* (noun)–*inSULT* (verb) etc. A clear distinction between the noun and verb forms was observed, the former having Ps peaks in the first syllable and the latter on the second syllable.

As for segment-related Ps variations (e.g., the stops in Fig. 16.7), there is wide agreement that such local perturbations are induced as automatic consequences of speech production aerodynamics. Explaining short-term ripple on Ps curves in terms of aerodynamics is consistent with the observation that the respiratory system is mechanically sluggish and therefore less suited to implement rapid Ps changes in individual phonetic segments. Basically, its primary task is to produce a Ps contour stable enough to maintain vocal intensity at a fairly constant level (Fig. 16.6).

Amazon
EX1013 Page 681

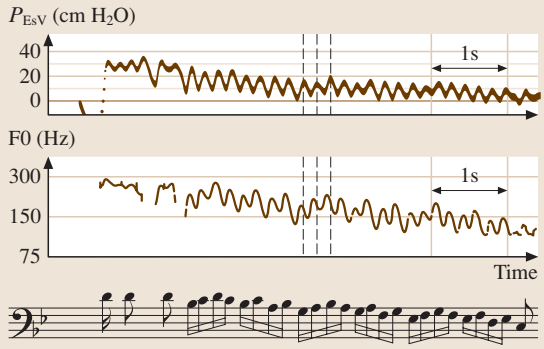


Fig. 16.8 Changes in esophageal pressure, corresponding to subglottic pressure changes (*upper curve*) and phonation frequency (*middle*) in a professional baritone singer performing the coloratura passage shown at the bottom. (After *Leanderson* et al. [16.17])

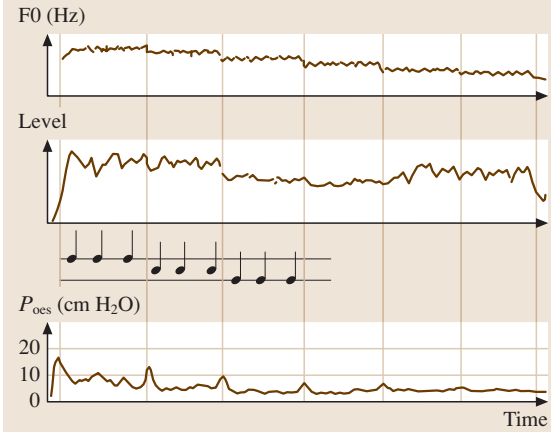


Fig. 16.9 The three *curves* show, from *top to bottom* F0 sound level and esophageal pressure in a professional baritone singer singing the music example shown in the *graph*, i. e. a descending scale with three tones on each scale tone. The singer marked the first beat in each bar by a pressure pulse. (After *Sundberg* et al. [16.18])

Singing provides experimental data on the behavior of the respiratory system that tend to reinforce the view that a time varying respiration activity can be an active participant in the sound generation process. Although there is ample justification for saying that the mechanical response of the respiratory structures is characterized by a long time constant, Ps in singing varies quickly and accurately. One example is presented in Fig. 16.8. It shows variations in pressure, captured as the pressure variations in the esophagus and thus mirroring the Ps variations [16.16], during a professional baritone singer's performance of the music example shown at the bottom. Each tone in the sequence of sixteenth notes is produced with a pressure pulse. Note also that in the fundamental frequency curve each note corresponds to a small rise and fall. The tempo of about six sixteenth notes per second implies that the duration of each rise–fall cycle is about 160 ms. It would take quite special skills to produce such carefully synchronized Ps and fundamental frequency patterns.

Synthesis experiments have demonstrated that this particular fundamental frequency pattern is what produces what is perceived as a sung legato coloratura [16.19, 20]. The voice organ seems incapable of producing a stepwise-changing F0 curve in legato, and such a performance therefore sounds as if it was produced by a music instrument rather than by a voice. In this sense this F0 variation pattern seems to be needed for eliciting the perception of a sung sequence of short legato tones.

The pressure variations are not likely to result from a modulation of glottal adduction. A weak glottal adduction should result in a voice source dominated by the fundamental and with very weak high spectrum partials, and mostly the amplitudes of the high overtones do not vary in coloratura sequences.

It is quite possible that the F0 variations are caused by the Ps variations. In the example shown in Fig. 16.8, the pressure variations amount to about 10 cm H_2O, which should cause a F0 modulation of about 30 Hz or so, corresponding to two semitones in the vicinity of 220 Hz. This means that the coloratura pattern may simply be the result of a ramp produced by the F0 regulating system which is modulated by the pressure pulses produced by the respiratory system.

As another example of the skilled use of Ps in singing, Fig. 16.9 shows the pressure variation in the esophagus in a professional baritone singer performing a descending scale pattern in 3/4 time, with three tones on each scale step as shown by the score fragment in the first three bars. The singer was instructed to mark the first tone in each bar. The pressure record demonstrates quite clearly that the first beat was produced with a marked pulse approximately doubling the pressure. Pulses of this magnitude must be produced by the respiratory apparatus. When the subject was instructed to avoid marking the first tone in each bar, no pressure pulses were observed [16.18].

The effect of Ps on fundamental frequency has been investigated in numerous studies. A commonly used

Amazon
EX1013 Page 682

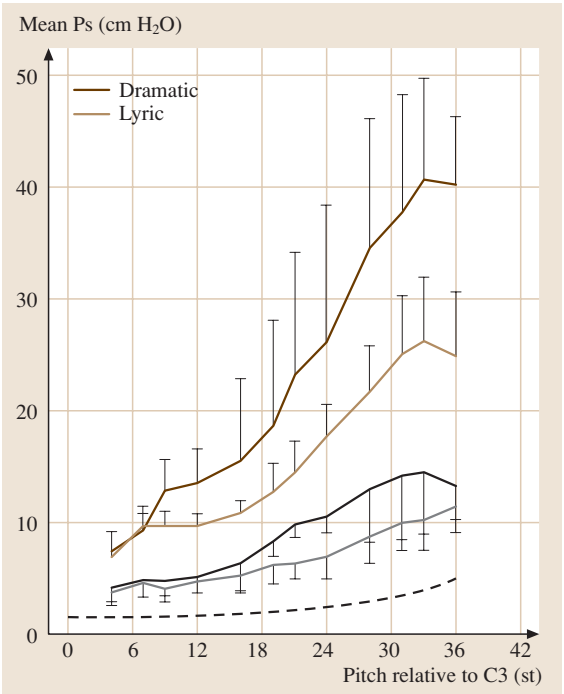


Fig. 16.10 Mean subglottal pressure, captured as the oral pressure during /p/-occlusion, in five lyric and six dramatic professional sopranos (*heavy* and *thin curves*, respectively) who sang as loudly and as softly as possible at different F0. The *dashed curve* shows Titze's prediction of threshold pressure, i. e., the lowest pressure that produces vocal fold vibration. The *bars* represent one SD (After [16.21])

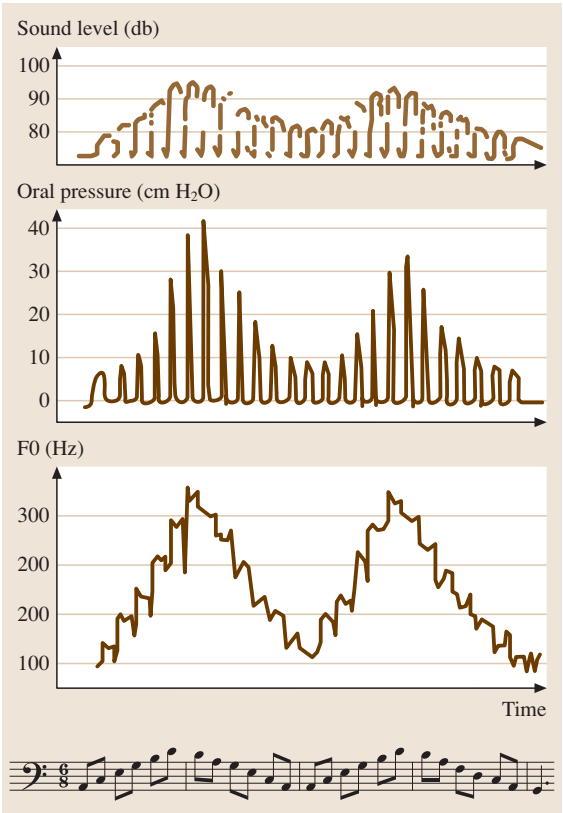


Fig. 16.11 Simultaneous recordings of sound level, subglottic pressure (captured as the oral pressure during /p/ occlusion) and F0 in a professional singer singing the note example shown at the bottom. (After *Leanderson* et al. [16.17])

method is to ask the subject to produce a sustained vowel at a given steady pitch and then, at unpredictable moments, change the Ps by applying a push to the subject's abdomen or chest. The assumption underlying these studies is that there will be an initial interval during which possible reflex action of laryngeal muscles will not yet come into play. Hence the data from this time segment should give a valid picture of the relationship between fundamental frequency and Ps.

In a study using the push method, *Baer* [16.22] established this relationship for a single male subject. Data on steady phonations at 94, 110, 220 Hz and a falsetto condition (240 Hz) were collected. From electromyograms from the *vocalis* and the *interarytenoid* it was possible to tell that the first 30 ms after the push onset were uncontaminated by reflex muscle responses. Fundamental frequency was plotted against Ps for the four F0 conditions. These plots showed linear data clusters with slopes of 4, 4, 3 and 9 Hz/cm H_2O respectively. Other studies which assumed a longer reflex response latency (100 ms) report that F0's dependence on Ps occurs between 2 to 7 Hz/cm H_2O.

From such observations phoneticians have concluded that, in producing the F0 variations of natural speech, Ps plays only a secondary role. F0 control is primarily based on laryngeal muscle activity. Nonetheless, the falling F0 contours of phrases with statement intonation tend to have Ps contours that are also falling. Moreover, in questions with final F0 increases, the Ps records are higher towards the end of the phrase than in the corresponding statements [16.15].

There is clear evidence that the Ps needs to be carefully adapted to the target fundamental frequency in singing, especially in the high range ([16.23], [16.24,

Amazon
EX1013 Page 683

p. 36]). An example is presented in Fig. 16.10 which shows mean Ps for five dramatic and five lyric sopranos when they sang pp and ff tones throughout their range [16.21]. The bars represent one standard deviation. The dashed curve is the threshold pressure, i. e. the lowest pressure that produced vocal fold vibration, for a female voice according to *Titze* [16.25, 26].

The singers mostly used more than twice as high Ps values when they were singing their loudest as compared with their softest tones. The subjects reach up to 40 cm H_2O. This can be compared with values typical of normal adult conversational speech, which tend to occur in the range of 5–10 cm H_2O. However, loud speech and particularly stage speech may occasionally approach the higher values for singing.

Also interesting is the fact that lyric sopranos use significantly lower Ps values than dramatic sopranos both in soft and loud phonation. It seems likely that this difference reflects difference in the mechanical properties of the vocal folds.

Figure 16.11 presents another example of the close link between Ps and fundamental frequency. From top to bottom it plots the time functions for sound level (SPL in dB), pressure (cm H_2O), fundamental frequency (Hz) for a professional baritone singing an ascending triad followed by a descending dominant-seventh triad (according to the score at the bottom). The pressure record was obtained by recording the oral pressure as the subject repeated the syllable [pæ] on each note. During the occlusion of the stop the peak oral pressure becomes equal to the Ps (as discussed in connection with Fig. 16.7). That means that the peak values shown in the middle diagram are good estimates of the actual Ps. It is important to note that, when the trace repeatedly returns to a value near zero, what we see is not the Ps, but the oral pressure reading for the [æ] vowel, which is approximately zero.

The exercise in Fig. 16.11 is often sung staccato, i. e. with short pauses rather than with a /p/ consonant between the tones. During these pauses the singer has to get ready for the next fundamental frequency value and must therefore avoid exhaling the pressurized air being built up in the lungs. Singers do this by opening the glottis between the tones and simultaneously reducing their Ps to zero, so that no air will be exhaled during the silent intervals between the tones. A remarkable fact demonstrated here is that, particularly when sung loudly – so that high Ps values are used – this exercise requires nothing less than a virtuoso mastering of both the timing and tuning of the breathing apparatus and the pitch control process. A failure to reach a target pressure is likely to result in a failure to reach the target F0.

16.2 The Glottal Sound Source

In speech and singing the general method to generate sound is to make a constriction and to let a strong flow of air pass through it. The respiratory component serves as the power source providing the energy necessary for sound production. At the glottis the steady flow of air generated by the respiratory component is transformed into a quasiperiodic series of glottal pulses. In the vocal tract, the glottally modified breath stream undergoes further modifications by the resonance characteristics of the oral, pharyngeal and nasal cavities.

Constrictions are formed at the glottis - by adjusting the separation of the vocal folds – and above the glottis – by positioning the articulators of the vocal tract. As the folds are brought close together, they respond to the air rushing through by initiating an open–close vibration and thereby imposing a quasiperiodic modulation of airflow. Thus, in a manner of speaking, the glottal structures operate as a device that imposes an AC modulation on a DC flow. This is basically the way that *voicing*, the sound source of voiced vowels and consonants and the carrier of intonation and melody, gets made.

A second mechanism is found in the production of *noise*, the acoustic raw materials for voiceless sounds (e.g., [f], [s], [p], [k]). The term refers to irregular turbulent fluctuations in airflow which arise when air comes out from a constriction at a high speed. This process can occur at the glottis – e.g., in [h] sounds, whisper and breathy voice qualities – or at various places of articulation in the vocal tract.

The framework for describing both singing and speech is that of the *source-filter theory of speech production* [16.27, 28]. The goal of this section is to put speech and singing side by side within that framework and to describe how the speaker and the singer coordinate respiration, phonation and articulation to shape the final product: the acoustic wave to be perceived by the listener.

Figure 16.12 [16.29] is an attempt to capture a few key aspects of vocal fold vibrations. At the center a sin-

Amazon
EX1013 Page 684

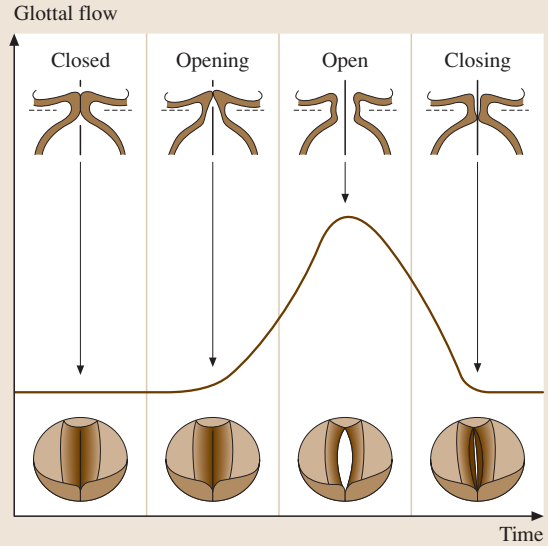


Fig. 16.12 Relationship between the flow glottogram, showing transglottal airflow versus time, and glottal configurations in a coronal plane (*upper series* of images) and from above, as through a laryngeal mirror (*lower series* of images). The airflow increases when the folds open the glottis and allow air to pass and decreases quickly when they close the glottis and arrest the airflow. (After [16.29])

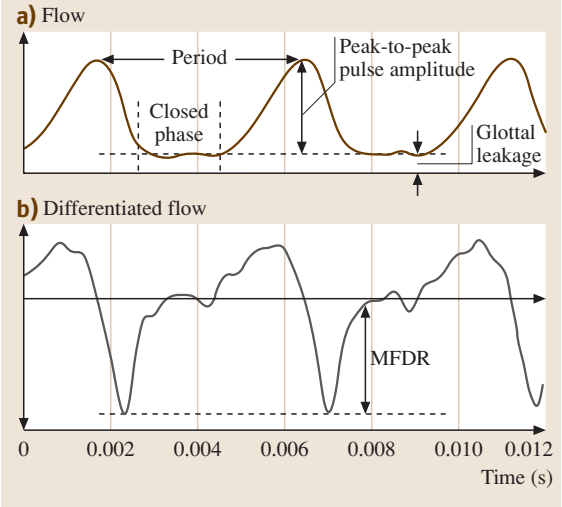


Fig. 16.13a,b Illustration of measures commonly used to characterize a flow glottogram **(a)** and its time derivative, the differentiated flow glottogram **(b)**. (The *wiggles* in the latter are artifacts caused by an imperfect inverse filtering)

gle cycle of a glottal waveform is seen. It plots airflow through the glottis as a function of time. Alternatively, the graph can be used to picture the time variations of the glottal area which present a pattern very similar to that for airflow. The top row shows stylized cross sections of the vocal folds at selected time points during the glottal cycle. From left to right they refer to the opening of the folds, the point of maximum area and the point of closure. Below is a view of the vocal folds from above corresponding to the profiles at the top of the diagram.

There are a number of different methods for visualizing vocal fold vibrations. By placing an electrode on each side of the thyroid cartilage, a minute current can be transferred across the glottis. This current increases substantially when the folds make contact. The resulting *electroglottogram*, also called a *laryngogram*, thus shows how the contact area varies with time. It is quite efficient in measurement of F0 and closed phase. *Optical glottograms* are obtained by illuminating the trachea from outside by means of a strong light source and capturing the light traveling through the glottis by means of an optical sensor in the pharynx. The signal therefore reflects the glottal area, but only as long as the light successfully finds it way to the sensor. A posterior tilting of the epiglottis may easily disturb or eliminate the signal.

Flow glottograms show transglottal airflow versus time and are derived by inverse filtering the audio signal, often picked up as a flow signal by means of a pneumotachograph mask [16.30]. Inverse filtering implies that the signal is passed through a filter with a transfer function equalling the inverted transfer function of the vocal tract. Therefore correct inverse filtering requires that the inverted resonance peaks of the inverse filter are tuned to the formant frequencies of the vowel being filtered.

As transglottal airflow is zero when the glottis is closed and nonzero when it is open, the flow glottogram is physiologically relevant. At the same time it is a representation of the sound of the voice source.

A typical example of a flow glottogram is given in the upper graph of Fig. 16.13. The classical parameters derived from flow glottograms are the durations of the period and of the closed phase, pulse peak-to-peak amplitude, and glottal leakage. The lower graph shows the differentiated glottogram. The negative peak amplitude is often referred to as the *maximum flow declination rate* (MFDR). As we shall see it has special status in the process of voice production.

In the study of both speech and singing, the acoustic parameter of main relevance is the time variations in sound pressure produced by the vocal system and received by the listener's ears. Theoretically, this signal

Amazon
EX1013 Page 685

is roughly proportional to the derivative of the output airflow at the lips [16.28, 31] and it is related to the derivative of the glottal waveform via the transfer function of the vocal tract. Formally, the excitation signal for voiced sounds is defined in terms of this differentiated signal. Accordingly, in source-filter theory, it is the derivative of glottal flow that represents the source and is applied to the filter or resonance system of the vocal tract. The amplitude of the vocal tract excitation, generally referred to as the *excitation strength*, is quantified by the maximum velocity of flow decrease during vocal-fold closing movement (the MFDR, Fig. 16.13) which is a determinant of the level of the radiated sound. At the moment of glottal closure a drastic modification of the air flow takes place. This change is what generates voicing for both spoken and sung sounds and produces a sound with energy across a wide range of frequencies.

The Liljencrants–Fant (LF) model [16.32, 33] is an attempt to model glottal waveforms using parameters such as fundamental frequency, excitation strength, dynamic leakage, open quotient and glottal frequency (defined by the time period of glottal opening phase). Other proposals based on waveform parameters have been made by Klatt and Klatt [16.34], Ljungqvist and Fujisaki [16.35], Rosenberg [16.36] and Rothenberg et al. [16.37]. A second line of research starts out from assumptions about vocal fold mechanics and applies aerodynamics to simulate glottal vibrations [16.38, 39]. Insights from such work indicate the importance of parameters such as Ps, the adducted/abducted position of vocal folds and their stiffness [16.28].

During the early days of speech synthesis it became clear that the simplifying assumption of a constant voice source was not sufficient to produce high-quality natural-sounding copy synthesis. Experimental work on the voice source and on speech synthesis has shown that, in the course of an utterance, source parameters undergo a great deal of variation. The determinants of this dynamics are in part prosodic, in part segmental. Figure 16.14 [16.32] presents a plot of the time variations of the *excitation strength* parameter (i. e. MFDR) during the Swedish utterance: *Inte i DETta århundrade* [ɪntɪ ˣdɛtːaoːrhəndradɛ]. The upper-case letters indicate that the greatest prominence was on the first syllable of *detta*. Vertical lines represent acoustic segment boundaries.

Gobl collected flow data using the mask developed by Rothenberg [16.30] and applied inverse filtering to obtain records of glottal flow which, after differentiation, enabled him to make measurements of excitation strength and other LF parameters.

Figure 16.14 makes clear that excitation strength is in no way constant. It varies depending on both prosodic and segmental factors. The effect of the segments is seen near the consonant boundaries. As the vocal tract is constricted, e.g., in [d] and [tː], and as transglottal pressure therefore decreases (cf. the pressure records of Fig. 16.7), excitation strength is reduced. In part these variations also occur to accommodate the voicing and the voicelessness of the consonant [16.28]. This influence of consonants on the voice source has been documented in greater detail by *Ni Chasaide* and *Gobl* [16.40] for German, English, Swedish, French, and Italian. Particularly striking effects were observed in the context of voiceless consonants.

Prosodically, we note in Fig. 16.14 that excitation strength exhibits a peak on the contrastively stressed syllable in *detta* and that the overall pattern of the phrase is similar to the falling declination contour earlier mentioned for declarative statements.

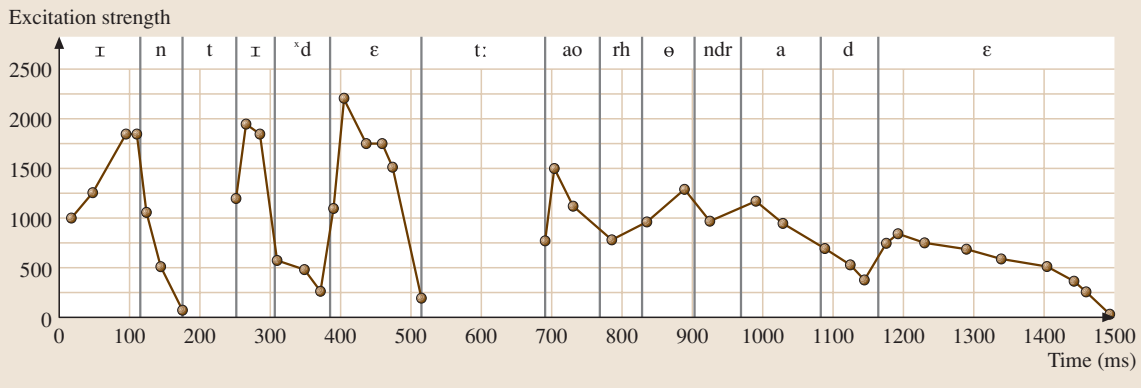


Fig. 16.14 Running speech data on the 'excitation strength parameter' of the LF model. (After [16.32])

Amazon
EX1013 Page 686

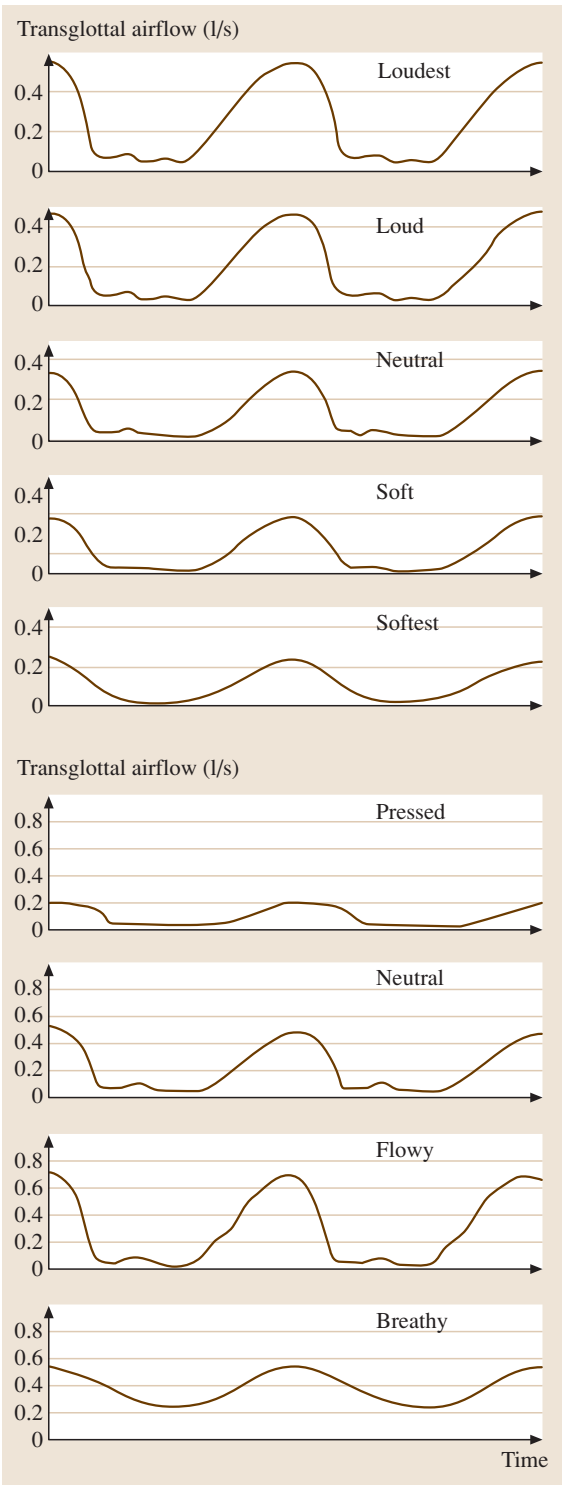


Table 16.1 Measurements of subglottal pressure (cm H_2O) and SPL at 0.3 m (dB) measured for the glottograms shown in Fig. 16.15

Table A	Ps (cm H_2O)	SPL at 0.3 m(dB)
loudest	14.3	84
loud	7.3	82
neutral	5.7	78
soft	3.9	75
softest	3	70
Table B	**Ps (cm H_2O)**	**SPL at 0.3 m(dB)**
pressed	11.4	83
neutral	5.1	79
flowy	8	88
breathy	6.6	84

Examples of the fact that the Ps has a strong influence on the flow glottogram are given in the upper set of graphs of Fig. 16.15, which shows a set of flow glottograms for phonations produced at the same pitch but with varying degrees of *vocal loudness*. As we examine the series of patterns from loudest to softest we note that both the peak flow and the maximum steepness of the trailing end of the pulse, i. e., MFDR, increase significantly with increasing Ps. These shape changes are lawfully related to the variations in Ps and are directly reflected in sound pressure levels as indicated by the numbers in Table 16.1.

Holmberg and colleagues [16.41] made acoustic and airflow recordings of 25 male and 20 female speakers producing repetitions of [pæ] at soft, normal and loud vocal efforts [16.42, p. 136]. Estimates of Ps and glottal airflow were made from recordings of oral pressure and oral airflow. Ps was derived by interpolating between peak oral pressures for successive [p] segments and then averaging across repetitions. A measure of average flow was obtained by low-pass filtering the airflow signal and averaging values sampled at the vowel midpoints.

Fig. 16.15 Typical flow glottogram changes associated with changes of loudness or mode of phonation (*top* and *bottom series*). As loudness of phonation is raised, the closing part of the curve becomes more steep. When phonation is pressed, the glottogram amplitude is low and the closed phase is long. As the adduction force is reduced, pulse amplitude grows and the closed phase becomes shorter (note that the flow scales are different for the *top* and *bottom series* of glottograms. Flowy phonation is the least adducted, and yet not leaky phonation

Amazon
EX1013 Page 687

A copy of the airflow signal was low-pass filtered and inverse filtered to remove the effect of F1 and other formants. The output was then differentiated for the purpose of determining the MFDR (Fig. 16.13).

Figure 16.15 also illustrates how the voice source can be continuously varied between different *modes of phonation*. These modes range from hyperfunctional, or pressed, over neutral and flowy to hypofunctional, or breathy. The corresponding physiological control parameter can be postulated to be glottal adduction, i. e. the force by which the folds press against each other. It varies from minimum in hypofunctional to extreme in hyperfunctional. Flowy phonation is produced with the weakest degree of glottal adduction compatible with a full glottal closure. The physiologically relevant property that is affected is the vibration amplitude of the vocal folds, which is small in hyperfunctional/pressed phonation and wide in breathy phonation.

As illustrated in Fig. 16.15 the flow glottogram is strongly affected by these variations in phonation mode [16.43]. In pressed phonation, the pulse amplitude is small and the closed phase is long. It is larger in neutral and even more so in flowy. In breathy phonation, typically showing a waveform similar to a sine wave, airflow is considerable, mainly because of a large leakage, so there is no glottal closure.

Phonation mode affects the relation between Ps and the SPL of the sound produced. As shown in Table 16.1B pressed phonation is less economical from an acoustic point of view: a Ps of 11.4 cm H_2O produces an SPL at 0.3 m of only 83 dB, while in flowy phonation a lower Ps produces a higher SPL.

Pitch, loudness and phonation mode are voice qualities that we can vary continuously. By contrast, vocal registers, also controlled by glottal parameters, appear more like toggles, at least in untrained voices. The voice is operating either in one or another register. There are at least three vocal registers, *vocal fry*, *modal* and *falsetto*. When shifting between the modal and falsetto registers, F0 discontinuities are often observed [16.44].

The definition of vocal registers is quite vague, a set of tones along the F0 continuum that sound similar and are felt to be produced in a similar way. As registers depend on glottal function, they produce different flow glottogram characteristics. Figure 16.16 shows typical examples of flow glottograms for the falsetto and modal registers as produced by professional baritone, tenor and countertenor singers. The pulses are more rounded, the closed phase is shorter, and the glottal leakage is greater in the falsetto than in the modal register. However, the waveform of a given register often varies substantially between individuals. Classically trained sopranos, altos, and tenors learn to make continuous transitions between the modal and the falsetto registers, avoiding abrupt changes in voice timbre.

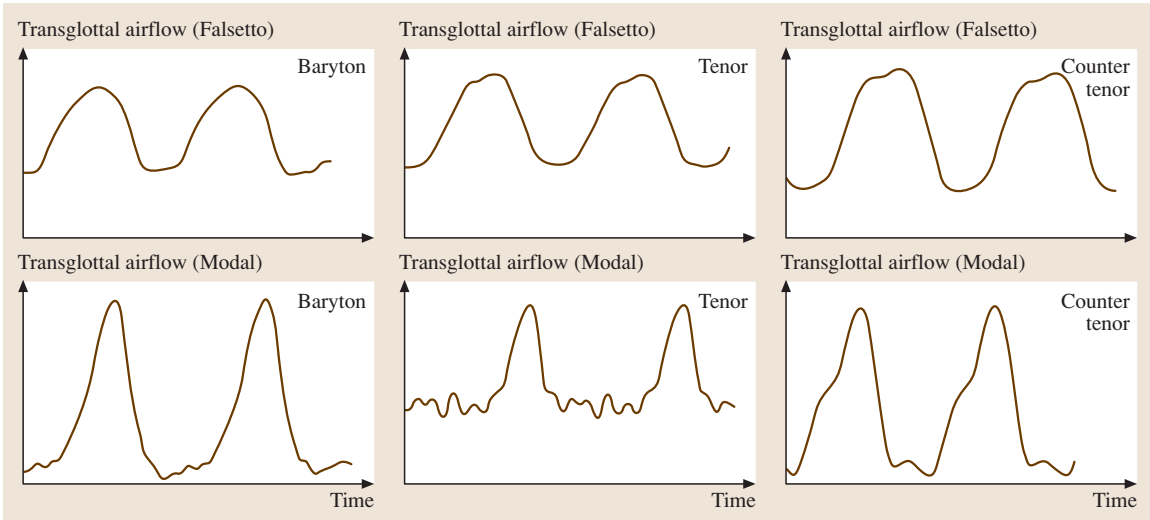


Fig. 16.16 Typical flow glottograms for falsetto and modal register in a baritone, tenor and countertenor singer, all professional. The flow scale is the same within subjects. The ripple during the closed phase in the lower middle glottogram is an artifact. In spite of the great inter-subject variability, it can be seen that the glottogram pulses are wider and more rounded in the falsetto register

Amazon
EX1013 Page 688

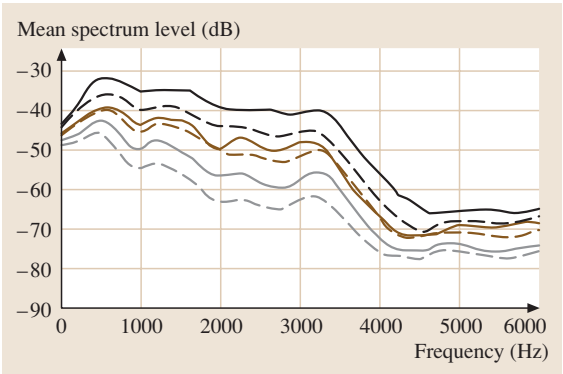


Fig. 16.17 Long-term-average spectra curves obtained from an untrained male speaker reading the same text at 6 different degrees of vocal loudness. From *top to bottom* the corresponding L_{eq} values at 0.3 m were 93 dB, 88 dB, 85 dB, 84 dB, 80 dB, and 76 dB

Variation of vocal loudness affects the spectrum slope as illustrated in Fig. 16.17, which shows long-term-average spectra (LTAS) from a male untrained voice. In the figure loudness is specified in terms of the so-called equivalent sound level L_{eq}. This is a commonly used time average of sound level, defined as

$$L_{eq} = 10 \log \frac{1}{T} \int_0^T \frac{p^2}{p_0^2} \, dt \, ,$$

where t is time and T the size of the time window. p and p_0 are the sound pressure and the reference pressure, respectively.

When vocal loudness is changed, the higher overtones change much more in sound level than the lower overtones. In the figure, a 14 dB change of the level near 600 Hz is associated with a 22 dB change near 3000 Hz, i. e., about 1.5 times the level change near 600 Hz. Similar relationships have been observed for professional singers [16.47]. In other words, the slope of the voice source spectrum decreases with increasing vocal loudness.

The physiological variable used for variation of vocal loudness is Ps. This is illustrated in the upper graph of Fig. 16.18, comparing averaged data observed in untrained female and male subjects and data obtained from professional operatic baritone singers [16.45, 46]. The relationship between the Ps and MFDR is approximately linear. It can be observed that the pressure range used by the singer is considerably wider than that used by the untrained voices. The MFDR produced with a given Ps by the untrained female and male subjects is mostly

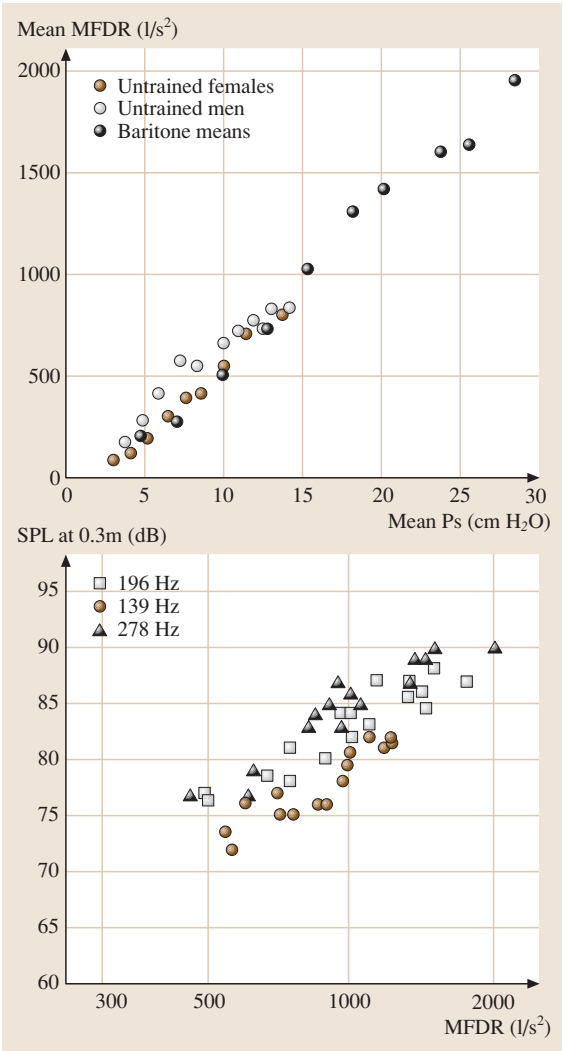


Fig. 16.18 The *top* graph shows the relationship between the mean subglottal pressure and the mean MFDR for the indicated subject groups. The *bottom* graph shows the relationship between MFDR and the SPL at 0.3 m for a professional baritone singing the vowels /a/ and /æ/ at different F0s. (After [16.45] p. 183, [16.46] p. 184)

higher than that produced by the baritones with the same pressure.This may depend on different mechanical characteristics of the vocal folds.

As we will see later, SPL depends on the strength of the excitation of the vocal tract, i. e. on MFDR. This variable, in turn, depends on Ps and F0; the higher the pressure, the greater the MFDR value and the higher the F0, the greater the MFDR. The top graph of Fig. 16.18

Amazon
EX1013 Page 689

shows how accurately MFDR could be predicted from Ps and F0 for previously published data for untrained male and female singers and for professional baritone singers [16.45, 46]. Both Ps and F0 are linearly related to MFDR. However, the singers showed a much greater variation with F0 than the untrained voices. This difference reflected the fact that unlike the untrained subjects the singers could sing a high F0 much more softly than the untrained voices. The ability to sing high notes also softly would belong to the essential expressive skills of a singer. Recalling that an increase of Ps increases F0 by a few Hz/cm H_2O, we realize that singing high tones softly requires more forceful contraction of the pitch-raising laryngeal muscles than singing such tones loudly.

16.3 The Vocal Tract Filter

The source-filter theory, schematically illustrated in Fig. 16.19, describes vocal sound production as a three-step process: (1) generation of a steady flow of air from the lungs (DC component); (2) conversion of this airflow into a pseudo-periodically pulsating transglottal airflow (DC-to-AC conversion), referred to as the voice source; and (3) response of the vocal tract to this excitation signal (modulation of AC signal) which is characterized by the frequency curve or transfer function of the vocal tract. So far the first two stages, respiration and phonation, have been considered.

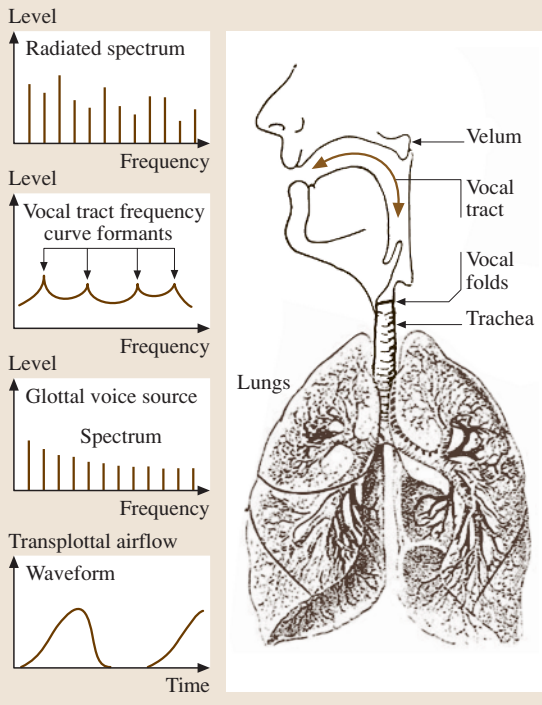


Fig. 16.19 Schematic illustration of the generation of voice sounds. The vocal fold vibrations result in a sequence of voice pulses (*bottom*) corresponding to a series of harmonic overtones, the amplitudes of which decrease monotonically with frequency (*second from bottom*). This spectrum is filtered according to the sound transfer characteristics of the vocal tract with its peaks, the formants, and the valleys between them. In the spectrum radiated from the lip opening, the formants are depicted in terms of peaks, because the partials closest to a formant frequency reach higher amplitudes than neighboring partials

In this section we will discuss the third step, viz. how the vocal tract filter, i. e. the resonance characteristics of the vocal tract, modifies, and to some extent interacts with, the glottal source and shapes the final sound output radiated from the talker's/singer's lips.

Resonance is a key feature of the filter response. The oral, pharyngeal and nasal cavities of the vocal tract form a system of resonators. During each glottal cycle the air enclosed by these cavities is set in motion by the glottal pulse, the main moment of excitation occurring during the closing of the vocal folds, more precisely at the time of the MFDR, the maximum flow declination rate (cf. the previous section on source).

The behavior of a vocal tract resonance, or *formant*, is specified both in the time and the frequency domains. For any transient excitation, the time response is an exponentially decaying cosine [16.27, p. 46]. The frequency response is a continuous amplitude-frequency spectrum with a single peak. The shape of either function is uniquely determined by two numbers (in Hz): the formant frequency F and the bandwidth B. The bandwidth quantifies the degree of damping, i. e., how fast the formant oscillation decays. Expressed as sound pressure variations, the time response is

$$p(t) = A\,\mathrm{e}^{-\pi Bt}\cos(2\pi Ft)\ . \tag{16.1}$$

For a single formant curve, the amplitude variations as a function of frequency f is given (in dB) by

$$L(f) = 20\log\frac{\left[F^2+\left(\frac{B}{2}\right)^2\right]}{\sqrt{(f-F)^2+\left(\frac{B}{2}\right)^2}\sqrt{(f+F)^2+\left(\frac{B}{2}\right)^2}}\ . \tag{16.2}$$

Amazon
EX1013 Page 690

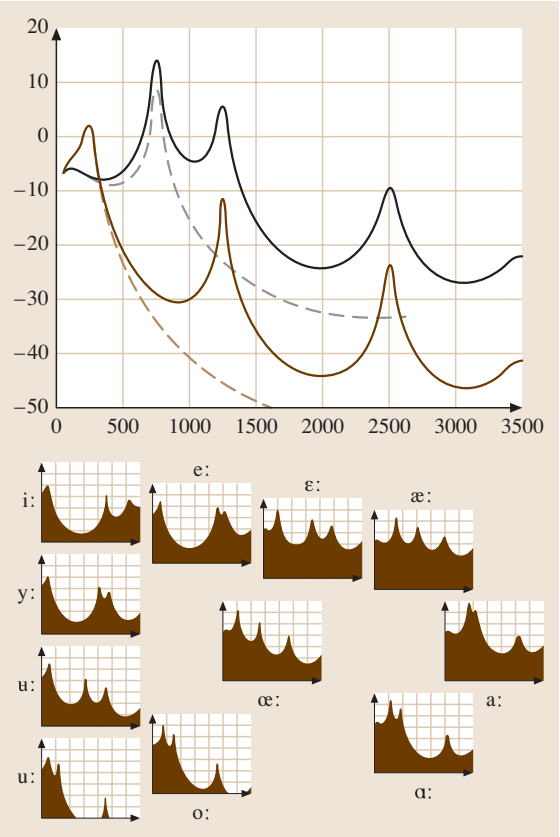


Fig. 16.20 Predictability of formant levels. Calculated spectral envelopes for a set of Swedish vowels. The *upper graph* illustrates how varying only the frequency of F1 affects the amplitudes of other formants

In the frequency domain the bandwidth is defined as the width of the formant 3 dB down from the peak, that is, at the half-power points. A large bandwidth produces a flat peak whereas a small value (less damping, reduced acoustic losses) makes the peak higher and sharper.

Figure 16.20 shows spectral envelopes for a set of Swedish vowels. They were calculated from formant frequency data [16.48, 49] in accordance with source-filter theory which assumes that a vowel spectrum can be decomposed into formants (all specified with respect to frequency and bandwidth), the spectrum of the voice source, the contribution of formants above F4 and radiation [16.27].

The individual panels are amplitude versus frequency plots. Vowels are portrayed in terms of their envelopes rather than as line spectra with harmonics. The panels are arranged in a *formant space*. From top to bottom the F2 in the panels decreases. From left to right F1 increases. The frequency scale is limited to showing the first three formant peaks.

We note that all envelopes have falling overall slopes and that the amplitudes of the formant peaks vary a great deal from vowel to vowel and depending on the relative configuration of formant frequency positions.

In acoustic phonetic specifications of vowels, it is customary to report no more than the frequencies of the first two or three formants ([16.50, p. 59], [16.51]). Experiments in speech synthesis [16.52] have indicated that this compact description is sufficient to capture the quality of steady-state vowels reasonably well. Its relative success is explained by the fact that most of the building blocks of a vowel spectrum are either predictable (bandwidths and formant amplitudes), or show only limited spectral variations (source, radiation, higher-formant correction) [16.27].

Formant bandwidths [16.28, 53] reflect acoustic losses. They depend on factors such as radiation, sound transmission through the vocal tract walls, viscosity, heat conduction, constriction size as well as the state of the glottis. For example, a more open glottis, as in a breathy voice, will markedly increase the bandwidth of the first formant.

Despite the complex interrelations among these factors, bandwidths pattern in regular ways as a function of formant frequency. Empirical formulas have been proposed [16.54] that summarize bandwidth measurements made using transient and sweep-tone excitation of the vocal tract for closed-glottis conditions [16.55, 56].

To better understand how formant levels vary let us consider the top diagram of Fig. 16.20. It compares envelopes for two vowel spectra differing only in terms of F1. It is evident that the lowering of F1 (from 750 to 250 Hz) reduces the amplitudes of F2 and F3 by about 15 dB. This effect is predicted by acoustic theory which derives the spectral envelope of an arbitrary vowel as a summation of individual formant curves on a dB scale [16.27]. Figure 16.20 makes clear that, as F1 is shifted, its contribution to the envelope is drastically changed. In this case the shift moves the entire F1 curve down in frequency and, as a result, its upper skirt (dashed line) provides less of a *lift* to the upper formants. This interplay between formant frequency and formant levels is the major determinant of the various envelope shapes seen in Fig. 16.20.

One of the main lessons of Fig. 16.20 is accordingly that, under normal conditions of a stable voice source, formant amplitudes are predictable. Another

Amazon
EX1013 Page 691

important consequence of the source-filter theory is that, since knowing the formants will enable us to reconstruct the vowel's envelope, it should also make it possible to derive estimates about a vowel's overall intensity.

A vowel's intensity can be calculated from its power spectrum as

$$I = 10 \log \left[\sum (A_i)^2 \right] , \qquad (16.3)$$

where A_i is the sound pressure of the i-th harmonic. This measure tends to be dominated by the strongest partial or partials. In very soft phonation the strongest partial is generally the fundamental while in neutral and louder phonation it is the partial that lies closest to the first formant. Thus, typically all partials which are more than a few dB weaker than the strongest partial in the spectrum do not contribute appreciably to the vowel's intensity.

As suggested by the envelopes in Fig. 16.20, the strongest harmonics in vowels tend to be found in the F1 region. This implies that a vowel's intensity is primarily determined by its F1. Accordingly, in the set shown in Fig. 16.20, we would expect [iːyːɯːuː] to be least intense and [æːɑɑː] the most intense vowels. This is in good agreement with experimental observations [16.57].

The intrinsic intensity of vowels and other speech sounds has been a topic of interest to phoneticians studying the acoustic correlates of stress [16.58]. It is related to sonority, a perceptual attribute of speech sounds that tends to vary in a regular way in syllables [16.59]. We will return below to that topic in Sect. 16.5.

Let us continue to illustrate how formant levels depend on formant frequencies with an example from singing: the singer's formant. This striking spectral phenomenon is characteristic of classically trained male singers. It is illustrated in Fig. 16.21. The figure shows a spectrogram of a commercial recording of an operatic tenor voice (Jussi Björling) performing a *recitativo* from Verdi's opera Aida. Apart from the vibrato undulation of the partials, the high levels of partials in the frequency range 2200–3200 Hz are apparent. They represent the singer's formant.

The singer's formant is present in all voiced sounds as sung by operatic male singers. It was first discovered by *Bartholomew* [16.60]. It manifests itself as a high, marked peak in the long-term-average spectrum (LTAS).

A second example is presented in Fig. 16.22 showing an LTAS of the same tenor voice as in Fig. 16.21, accompanied by an orchestra of the traditional western

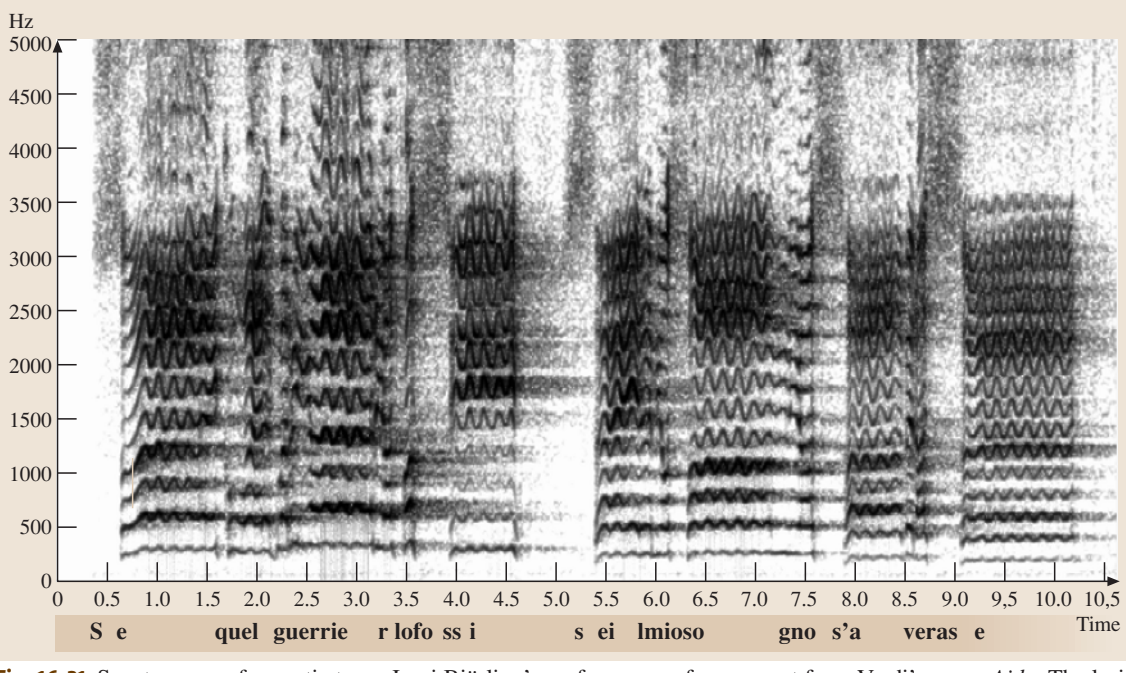


Fig. 16.21 Spectrogram of operatic tenor Jussi Björling's performance of an excerpt from Verdi's opera *Aida*. The lyrics are given below the graph

Amazon
EX1013 Page 692

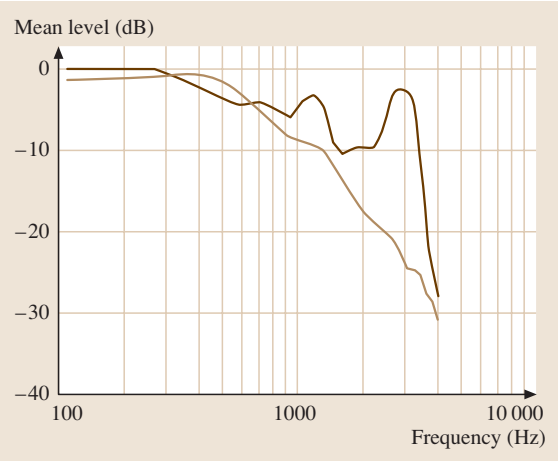


Fig. 16.22 LTAS of the sound of a symphonic orchestra with and without a singer soloist (*dark* and *light curves*). The singer's formant constitutes a major difference between the orchestra with and without the singer soloist. (After [16.61])

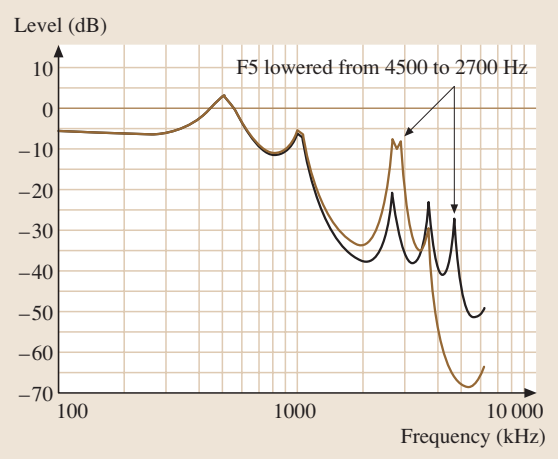


Fig. 16.23 Effect on the spectrum envelope of lowering formants F4 and F5 from 3500 Hz and 4500 Hz to 2700 Hz and 3500 Hz, respectively. The resulting gain in level at F3 is more than 10 dB

opera type. The LTAS peak caused by the singer's formant is just a few dB lower than the main peak in the low-frequency range.

In the same figure an LTAS of a classical symphony orchestra is also shown. On the average, the partials in the low-frequency range are quite strong but those above 600 Hz decrease by about 10 dB/octave with rising frequency. Even though this decrease varies depending on how loudly the orchestra is playing, this implies that the level of the orchestral accompaniment is much lower at 3000 Hz than at about 600 Hz. In other words, the orchestral sound offers the singer a rather reasonable competition in the frequency range of the singer's formant. The fact of the matter is that a voice possessing a singer's formant is much easier to hear when the orchestral accompaniment is loud than a voice lacking a singer's formant. Thus, it helps the singer's voice to be heard when the orchestral accompaniment is loud.

The singer's formant can be explained as a resonance phenomenon [16.62]. It is a product of the same rules that we invoked above to account for the formant amplitudes of vowels and for intrinsic vowel intensities. The strategy of a classically trained male singer is to shape his vocal tract so as to make F3, F4, and F5 form a tight cluster in frequency. As the frequency separations among these formants are decreased, their individual levels increase, and hence a high spectral peak is obtained between 2500 and 3000 Hz.

Figure 16.23 shows the effects on the spectrum envelope resulting from lowering F5 and F4 from 3500 Hz and 4500 Hz to 2700 Hz and 3500 Hz, respectively. The resulting increase of the level of F3 amounts to 12 dB, approximately. This means that male operatic singers produce a sound that can be heard more easily through a loud orchestral accompaniment by tuning vocal tract resonances rather than by means of producing an excessive Ps.

The acoustical situation producing the clustering of F3, F4, and F5 is obtained by acoustically mismatching the aperture of the larynx tube, also referred to as the epilaryngeal tube, with the pharynx [16.62]. This can be achieved by narrowing this aperture. Then, the larynx tube acts as a resonator with a resonance that is not much affected by the rest of the vocal tract but rather by the shape of the larynx tube. Apart from the size of the aperture, the size of the laryngeal ventricle would be influential: the larger the ventricle, the lower the larynx tube resonance. Presumably singers tune the larynx tube resonance to a frequency close to F3. The articulatory means used to establish this cavity condition seems mainly to be a lowering of the larynx, since this tends to widen both the pharynx and the laryngeal ventricle. Many singing teachers recommend students to sing with a comfortably low larynx position.

The level of the singer's formant is influenced also by the slope of the source spectrum which, in turn, depends on vocal loudness, i. e. on Ps, as mentioned. Thus, the singer's formant tends to increase by about 15 dB for a 10 dB change of the overall SPL [16.47].

Amazon
EX1013 Page 693

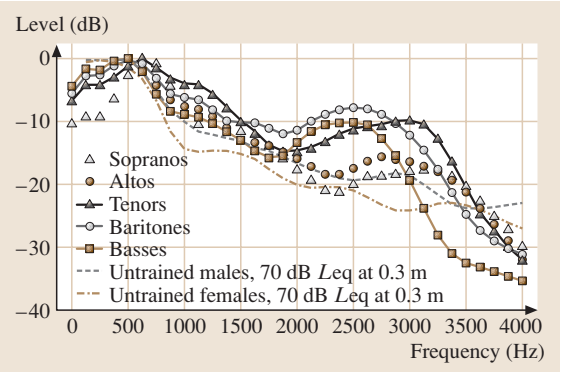


Fig. 16.24 Mean LTAS derived from commercial recordings of four representatives of each of the indicated voice classifications. The *dashed curves* show LTAS of untrained female (red) and male (blue) voices' speech with an L_{eq} of 70 dB at 0.3 m. Note that the singer's formant produces a prominent peak with a level well above the level of the untrained voices' LTAS only for the male singers

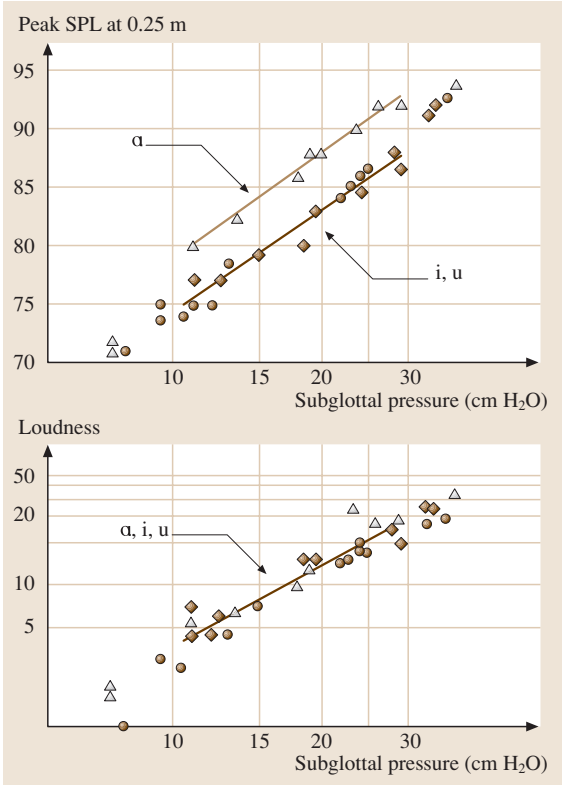


Fig. 16.25 Perceived loudness of a vowel (*bottom panel*) correlates better with subglottal pressure Ps than does SPL (*top panel*) which depends on the intrinsic intensity of the vowel. This, in turn, is determined by its formant pattern. (After [16.66])

The center frequency of the singer's formant varies slightly between voice classifications, as illustrated in the mean LTAS in Fig. 16.24 which shows mean LTAS derived from commercial recordings of singers classified as soprano, alto, tenor, baritone and bass [16.63]. Each group has four singers. The center frequency of the singer's formant for basses, baritones and tenors are about 2.4, 2.6, and 2.8 kHz, respectively. These small differences are quite relevant to the typical voice timbres of these classifications [16.64]. Their origin is likely to be vocal tract length differences, basses tending to have longer vocal tracts than baritones who in turn have longer vocal tracts than tenors [16.65]. On the other hand, substantial variation in the center frequency of the singer's formant occurs also within classifications.

Also shown for comparison in Fig. 16.24 is a mean LTAS of untrained voices reading at an L_{eq} of 70 dB at 0.3 m distance. The singer's formant produces a marked peak some 15 dB above the LTAS of the untrained voices for the male singers. The female singers, on the other hand, do not show any comparable peak in this frequency range. This implies that female singers do not have a singer's formant [16.67–69].

Female operatic singers' lack of a singer's formant is not surprising, given the fact that: (1) they sing at high F0, i. e. have widely spaced spectrum partials, and (2) the F3, F4, F5 cluster that produces a singer's formant is rather narrow in frequency. The latter means that a narrow formant cluster will be hit by a partial only in some tones of a scale while in other scale tones there will be no partial in the formant cluster. This would lead to salient and quasi-random variation of voice timbre between scale tones.

The singer's formant is a characteristic of classically trained male singers. It is not found in nonclassical singing, e.g. in pop or musical theater singing, where audibility is the responsibility of the sound engineer rather than of the singer. Likewise, choir singers generally do not possess a singer's formant.

From the evidence in Fig. 16.18 we concluded that the logarithm of Ps does a good job of predicting the SPL over a large range of vocal intensities. This was shown for untrained male and female speakers as well as a professional baritone singer. In the next few paragraphs we will look at how vowel identity affects that prediction and how listeners perceive loudness in the presence of vowel-dependent variations in SPL at constant Ps.

Amazon
EX1013 Page 694

This topic was addressed by Ladefoged [16.66] who measured the peak SPL and the peak Ps in 12 repetitions of *bee*, *bay*, *bar*, *bore* and *boo* spoken by a British talker at varying degrees of loudness. SPL values were plotted against log(Ps) for all tokens. Straight lines could readily be fitted to the vowels individually. The vowels of *bay* and *bore* tended to have SPL values in between those of *bar* and *bee*/*boo*. The left half of Fig. 16.25 replots the original data for [ɑ] and [i]/[u] pooled. The [ɑ] line is higher by 5–6 dB as it should be in view of F1 being higher in [ɑ] than in [i] and [u] (cf. preceding discussion).

In a second experiment listeners were asked to judge the loudness of the test words. Each item was presented after a carrier phrase: *Compare the words: bar and* __. The word bar served as reference. The subjects were instructed to compare the test syllable and the reference in terms of loudness, to give the value of 10 to the reference and another relative number to the test word. The analysis of the responses in terms of SPL indicated that, for a given SPL, it was consistently the case that *bee* and *boo* tended to be judged as louder than *bar*. On the other hand, there were instances of *bee* and *bar* with similar Ps that were judged to be equally loud. This effect stood out clearly when loudness judgements were plotted against log(Ps) as in the bottom half of Fig. 16.25. Ladefoged concludes that " ... in the case of speech sounds, loudness is directly related to the physiological effort" – in other words, Ps – rather than the SPL as for many other sounds.

Speech Sounds with Noise and Transient Excitation

A comprehensive quantitative treatment of the mechanisms of noise production in speech sounds is found in *Stevens* [16.28, pp. 100–121]. This work also provides detailed analyses of how the noise source and vocal tract filtering interact in shaping the bursts of stops and the spectral characteristics of voiced and voiceless fricatives. While the normal source mechanism for vowels always involves the glottis, noise generation may take place not only at the glottis but at a wide range of locations along the vocal tract. A useful rule of thumb for articulations excited by noise is that the output will spectrally be dominated by the cavity in front of the noise source. This also holds true for sounds with transient excitation such as stop releases and click sounds [16.28, Chapts. 7,8].

While most of the preceding remarks were devoted to the acoustics of vowels, we should stress that the source-filter theory applies with equal force to the production of consonants.

16.4 Articulatory Processes, Vowels and Consonants

X-ray films of normal speech movements reveal a highly dynamic process. Articulatory activity comes across as a complex flow of rapid, parallel lip, tongue and other movements which shows few, if any, steady states. Although the speaker may be saying no more than a few simple syllables, one nonetheless has the impression of a virtuoso performance of a *polyphonic* motor score. As these events unfold, they are instantly reflected in the acoustic output. The articulatory movements modify the geometry of the vocal tract. Accordingly, the filter (transfer function) undergoes continual change and, as a result, so do the output formant frequencies.

Quantitative modeling is a powerful tool for investigating the speech production process. It has been successfully applied to study the relations between formant and cavities. Significant insights into this mapping have been obtained by representing a given vocal tract configuration as an *area function* – i. e., a series of cylindrical cross sections of variable lengths and cross-sectional areas – and by simulating the effects of changes in the area function on the formant frequencies [16.27].

In the past, pursuing this approach, investigators have used lateral X-ray images – similar to the magnetic

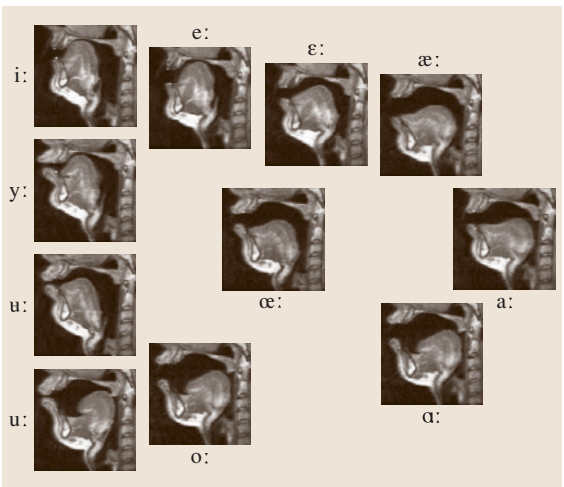


Fig. 16.26 Magnetic resonance images of the Swedish vowels of Fig. 16.20

Amazon
EX1013 Page 695

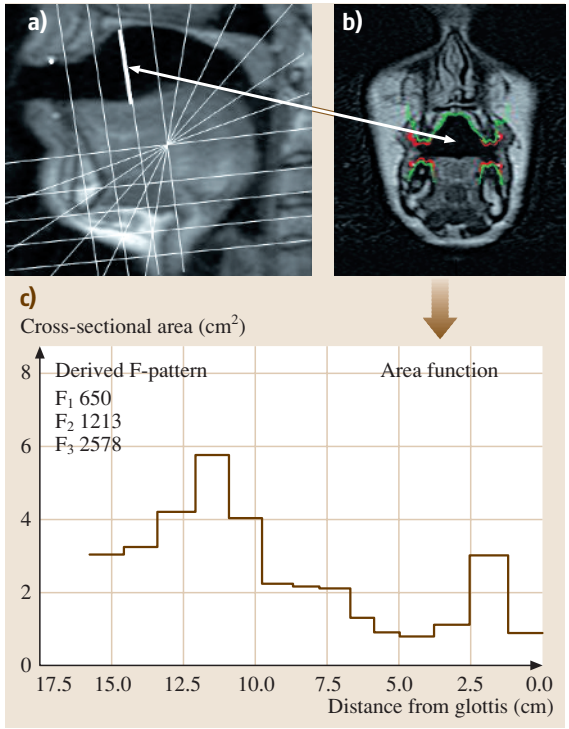


Fig. 16.27a–c From articulatory profile to acoustic output. Deriving the formant patterns of a given vowel articulation involves **(a)** representing the vocal tract profile in terms of the cross-distances from the glottis to the lips; **(b)** converting this specification into a cross-sectional area function; and **(c)** calculating the formant pattern from this area function. To go from **(a)** to **(b)**, profile data (*top left*) need to be supplemented by area measurements obtained from transversal (coronal and axial) images of the cross sections

resonance imaging (MRI) pictures in Fig. 16.26 – to trace the outlines of the acoustically relevant articulatory structures. To make estimates of cross-sectional areas along the vocal tract such lateral profiles need to be supplemented with information on the transverse geometry of the cross sections, e.g., from casts of the vocal tract [16.70] and tomographic pictures [16.27, 71].

More currently, magnetic resonance imaging methods have become available, making it possible to obtain three-dimensional (3-D) data on sustained steady articulations [16.72–76]. Figure 16.26 presents MRI images taken in the mid-sagittal plane of a male subject during steady-state productions of a set of Swedish vowels. These data were collected in connection with work on APEX, an articulatory model developed for studying the acoustic consequences of articulatory movements, e.g., the lips, the tongue and the jaw [16.77–81].

Figure 16.27 highlights some of the steps involved in deriving the formant pattern from lateral articulatory profiles such those of Fig. 16.26. The top-left picture is the profile for the subject's [a]. The white lines indicate the coronal, coronal oblique and axial planes where ad-

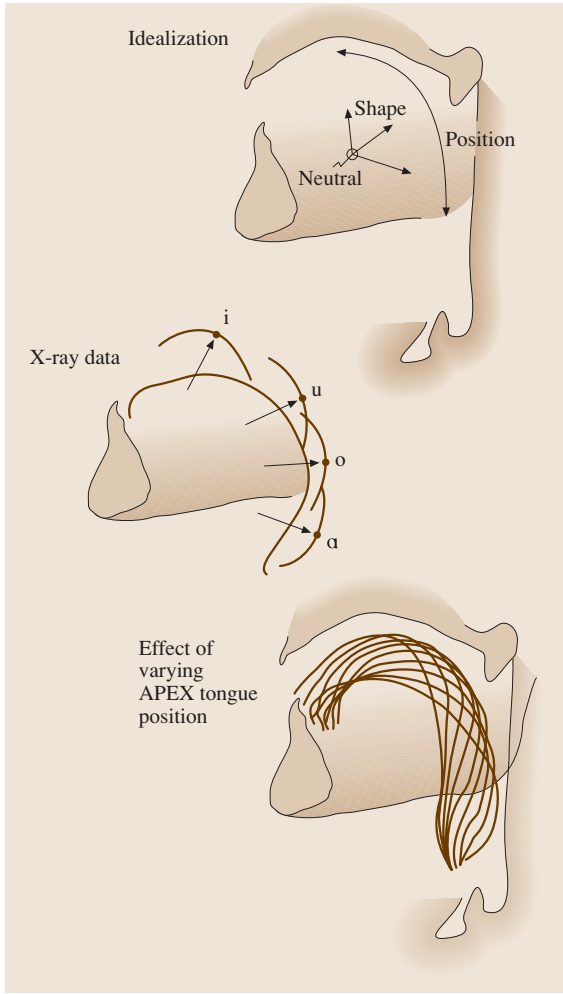


Fig. 16.28 The top drawing shows superimposed observed tongue shapes for [i], [u], [o] and [ɑ] differing in the place of the main constriction. Further analyses of similar data suggest that, for vowels, two main dimensions are used: The anterior–posterior location of the tongue body and its displacement from neutral which controls the degree of vocal tract constriction (*lower right*). These parameters are implemented numerically to produce vowels in the APEX articulatory model

Amazon
EX1013 Page 696

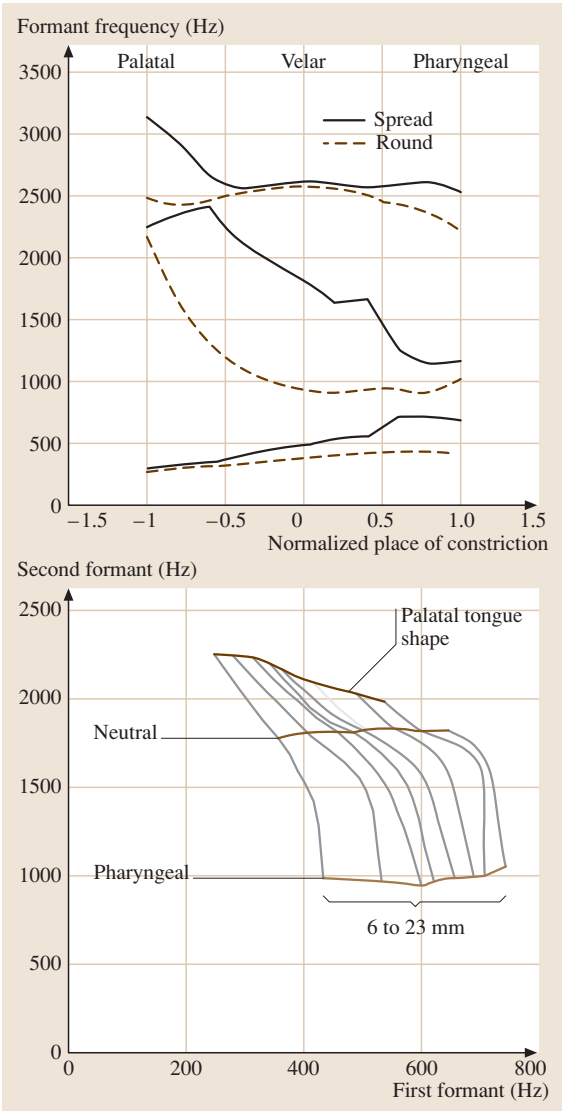


Fig. 16.29 Acoustic consequences of tongue body, jaw and lip movements as examined in APEX simulations. The tongue moves continuously between palatal, velar and pharyngeal locations while maintaining a minimum constriction area A_{min} of 0.25 cm, a 7 mm jaw opening and a fixed larynx height. The *dashed* and *solid lines* refer to rounded and spread conditions. The *lower diagram* illustrates the effect of varying the jaw. The *thin lines* show tongue movement at various degrees of jaw opening (6, 7, 8, 9, 12, 15, 19 and 23 mm). The *bold lines* pertain to fixed palatal, neutral or pharyngeal tongue contours

ditional MR images were taken to get information on the cross sections in the transverse dimension.

The location of the transverse cut is indicated by the bold white line segment in the lateral profile. In the coronal section to the right the airway of the vocal tract is the dark area at the center. Below it, the wavy surface of the tongue is evident. Since the teeth lack the density of hydrogen nuclei needed to show up on MR images, their intersection contours (green) were reconstructed from casts and added computationally to the image. The red lines indicate the outline of a dental plate custom-fitted to the subject's upper and lower teeth and designed to carry a contrast agent used to provide reference landmarks (tiny white dots) [16.82].

At any given point in the vocal tract, it would appear that changes in cross-sectional area depend primarily on the midsagittal distance from the tongue surface to the vocal tract wall. An empirically adequate approach to capturing distance-to-area relations is to use power functions of the form $A_x = \alpha d_x^{\beta}$, where d_x is the mid-sagittal cross distance at a location x in the vocal tract and α and β are constants whose values tend to be different in different regions and depend on the geometry of the vocal tract walls [16.24,70,72]. The cross-sectional area function can be obtained by applying a set of d-to-A rules of this type to cross distances measured along the vocal tract perpendicular to the midline. The final step consists in calculating the acoustic resonance frequencies of the derived area function.

In producing a vowel and the vocal tract shape appropriate for it, what are the main articulatory parameters that the speaker has to control? The APEX model [16.80] assumes that the significant information about the vowel articulations in Fig. 16.26 is the following. All the vowels exhibit tongue shapes with a single constriction. They differ with respect to how narrow this constriction is and where it is located. In other words, the vowels appear to be produced with control of two degrees of freedom: the palatal–pharyngeal dimension (*position*) and tongue height, or, in APEX terminology, *displacement* from neutral. This interpretation is presented in stylized form in Fig. 16.28 together with APEX tongue shapes sampled along the palatal–pharyngeal continuum. The choice of tongue parameters parallels the degree and place of constriction in the area-function models of *Stevens* and *House* [16.83] and *Fant* [16.27].

A useful application of articulatory models is that they can be set up to change a certain variable while keeping others constant [16.84]. Clearly, asking a human subject to follow such an instruction does not create an easily controlled task.

Amazon
EX1013 Page 697

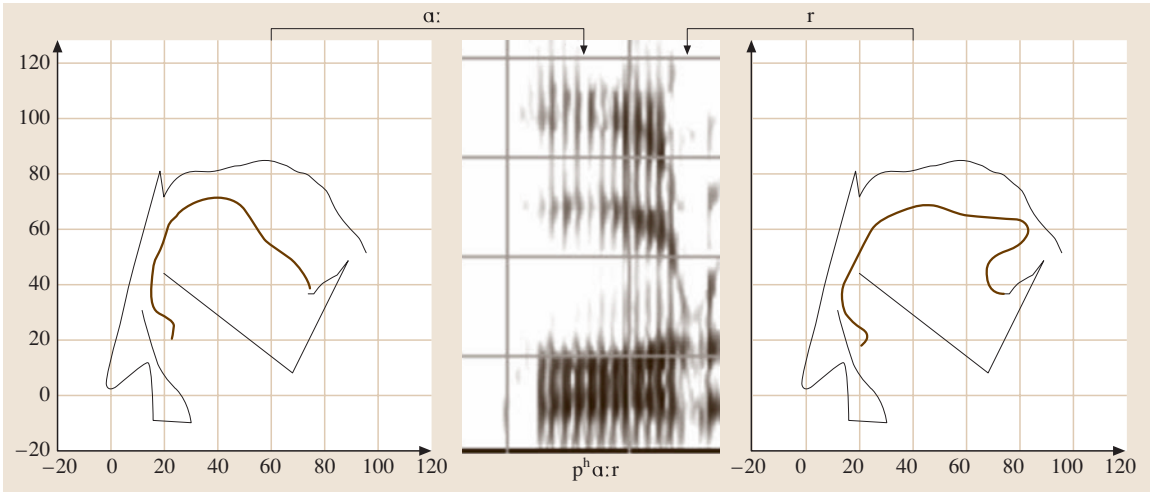


Fig. 16.30 Role of the tongue tip and blade in tuning F3. As the tongue is raised to produce the retroflex configuration required for the Swedish [r] sound an acoustically significant volume is created under the tongue blade. Simulations using APEX indicate that this cavity is responsible for the considerable lowering of the third formant associated with the [r]

Figure 16.29 plots results from APEX simulation experiments. It demonstrates how F1, F2 and F3 vary in response to the tongue moving continuously between palatal, velar and pharyngeal locations and maintaining a minimum constriction area A_{min} of 0.25 cm^2, a 7 mm jaw opening and keeping the larynx height constant. The dashed and solid lines refer to rounded and spread conditions. The calculations included a correction for the impedance of the vocal tract walls ([16.54], [16.28, p. 158]).

It is seen that the tongue movement has strong effect on F2. As the constriction becomes more posterior F2 decreases and F1 rises. In general, rounding lowers formants by varying degrees that depends on the formant-cavity relations that apply to the articulation in question. However, for the most palatal position in Fig. 16.29 – an articulation similar to an [i] or an [y], it has little effect on F2 whereas F3 is affected strongly.

The F2 versus F1 plot of the lower diagram of Fig. 16.29 was drawn to illustrate the effect of varying the jaw. The thin lines show how, at various degrees of jaw opening (6, 7, 8, 9, 12, 15, 19 and 23 mm), the tongue moves between palatal and pharyngeal constrictions by way of the neutral tongue shape. The bold lines pertain to fixed palatal, neutral or pharyngeal tongue contours. We note that increasing the jaw opening while keeping the tongue configuration constant shifts F1 upward.

Figure 16.30 exemplifies the role of the tongue tip and blade in tuning F3. The data come from an X-ray study with synchronized sound [16.79,85]. At the center a spectrogram of the syllable "*par*" [pʰɑ:r]. Perhaps the most striking feature of the formant pattern is the extensive lowering of F3 and F4 into the [r].

The articulatory correlates are illustrated in the tracings. As we compare the profiles for [ɑ:] and [r], we see that the major change is the raising of the tongue blade for [r] and the emergence of a significant cavity in front of, and below, the tongue blade. Simulations us-

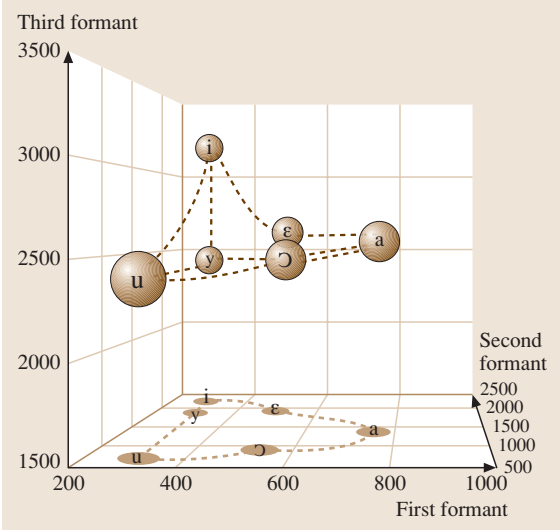


Fig. 16.31 The acoustic vowel space: *possible* vowel formant combinations according to the APEX model

Amazon
EX1013 Page 698

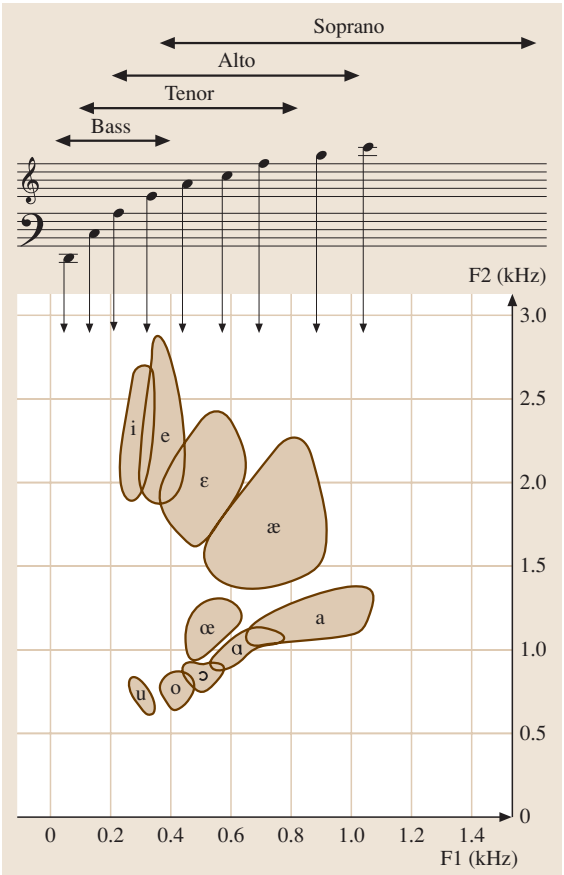


Fig. 16.32 Schematic illustration of the variation ranges of F1 and F2 for various vowels as pronounced by female and male adults. Above the graph F1 is given in musical score notation together with typical F0 ranges for the indicated voice classifications

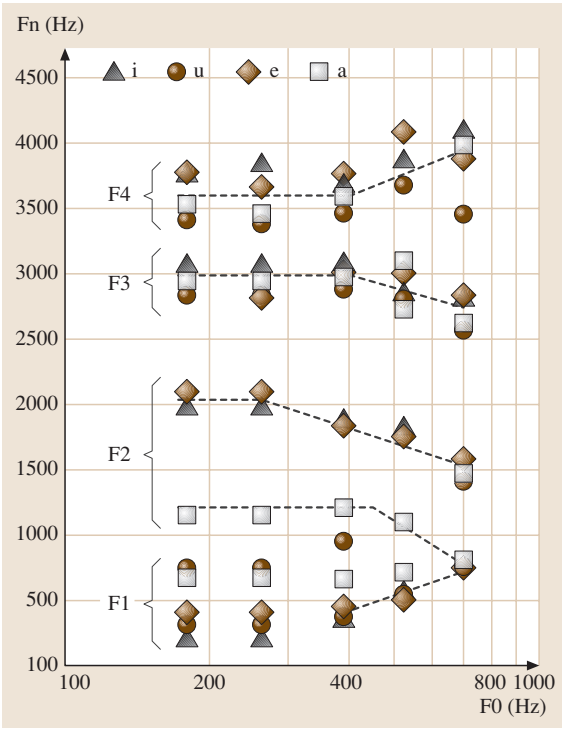


Fig. 16.33 Values of F1, F2, F3, and F4 for the indicated vowels in a professional soprano as function of F0. F1 and F3 are represented by open symbols and F2 and F4 by filled symbols. The values for F0 ≈180 Hz were obtained when the singer sustained the vowels in a speech mode. (After *Sundberg* [16.87])

ing APEX and other models [16.28,86] indicate that this subapical cavity is responsible for the drastic lowering of F3.

The curves of Fig. 16.30 are qualitatively consistent with the nomograms published for three-parameter area-function models, e.g., Fant [16.27]. An advantage of more-realistic physiological models is that the relationships between articulatory parameters and formant patterns become more transparent. We can summarize observations about APEX and the articulation-to-formants mapping as:

- F1 is controlled by the jaw in a direct way;
- F2 is significantly changed by front–back movement of the tongue;
- F3 is influenced by the action of the tongue blade.

Another way of summarizing the acoustic properties of a speech production model is to translate all of its articulatory capabilities into a formant space. By definition that contains all the formant patterns that the model is capable of producing (and specifying articulatorily). Suppose that a model uses n independent parameters and each parameter is quantized into a certain number of steps. By forming all *legal* combinations of those parametric values and deriving their vocal tract shapes and formant patterns, we obtain the data needed to represent that space graphically.

The APEX vowel space is shown in 3-D in Fig. 16.31. The x-axis is F1. The depth dimension is F2 and the vertical axis is F3. Smooth lines were drawn to enclose individual data points (omitted for clarity).

A cloud-like structure emerges. Its projection on the F2/F1 floor plane takes the form of the familiar trian-

Amazon
EX1013 Page 699

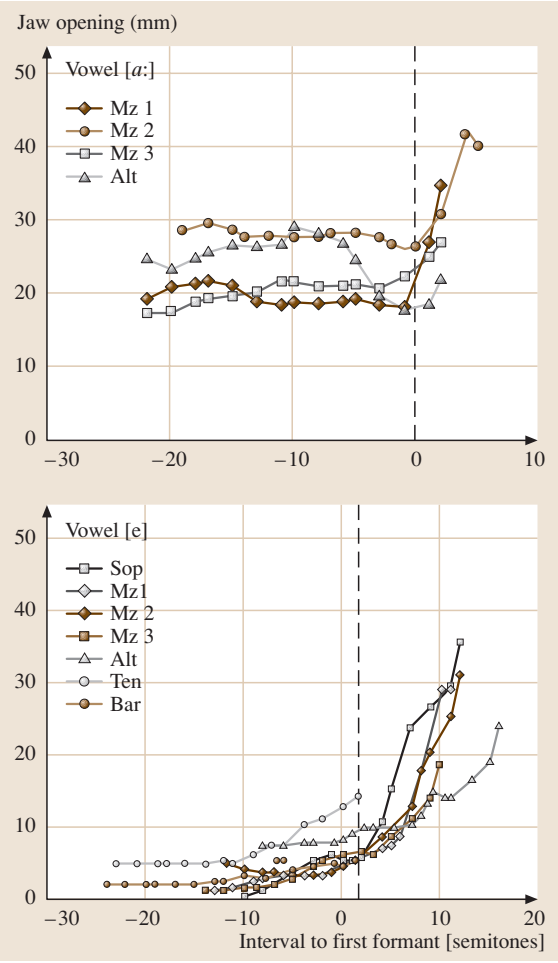


Fig. 16.34 Singers' jaw opening in the vowels [ɑ:] and [e] plotted as functions of the distance in semitones between F0 and the individual singer's normal F1 value for these vowels. The singers belonged to different classifications: Sop = soprano, Mz = mezzosoprano, Alt = alto, Ten = tenor, Bar = baritone. Symbols refer to subjects. (After *Sundberg* and *Skoog* [16.88])

gular pattern appears with [i], [a] and [u] at the corners. Adding F3 along the vertical axis offers additional room for vowel timber variations especially in the [i/y] region.

Figure 16.32 shows typical values for F1 and F2 for various vowels. At the top of the diagram, F1 is given also on the musical staff. The graph thus demonstrates that the fundamental frequency in singing is often higher than the normal value of F1. For example, the first formant of [i:] and [u:] is about 250 Hz (close to the pitch of C4), which certainly is a quite low note for a soprano.

Theory predicts that the vocal fold vibration will be greatly disturbed when F0 = F1 [16.89], and singers seem to avoid allowing F0 to pass F1 [16.87, 90]. Instead they raise F1 to a frequency slightly higher than F0. Some formant frequency values for a soprano singer are shown in Fig. 16.33. The results can be idealized in terms of lines, also shown in the figure, relating F1, F2, F3, and F4 to F0. Also shown are the subject's formant frequencies in speech. The main principle seems to be as follows. As long as fundamental frequency is lower than the normal value of the vowel's first formant frequency, this formant frequency is used. At higher pitches, the first formant is raised to a value somewhat higher than the fundamental frequency. In this way, the singer avoids having the fundamental frequency exceed the first formant frequency. With rising fundamental frequency, F2 of front vowels is lowered, while F2 of back vowels is raised to a frequency just above the second spectrum partial; F3 is lowered, and F4 is raised.

A commonly used articulatory trick for achieving the increase of F1 with F0 is a widening of the jaw opening [16.85]. The graphs of Fig. 16.34 show the jaw opening of professional singers as a function of the frequency separation in semitones between F0 and the F1 value that the singer used at low pitches. The graph referring to the vowel [ɑ:] shows that most singers began to widen their jaw opening when the F0 was about four semitones below the normal F1 value. The lower graph of Fig. 16.34 shows the corresponding data for the vowel [e]. For this vowel, most subjects started to widen the jaw opening when the fundamental was about four semitones above the normal value of F1. It is likely that below this pitch singers increase the first formant by other articulatory means than the jaw opening. A plausible candidate in front vowels is the degree of vocal tract constriction; a reduced constriction increases the first formant. Many singing teachers recommend their students to *release the jaw* or to *give space to the tone*; it seems likely that the acoustic target of these recommendations is to raise F1.

There are also other articulators that can be recruited for the purpose of raising F1. One is the lip opening. By retracting the mouth corners, the vocal tract is shortened; and hence the frequencies of the formants will increase. The vocal tract can also be shortened by raising the larynx, and some professional female singers take advantage of this tool when singing at high pitches.

This principle of tuning formant frequencies depending on F0 has been found in all singers who encounter the situation that the normal value of the first formant is lower than their highest pitches. In fact, all singers except basses encounter this situation at least for some

Amazon
EX1013 Page 700

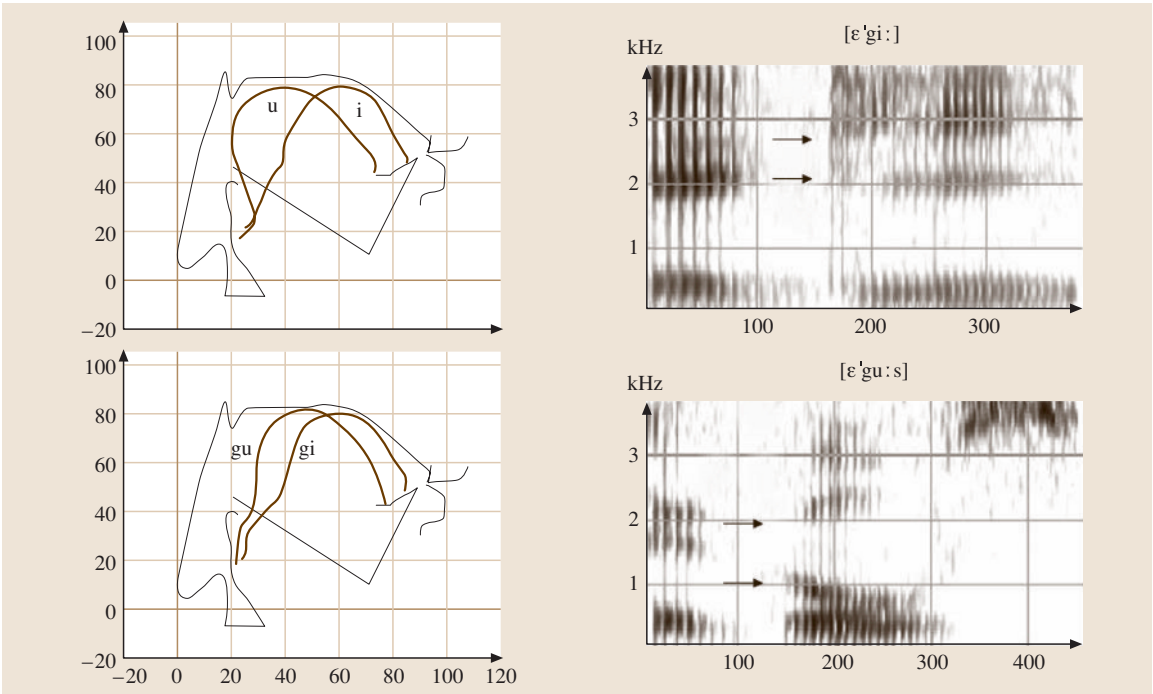


Fig. 16.35 The universal phenomenon of coarticulation illustrated with articulatory and acoustic data on Swedish [g]. The influence of the following vowel extends throughout the articulation of the stop closure. There is no point in time at which a *pure* (context-free) sample of the [g] could be obtained

vowels sung at high pitches. The benefit of these arrangements of the formant frequencies is an enormous increase of sound level, gained by sheer resonance.

Vowel quality is determined mainly by F1 and F2, as mentioned. Therefore, one would expect drastic consequences regarding vowel intelligibility in these cases. However, the vowel quality of sustained vowels seems to survive these pitch-dependent formant frequency changes surprisingly well, except when F0 exceeds the pitch of F5 (about 700 Hz). Above that frequency no formant frequency combination seems to help, and below it, the vowel quality would not be better if normal formant frequencies were chosen. The amount of text intelligibility which occurs at very high pitches relies almost exclusively on the consonants surrounding the vowel. Thus, facing the choice between inaudible tones with normal formant frequencies or audible tones with strange vowel quality, singers probably make a wise choice.

One of the major challenges both for applied and theoretical speech research is the great variability of the speech signal. Consider a given utterance as pronounced by speakers of the same dialect. A few moments' reflection will convince us that this utterance is certain to come in a large variety of physical shapes. Some variations reflect the speaker's age, gender, vocal anatomy as well as emotional and physiological state. Others are stylistic and situational as exemplified by speaking clearly in noisy environments, speaking formally in public, addressing large audiences, chatting with a close friend or talking to oneself while trying to solve a problem. Style and situation make significant contributions to the variety of acoustic waveforms that instantiate what we linguistically judge as the *same utterance*.

In phonetic experimentation investigators aim at keeping all of these stylistic and situational factors constant. This goal tends to limit the scope of the research to a style labeled *laboratory speech*, which consists of test items that are chosen by the experimenter and are typically read from a list. Despite the marked focus on this type of material, laboratory speech nonetheless presents several variability puzzles. We will mention two: segmental interaction, or *coarticulation*, and prosodic modulation. Both exemplify the ubiquitous context dependence of phonetic segments. First a few remarks on coarticulation.

Amazon
EX1013 Page 701

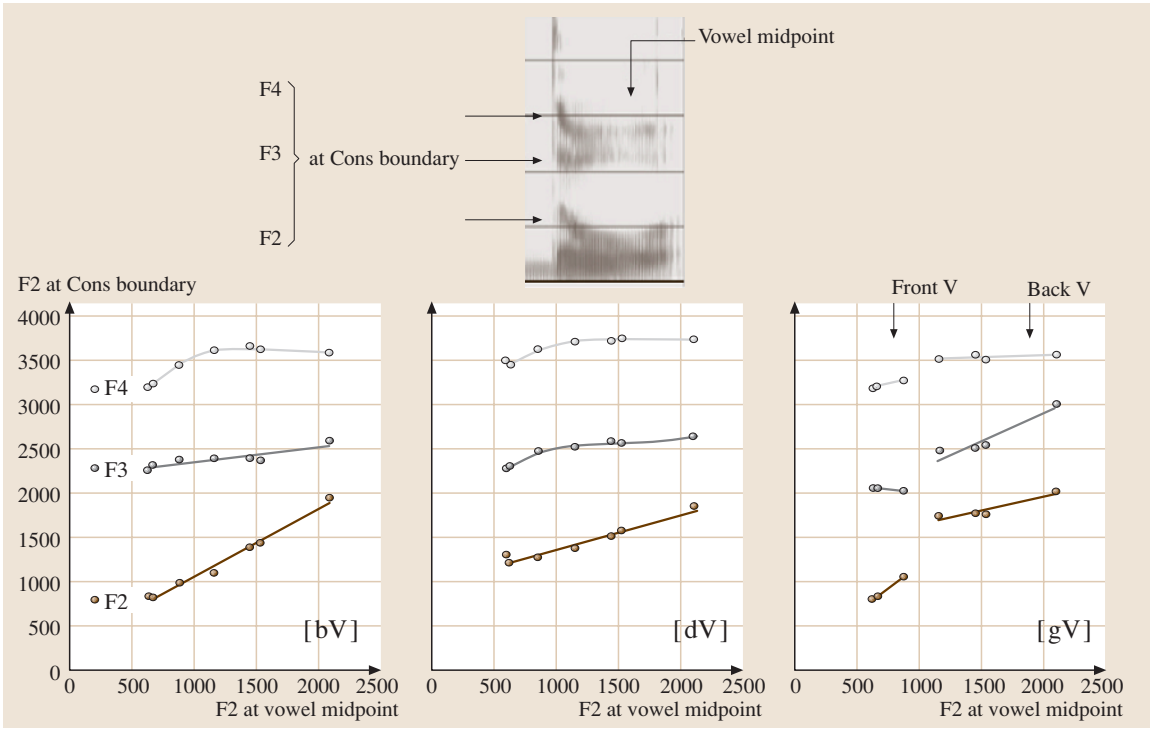


Fig. 16.36 Coarticulation in [bV], [dV] and [gV] as reflected by measurements of F2, F3 and F4 onsets plotted against an acoustic correlate of front back tongue movement, viz., F2 at the vowel midpoint

In Fig. 16.35 the phoneme [g] occurs in two words taken from an X-ray film [16.79, 85] of a Swedish speaker: [ɛˈgiː] and [ɛˈguːs]. In these words the first three phonemes correspond to the first three segments on the spectrogram: an initial [ɛ] segment, the [g] stop gap and then the final vowel. So far, so good. However, if we were to try to draw a vertical line on the spectrogram to mark the point in time where [ɛ] ends and [g] begins, or where [g] ends and the final vowel begins, we would soon realize that we have an impossible task. We would perhaps be able to detect formant movements in the [ɛ] segment indicating that articulatory activity towards the [g] had been initiated. However, that point in time occurs during the acoustic [ɛ] segment. Similarly, we could identify the endpoint of the formant transitions following [g] but that event occurs when the next segment, the final vowel, is already under way.

What we arrive at here is the classical conclusion that strings of phonemes are not organized as *beads on a necklace* [16.91–93]. The acoustic correlates of phonemes, the acoustic segments, are produced according to a motor schema that requires parallel activity in several articulatory channels and weaves the sequence of phonemes into a smooth fabric of overlapping movements. We are talking about *coarticulation*, the overlap of articulatory gestures in space and time.

Not only does this universal of motor organization give rise to the segmentation problem, i. e., make it impossible to chop up the time scale of the speech wave into phoneme-sized chunks, it creates another dilemma known as the *invariance issue*. We can exemplify it by referring to Fig. 16.35 again and the arrows indicating the frequencies of F2 and F3 at the moment of [g] release. With [iː] following they are high in frequency. Next to [uː] they are lower. What acoustic attributes define the [g] phoneme? How do we specify the [g] phoneme acoustically in a context-independent way?

The answer is that, because of coarticulation, we cannot provide an acoustic definition that is context independent. There is no such thing as an acoustic *pure sample* of a phoneme.

Articulatory observations confirm this account. There is no point in time where the shape of the tongue shows zero influence from the surrounding vowels. The tracings at the top left of Fig. 16.35 show the tongue con-

Amazon
EX1013 Page 702

tours for [iː] and [uː] sampled at the vowel midpoints. The bottom left profiles show the tongue shapes at the [g] release. The effect of the following vowel is readily apparent.

So where do we look for the invariant phonetic correlates for [g]? Work on articulatory modeling [16.49, 94] indicates that, if there is such a thing as a single context-free target underlying the surface variants of [g], it is likely to occur at a deeper level of speech production located before the motor commands for [g] closure and for the surrounding vowels blend.

This picture of speech production raises a number of questions about speech perception that have been addressed by a large body of experimental work [16.95, 96] but which will not be reviewed here.

It should be remarked though that, the segmentation and invariance issues notwithstanding, the context sensitivity of phonetic segments is systematic. As an illustration of that point Fig. 16.36 is presented. It shows average data on formant transitions that come from the Swedish speaker of Fig. 16.36 and Figs. 16.26–16.28. The measurements are from repetitions of CV test words in which the consonants were [b], [d] or [g] and were combined with [iː] [ɛː] [æː] [a] [ɑː] [oː] and [uː]. Formant transition onsets for F2, F3 and F4 are plotted against F2 midpoints for the vowels.

If the consonants are coarticulated with the vowels following, we would expect consonant onset patterns to co-vary with the vowel formant patterns. As shown by Fig. 16.36 that is also what we find. Recall that, in the section on articulatory modeling, we demonstrated that F2 correlates strongly with the front–back movement of the tongue. This implies that, in an indirect way, the *x*-axis labeled 'F2 at vowel midpoint' can be said to range from back to front. The same reasoning applies to F2 onsets.

Figure 16.36 shows that the relationship between F2 onsets and F2 at vowel midpoint is linear for bV and dV. For gV, the data points break up into back (low F2) and front (high F2) groups. These straight lines – known as *locus equations* [16.97] – have received considerable attention since they provide a compact way of quantifying coarticulation. Data are available for several languages showing robustly that slopes and intercepts vary in systematic ways with places of articulation.

Furthermore, we see from Fig. 16.36 that lawful patterns are obtained also for F3 and F4 onsets. This makes sense if we assume that vocal tract cavities are not completely uncoupled and that hence, all formants – not only F2 – are to some extent influenced by where along the front–back dimension the vowel is articulated.

16.5 The Syllable

A central unit in both speech and singing is the syllable. It resembles the phoneme in that it is hard to define but it can be described in a number of ways.

Linguists characterize it in terms of how vowels and consonants pattern within it. The central portion, the nucleus, is normally a vowel. One or more consonants can precede and/or follow forming the onset and the coda respectively. The vowel/nucleus is always there; the onset and coda are optional.

Languages vary with respect to the way they combine consonants and vowels into syllables. Most of them favor a frame with only two slots: the CV syllable. Others allow more-elaborated syllable structures with up to three consonants initially and the mirror image in syllable final position. If there is also a length distinction in the vowel and/or consonant system, syllables frames can become quite complex. A rich pattern with consonant clusters and phonological length usually implies that the language has a strong contrast between stressed and unstressed syllables.

In languages that allow consonant sequences, there is a universal tendency for the segments to be serially ordered on an articulatory continuum with the consonants compatible with the vowel's greater jaw opening occurring next to the vowel, e.g., [l] and [r], while those less compatible, e.g. [s], are recruited at the syllable margins [16.98, 99]. In keeping with this observation, English and other languages use [spr] as an initial, but not final, cluster. The reverse sequence [rps] occurs in final, but not initial, position, cf. *sprawl* and *harps*. Traditionally and currently, this trend is explained in terms of an auditory attribute of speech sounds, sonority. The *sonority principle* [16.59] states that, as the most *sonorous* segments, vowels take the central nucleus position of the syllable and that the *sonority* of the surrounding consonants must decrease to the left and to the right starting from the vowel. Recalling that the degree of articulatory opening affects F1 which in turn affects sound intensity, we realize that these articulatory and auditory accounts are not incompatible. However,

Amazon
EX1013 Page 703

the reason for the syllabic variations in sonority is articulatory: the tendency for syllables to alternate close and open articulations in a cyclical manner.

The syllable is also illuminated by a developmental perspective. An important milestone of normal speech acquisition is *canonical babbling*. This type of vocalization makes its appearance sometime between 6–10 months. It consists of sequences of CV-like events, e.g., [dædæ], [baba] [16.101–105]. The phonetic output of deaf infants differs from canonical babbling both quantitatively and quantitatively [16.106–108], suggesting that auditory input from the ambient language is a prerequisite for canonical babbling [16.109–112]. What babbling shares with adult speech is its "syllabic" organization, that is, the alternation of open and close articulations in which jaw movement is a major component [16.113].

As mentioned, the regular repetition of open-close vocal tract states gives rise to an amplitude modulation of the speech waveform. Vowels tend to show the highest amplitudes contrasting with the surrounding consonants which have various degrees of constriction and hence more reduced amplitudes. At the acoustic boundary between a consonant and a vowel, there is often an abrupt rise in the amplitude envelope of the waveform.

When a Fourier analysis is performed on the waveform envelope, a spectrum with primarily low, sub-audio frequency components is obtained. This is to be expected given the fact that amplitude envelopes vary slowly as a function of time. This representation is known as the modulation spectrum [16.114]. It reflects recurring events such as the amplitude changes at consonant-vowel boundaries. It provides an approximate record of the rhythmic pulsating stream of stressed and unstressed syllables.

The time envelope may at first appear to be a rather crude attribute of the signal. However, its perceptual importance should not be underestimated. Room acoustics and noise distort speech by modifying and destroying its modulation spectrum. The modulation transfer function was proposed by Houtgast and Steeneken as a measure of the effect of the auditorium on the speech signal and as a basis for an index, the speech transmission index (STI), used to predict speech intelligibility under different types of reverberation and noise. The success of this approach tells us that the modulation spectrum, and hence the waveform envelope, contains information that is crucial for robust speech perception [16.115]. Experimental manipulation of the temporal envelope has been performed by Drullman et al. [16.116, 117] whose work reinforces the conclusions reached by Houtgast and Steeneken.

There seems to be something special about the front ends of syllables. First, languages prefer CVs to VCs. Second, what children begin with is strings of CV-like pseudosyllables that emulate the syllable onsets of adult speech. Third there is perceptually significant information for the listener in the initial dynamics of the syllable. Let us add another phenomenon to this list: the *syllable beat*, or the syllable's P-center [16.118, 119].

In reading poetry, or in singing, we have a very strong sense that the syllables are spoken/sung in accordance with the rhythmic pattern of the meter. Native speakers agree more or less on how many syllables there are in a word or a phrase. In making such judgments, they seem to experience syllables as unitary events. Although it may take several hundred milliseconds to pronounce, subjectively a syllable appears to occur at a specific moment in time. It is this impression to which the phonetic term *syllable beat* refers and that has been studied experimentally in a sizable number of publications [16.120–126].

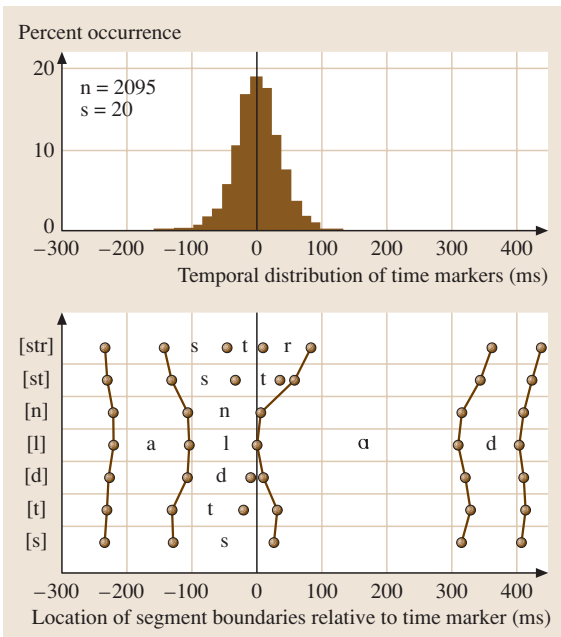


Fig. 16.37 In *Rapp* [16.100] three male Swedish subjects were asked to produce test words based on an [a_ˈɑːd] frame with [s, t, d, l, n, st, str] as intervocalic segment(s). These items were pronounced in synchrony with metronome beats presented over headphones. The figure shows the temporal distribution of the metronome beats (*upper panel*) in relation to the average time location of acoustic segment boundaries (*lower panel*)

Amazon
EX1013 Page 704

Rapp [16.100] asked three native speakers of Swedish to produce random sets of test words built from [aC'ɑːd] were the consonant C was selected from [s, t, d, l, n, st, str]. The instruction was to synchronize the stressed syllable with a metronome beat presented over earphones.

The results are summarized in Fig. 16.37. The *x*-axis represents distance in milliseconds from the point of reference, the metronome beat. The top diagram shows the total distribution of about 2000 time markers around the mean. The lower graph indicates the relative location of the major acoustic segment boundaries.

Several phonetic correlates have been proposed for the syllable beat: some acoustic/auditory, others articulatory. They all hover around the vowel onset, e.g., the amplitude envelope of a signal [16.127], rapid increase in energy in spectral bands [16.128, 129] or the onset of articulatory movement towards the vowel [16.130].

Rapp's data in Fig. 16.37 indicate that the mean beat time tends to fall close to the release or articulatory opening in [t, d, l, n] but that it significantly precedes the acoustic vowel onsets of [str-], [st-] and [s-]. However, when segment boundaries were arranged in relation to a fixed landmark on the F0 contour and vowel onsets were measured relative to that landmark, the range of vowel onsets was reduced. It is possible that the syllable beat may have its origin, not at the acoustic surface, nor at some kinematic level, but in a deeper motor control process that coordinates and imposes coherence on respiratory, phonatory and articulatory activity needed to produce a syllable.

Whatever the definitive explanation for the syllable's psychological *moment of occurrence* will be, the syllable beat provides a useful point of entry for attempts to understand how the control of rhythm and pitch works in speech and singing. Figure 16.38 compares spectrograms of the first few bars of *Over the rainbow*, spoken (left) and sung (right).

Vertical lines have been drawn at vowel onsets and points where articulators begin to move towards a more open configuration. The lines form an isochronous temporal pattern in the sung version which was performed at a regular rhythm. In the spoken example, they occur at intervals that seem more determined by the syllable's degree of prominence.

The subject reaches F0 targets at points near the beats (the vertical lines). From there target frequencies are maintained at stationary values until shortly before it is time to go to the next pitch. Thus the F0 curve resembles a step function with some smoothing applied to the steps.

On the other hand, the F0 contour for speech shows no such steady states. It makes few dramatic moves as it

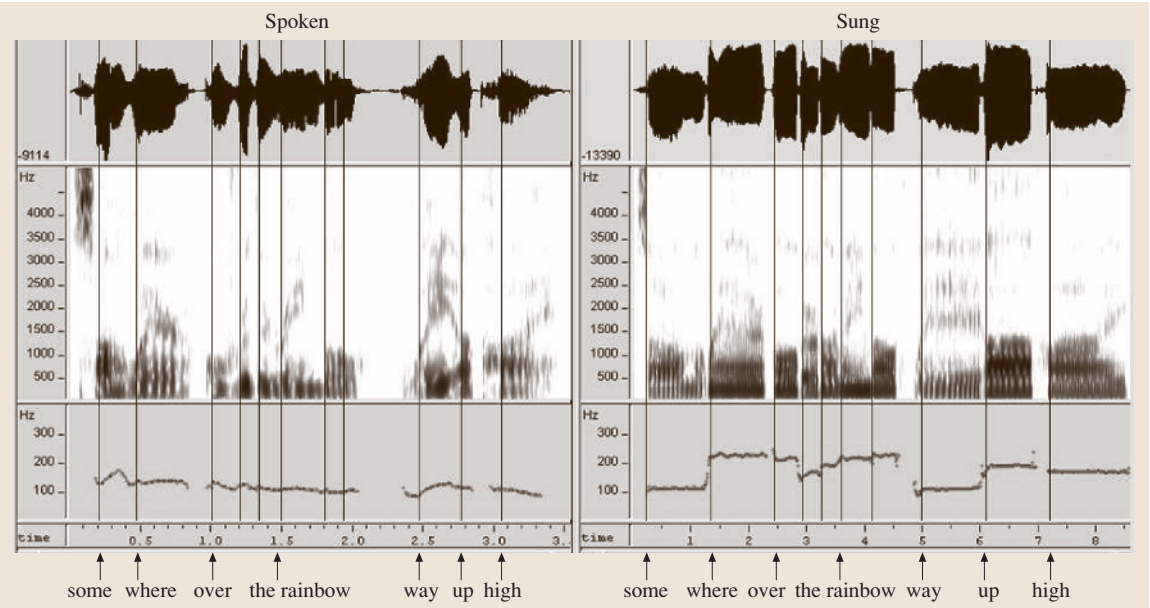


Fig. 16.38 Waveforms and spectrograms of the first few bars of Over the rainbow, spoken (*left*) and sung (*right*). Below: F0 traces in Hz. Vertical lines were drawn at time points corresponding to "vowel onsets" defined in terms of onset of voicing after a voiceless consonant, or the abrupt vocal tract opening following a consonant

Amazon
EX1013 Page 705

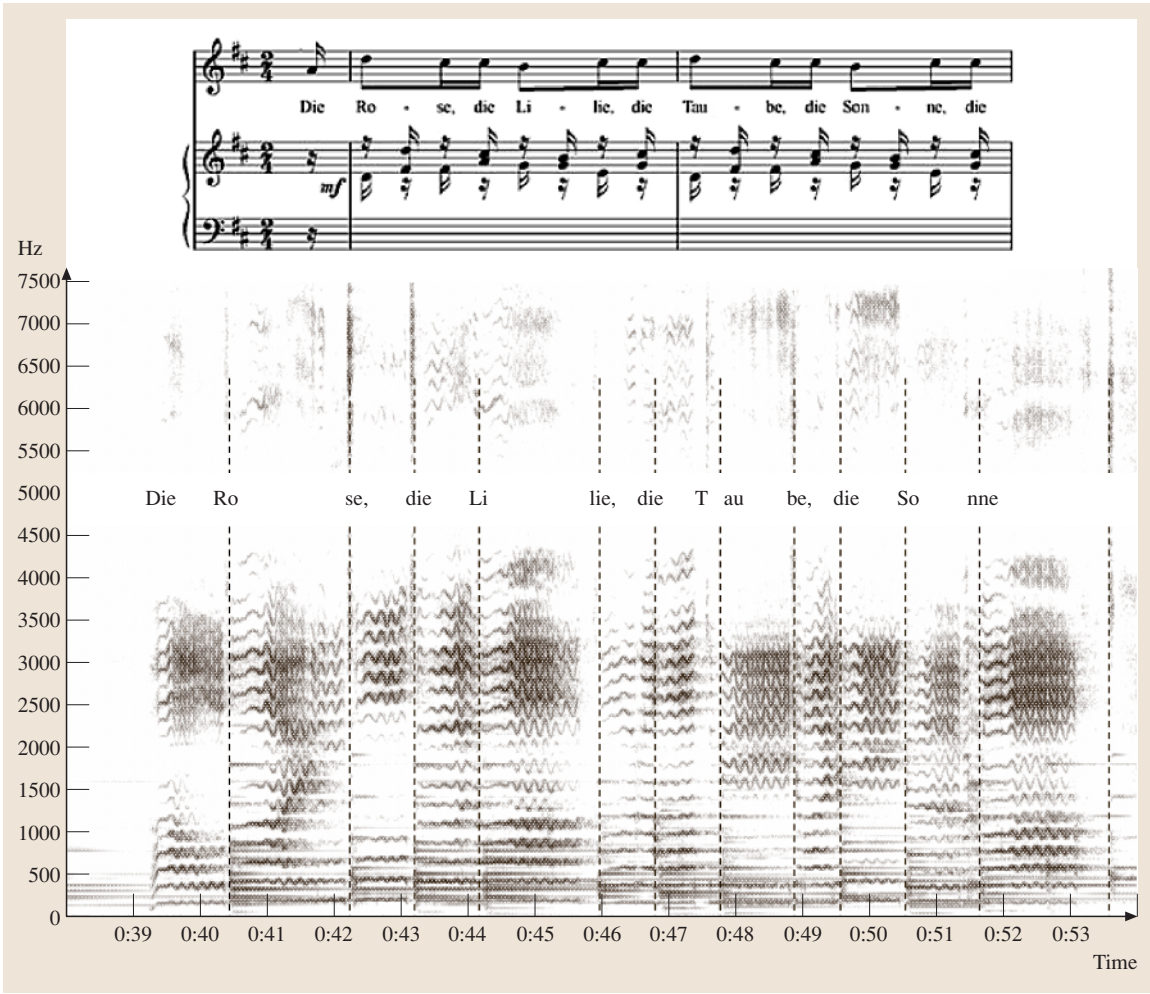


Fig. 16.39 Spectrogram of Dietrich Fischer-Dieskau's recording of song number 4 from Robert Schumann's Dichterliebe, op. 48. The *vertical bars* mark the onset of the tones of the piano accompaniment. Note the almost perfect synchrony between vowel onset and the piano. Note also the simultaneous appearance of high and low partials after the consonants also after the unvoiced /t/ in Taube

gradually drifts downward in frequency (the declination effect).

Figure 16.39 shows a typical example of classical singing, a spectrogram of a commercial recording of Dietrich Fischer-Dieskau's rendering of the song "*Die Rose, die Lilie*" from Robert Schumann's *Dichterliebe*, op. 48. The vertical dashed lines show the onsets of the piano accompaniment. The wavy patterns, often occurring somewhat after the vowel onset, reflect the vibrato. Apart from the vibrato undulation of the partials the gaps in the pattern of harmonics are quite apparent. At these points we see the effect of the more constricted articulations for the consonants. Note the rapid and crisply synchronous amplitude rise in all partials at the end of the consonant segments, also after unvoiced consonants, e.g. the /t/ in *Taube*. These rises are synchronized with the onset of the piano accompaniment tones, thus demonstrating that in singing a beat is marked by the vowel. The simultaneous appearing of low and high partials seems to belong to the characteristics of classical singing as opposed to speech, where the higher partials often arrive with a slight delay after unvoiced consonants. As mentioned before, such consonants are produced with abduction of the vocal folds. To generate

Amazon
EX1013 Page 706

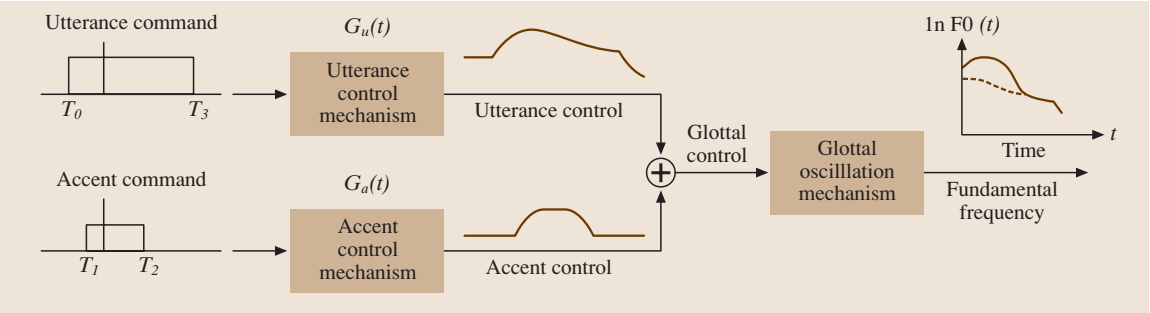


Fig. 16.40 A quantitative approach to synthesizing F0 contours by rule (After *Fujisaki* [16.131]). The framework shown here is an example of the so-called *superposition models* which generate F0 patterns by adding strongly damped responses to input step function target commands for phrase and accents

also high partials in this context, the vocal folds need to close the glottis at the very first vibratory cycle at the vowel onset. A potential benefit of this might be that this enhances the higher formants which are important to text intelligibility.

Several quantitative frameworks have been proposed for generating speech F0 contours by rule (for overview see Frid [16.132]). A subgroup of these has been called *superposition models* [16.131, 133]. A characteristic of such models is that they decompose the F0 curve into separate phrase and accent components. The accent commands are F0 step or impulse functions temporally linked to the syllables carrying stress. Accent pulses are superimposed on the phrase component. For a declarative sentence, the phrase component is a falling F0 contour produced from rectangular step function *hat patterns* [16.134]. The phrase and accent commands are passed through critically damped filters to convert their sum into the smoothly varying output F0 contour. An example of this approach is shown in Fig. 16.40 [16.131].

The last few figures suggest that the F0 patterns of singing and speech may be rather different if compared in terms of the raw F0 curves. However, the superposition models suggest that speech F0 contours have characteristics in part attributable to the passive response characteristics of the neuro-mechanical system that produces them, and in part due to active control signals. These control commands take the form of stepwise changes, some of short, others of longer duration. This representation is not unlike the sequence of F0 targets of a melody.

The implication of this analysis is that F0 is controlled in similar ways in speech and singing in the sense that both are based on sequences of underlying steady-state targets. On the other hand, a significant difference is that in singing high accuracy in attainment of acoustic target frequencies is required whereas in speech such demands are relaxed and smoothing is stronger.

16.6 Rhythm and Timing

A striking characteristic of a foreign language is its rhythm. Phoneticians distinguish between stress-timed and syllable-timed languages. English, Russian, Arabic and Thai are placed in the first group [16.135]. French, Spanish, Greek, Italian, Yoruba and Telugu are examples of the second.

Stress timing means that stressed syllables recur at approximately equal intervals. For instance, it is possible to say "*in ENGlish STRESses reCUR at EQual INtervals*", spacing the stressed syllables evenly in time and without sounding too unnatural. Stress increases syllable duration. In case several unstressed syllables occur in between stresses they are expected to undergo compensatory shortening. Syllable, or *machine-gun*, timing [16.136] implies that syllables recur at approximately equal intervals. In the first case it is the stresses that are isochronous, in the second the syllables.

To what extent are speech rhythms explicitly controlled by the speaker? To what extent are they fortuitous by-products of other factors? Does acquiring a new language involve learning new rhythmic target patterns?

The answer depends on how speech is defined. In the reading of poetry and nursery rhymes it is

Amazon
EX1013 Page 707

clear that an external target, viz. the metric pattern, is a key determinant of how syllables are produced. Similarly, the groupings and the pausing in narrative prose, as read by a professional actor, can exhibit extrinsic rhythmic stylization compared with unscripted speech [16.137].

Dauer [16.138] points the way towards addressing such issues presenting statistics on the most frequently occurring syllable structures in English, Spanish and French. In Spanish and French, syllables were found to be predominantly open, i. e. ending with a vowel, whereas English showed a preference for closed syllables, i. e. ending with a consonant, especially in stressed syllables. Duration measurements made for English and Thai (stress-timed) and Greek, Spanish and Italian (syllable-timed) indicated that the duration of the interstress intervals grew at the same constant rate as a function of the number of syllables between the interstress intervals. In other words, no durational evidence was found to support the distinction between stress timing and syllable timing. How do we square that finding with the widely shared impression that some languages do indeed sound *stress-timed* and others *syllable-timed*?

Dauer [16.138, p. 55] concludes her study by stating that: " . . . *the rhythmic differences we feel to exist between languages such as English and Spanish are more a result of phonological, phonetic, lexical and syntactic facts about that language than any attempt on the part of the speaker to equalize interstress or intersyllable intervals*."

It is clear that speakers are certainly capable of imposing a rhythmic template in the serial read-out of syllables. But do they put such templates in play also when they speak spontaneously? According to *Dauer* and others [16.139, 140] rhythm is not normally an explicitly controlled variable. It is better seen as an emergent product of interaction among the choices languages make (and do not make) in building their syllables: e.g., from open versus closed syllable structures, heavy versus weak clusters, length distinctions and strong stressed/unstressed contrast. We thus learn that the distinction between syllable timing and stress timing may be a useful descriptive term but should primarily be applied to the phonetic output, more seldom to its input control.

Do speech rhythms carry over into music? In other words, would music composed by speakers of syllable-timed or stress-timed languages also be syllable timed or stress timed? The question has been addressed [16.141, 142] and some preliminary evidence has been reported.

There is more to speech timing than what happens inside the syllable. Processes at the word and phrase levels also influence the time intervals between syllables. Consider the English words *digest*, *insult* and *pervert*. As verbs their stress pattern can be described as a sequence of *weak–strong* syllables. As nouns the order is reversed: *strong–weak*. The syllables are the same but changing the word's stress contour (the lexical stress) affects their timing.

The *word length effect* has been reported for several languages [16.58, 143]. It refers to the tendency for the stressed vowel of the word to shorten as more syllables are appended, cf. English *speed*, *speedy*, *speedily*. In Lehiste's formulation: "It appears that in some languages the word as a whole has a certain duration that tends to remain relatively constant, and if the word contains a greater number of segmental sounds, the duration of the segmental sounds decreases as their number in the word increases."

At the phrase level we find that rhythmic patterns can be used to signal differences in syntactic structure. Compare:

1. *The 2000-year-old skeletons*,
2. *The two 1000-year-old skeletons*.

The phrases contain syllables with identical phonetic content but are clearly timed differently. The difference is linked to that fact that in (1) *2000-year-old* forms a single unit, whereas in (2) *two 1000-year-old* consists of two constituents.

A further example is:

3. *Johan greeted the girl with the flowers*.

Hearing this utterance in a specific context a listener might not find it ambiguous. But it has two interpretations: (a) either Johan greeted the girl who was carrying flowers, or (b) he used the flowers to greet her. If spoken clearly to disambiguate, the speaker is likely to provide temporal cues in the form of shortening the syllables within each syntactic group to signal the coherence of the constituent syllables (cf. the word length effect above) and to introduce a short pause between the two groups. There would be a lengthening of the last syllable before the pause and of the utterance-final syllable.

Version (a) can be represented as

4. [*Johan greeted*] # [*the girl with the flowers*].

In (b) the grouping is

5. [*Johan greeted the girl*] # [*with the flowers*].

Amazon
EX1013 Page 708

The # symbol indicates the possibility of a short juncture or pause and a lengthening of the segments preceding the pause. This boundary cue is known as *pre-pausal lengthening*, or more generally as *final lengthening* [16.143, 144], a process that has been observed for a great many languages. It is not known whether this process is a language universal. It is fair to say that is typologically widespread.

Curiously it resembles a phenomenon found in poetry, folk melodies and nursery tunes, called *catalexis*. It consists in omitting the last syllable(s) in a line or other metrical unit of poetry. Instead of four mechanically repeated trochaic feet as in:

|– ⏑ |– ⏑ |– ⏑ |– ⏑ |
|– ⏑ |– ⏑ |– ⏑ |– ⏑ |

we typically find catalectic lines with durationally implied but deleted final syllables as in:

|– ⏑ |– ⏑ |– ⏑ |– |

Old McDonald had a farm

|– ⏑ |– ⏑ |– | |

ee-i ee-i oh

Final lengthening is an essential feature of speech prosody. Speech synthesized without it sounds both highly unnatural and is harder to perceive. In music performance frameworks, instrumental as well as sung, final lengthening serves the purpose of grouping and constituent marking [16.145, 146].

16.7 Prosody and Speech Dynamics

Degree of prominence is an important determinant of segment and syllable duration in English. Figure 16.41 shows spectrograms of the word '*squeal*' spoken with four degrees of stress in sentences read in response to a list of questions (source: [16.147]). The idea behind this method was to elicit tokens having emphatic, strong (as in focus position), neutral and weak (unfocused) stress on the test syllable. The lower row compares the strong and the emphatic pronunciations (left and right respectively). The top row presents syllables with weaker degrees of stress.

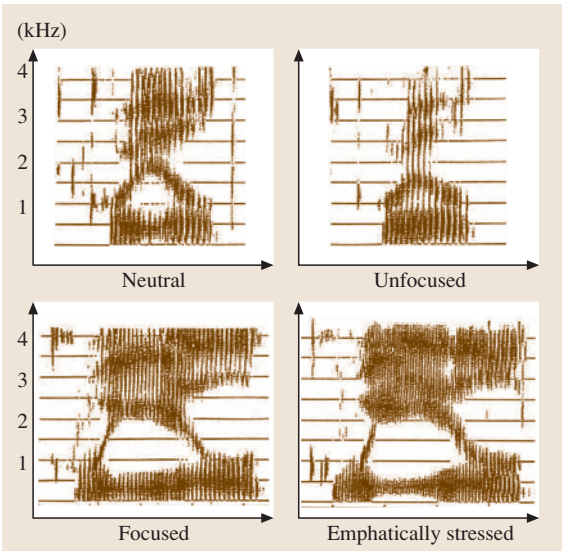


Fig. 16.41 Spectrograms of the word 'squeal' spoken with four degrees of stress in sentences read in response to a list of questions. (After *Brownlee* [16.147])

The representations demonstrate that the differences in stress have a marked effect on the word's spectrographic pattern. Greater prominence makes it longer. Formants, notably F2, show more extensive and more rapid transitions.

A similar experimental protocol was used in two studies of the dynamics of vowel production [16.147, 148]. Both report formant data on the English front vowels [i], [ɪ], [ɛ] and [eɪ] occurring in syllables selected to maximize and minimize contextual assimilation to the place of articulation of the surrounding consonants. To obtain a maximally assimilatory frame, words containing the sequence [w_ɬ] were chosen, e.g., as in *wheel*, *will*, *well* and *wail*. (The [w] is a labio-velar and English [ɬ] is velarized. Both are thus made with the tongue in a retracted position). The minimally assimilatory syllable was [h_d].

Moon's primary concern was *speaking style* (clear versus casual speech). Brownlee investigated changes due to *stress variations*. In both studies measurements were made of vowel duration and extent of formant transitions. Vowel segment boundaries were defined in terms of fixed transition onset and endpoint values for [w] and [ɬ] respectively. The [h_d] frame served as a *null context* reference. The [w_ɬ] environment produced formant transitions large enough to provide a sensitive and robust index of articulatory movement (basically the front–back movement of the tongue).

Amazon
EX1013 Page 709

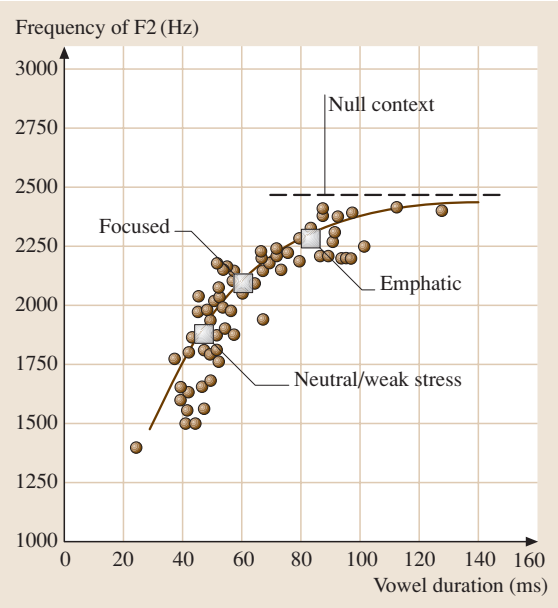


Fig. 16.42 Measurements of F2 as function of vowel duration from tokens of [i] in the word 'squeal' spoken by a male subject under four conditions of stress: emphatically stressed, focused, neutral stress and weak stress (out of focus). *Filled circles* pertain to the individual measurements from all the four stress conditions. The *dashed horizontal line*, is the average F2 value of citation form productions of [h_d]. The averages for the four stress conditions are indicated by the *open squares*. As duration increases – which is approximately correlated with increasing stress – the *square symbols* are seen to move up along the smooth curve implying decreasing undershoot. (After *Brownlee* [16.147])

Figure 16.42 presents measurements of F2 and vowel duration from tokens of [i] in the word '*squeal*' spoken by a male subject under four conditions of stress: emphatically stressed, focused, neutral stress and weak stress (out of focus). The subject was helped to produce the appropriate stress in the right place by reading a question before each test sentence.

Filled circles pertain to individual measurements pooled for all stress conditions. The points form a coherent cluster which is fairly well captured by an exponential curve. The asymptote is the dashed horizontal line, i. e., the average F2 value of the [h_d] data for [i]. The averages for the stress conditions are indicated by the unfilled squares. They are seen to move up along the curve with increasing stress (equivalently, vowel duration).

To understand the articulatory processes underlying the dynamics of vowel production it is useful to adopt a biomechanical perspective. In the [w_ɬ] test words the tongue body starts from a posterior position in [w] and moves towards a vowel target located in the front region of the vocal tract, e.g., [i], [ɪ], [ɛ] or [e]. Then it returns to a back configuration for the dark [ɬ]. At short vowel durations there is not enough time for the vowel gesture to be completed. As a result, the F2 movement misses the reference target by several hundred Hz. Note that in unstressed tokens the approach to target is off by almost an octave. As the syllable is spoken with more stress, and the vowel accordingly gets longer, the F2 transition falls short to a lesser degree. What is illustrated here is the phenomenon known as *formant undershoot* [16.149, 150].

Formant undershoot has received a fair amount of attention in the literature [16.151–155]. It is generally seen as an expression of the sluggish response characteristics of the speech production system. The term *sluggish* here describes both neural delays and mechanical time constants. The response to the neural commands for a given movement is not instantaneous. It takes time for neural impulses to be transformed into muscular contractions and it takes time for the tongue, the jaw and other articulatory structures to respond mechanically to the forces generated by those contractions. In other words, several stages of filtering take place between control signals and the unfolding of the movements. It is this filtering that makes the articulators sluggish. When commands for successive phonemes arrive faster than the articulatory systems are able to respond, the output is a set of incomplete movements. There is undershoot and failure to reach spatial, and thus also acoustic, targets.

However, biomechanics tells us that, in principle, it should be possible to compensate for system characteristics by applying greater force to the articulators thereby speeding up movements and improving target attainment [16.156]. Such considerations make us expect that, in speech movement dynamics, a given trajectory is shaped by

1. the distance between successive articulatory goals (the extent of movement);
2. articulatory effort (input force); and
3. duration (the time available to execute the movement).

The data of Brownlee are compatible with such a mechanism in that the stress-dependent undershoot observed is likely to be a joint consequence of: (1) the large distances between back and front targets; (2) the stress differences corresponding to variations in articulatory effort; and (3) durational limitations.

Amazon
EX1013 Page 710

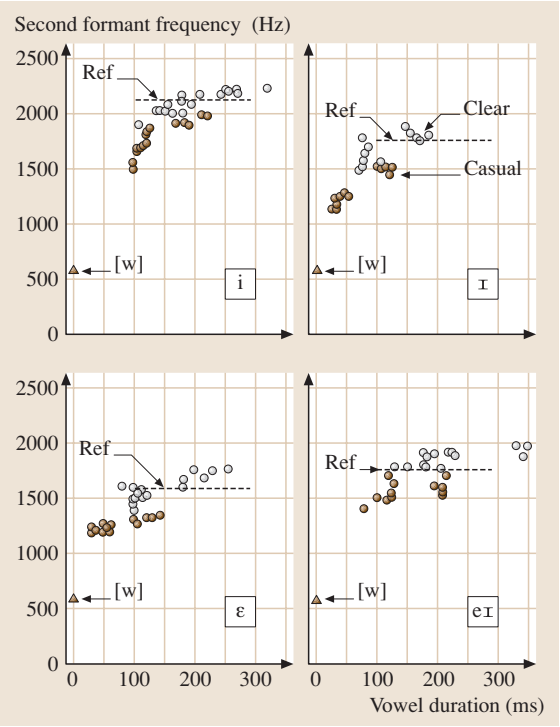


Fig. 16.43 Measurements F2 as function of vowel duration for five male subjects' casual citation-form style pronunciation of the words *wheel*, *wheeling*, *Wheelingham*, *will*, *willing*, *willingly* and for their pronunciation of the same words when asked to say them clearly in an overarticulated way. Casual style is represented by the *solid circles*, clear speech by *open circles*. The frequency of F2 in [w] is indicated by the arrows at 600 Hz. The mean F2 for the [h_d] data is entered as dashed horizontal lines. (After [16.157], Fig. 4)

The three factors – movement extent, effort and duration – are clearly demonstrated in Moon's research on clear and casual speech. As mentioned, Moon's test syllables were the same as Brownlee's. He varied speaking style while keeping stress constant. This was achieved by taking advantage of the word-length effect.The segment strings [wiɬ-], [wɪɬ-], [wɛɬ-] and [weɬ-] were used as the first main-stressed syllable in mono-, bi- and trisyllabic words to produce items such as *wheel*, *wheeling*, *Wheelingham, will*, *willing*, *willingly*, etc.

In the first part of his experiment, five male subjects were asked to produce these words in casual citation-form style. In the second section the instruction was to say them clearly in an *overarticulated* way. For the citation-form pronunciation the results replicated previously observed facts in that: (a) all subjects exhibited duration-dependent formant undershoot, and (b) the magnitude of this effect varied in proportion to the extent of the [w]–[vowel] transition. In the clear style, undershoot effects were reduced. The mechanism by which this was achieved varied somewhat from subject to subject. It involved combinations of increased vowel duration and more rapid and more extensive formant transitions. Figure 16.43 shows the response of one of the subjects to the two tasks [16.157].

Casual style is represented by the solid circles, clear speech by open circles. The frequency of F2 in [w] is indicated by the arrows at 600 Hz. The mean F2 for the [h_d] data is entered as dashed horizontal lines.

The solid points show duration-dependent undershoot effects in relation to the reference values for the [h_d] environment. The open circles, on the other hand, overshoot those values suggesting that this talker used more extreme F2 targets for the clear condition. The center of gravity of the open circles is shifted to the right for all four vowels showing that clear forms were longer than citation forms. Accordingly, this is a subject who used all three methods to decrease undershoot: clear pronunciations exhibited consistently more extreme targets (F2 higher), longer durations and more rapid formant movements.

These results demonstrate that duration and context are not the only determinants of vowel reduction since, for any given duration, the talker is free to vary the degree of undershoot by choosing to articulate more forcefully (as in overarticulated *hyperspeech*) or in a more relaxed manner (as in casual hypospeech). Taken together, the studies by Moon and Brownlee suggest that the dynamics of articulatory movement, and thus of formant transitions, are significantly constrained by three factors: extent of movement, articulatory effort and duration.

16.8 Control of Sound in Speech and Singing

When we perform an action – walk, run, or reach for and manipulate objects – our motor systems are faced with the fact that the contexts under which movements are made are never exactly the same. They change significantly from one situation to the next. Nonetheless, motor systems adapt effortlessly to the continually changing

Amazon
EX1013 Page 711

conditions presumably because, during evolution, they were shaped by the need to cope with unforeseen events and obstacles [16.158]. Their default mode of operation is compensatory [16.159].

Handwriting provides a good illustration of this ability. A familiar fact is that it does not matter if something is written on the blackboard or on a piece of paper. The characteristic features of someone's handwriting are nevertheless easily recognized. Different sets of muscles are recruited and the size of the letters is different but their shapes remain basically similar. What this observation tells us is that movements are not specified in terms of fixed set of muscles and constant contraction patterns. They are recruited in functionally defined groups, *coordinative structures* [16.160]. They are planned and executed in an external coordinate space, in other words in relation to the 3-D world in which they occur so as to attain goals defined outside the motor system itself. The literature on motor mechanisms teaches us that voluntary movements are prospectively organized or future-oriented.

Speech and singing provide numerous examples of this output-oriented mode of motor control [16.161–163]. Earlier in the chapter we pointed out that, in upright position, the diaphragm and adjacent structures are influenced by gravity and tend to be pulled down, thereby causing the volume of the thoracic cavity to increase. In this position, gravity contributes to the inspiratory forces. By contrast, in the supine position, the diaphragm tends to get pushed up into the rib cage, which promotes expiration [16.1, p. 40].

Sundberg et al. [16.164] investigated the effect of upright and supine positions in two baritone singers using synchronous records of esophageal and gastric pressure, EMG from inspiratory and expiratory muscles, lung volume and sound. Reorganization of respiratory activity was found and was interpreted as compensation for the different mechanical situations arising from the upright and supine conditions.

This finding is closely related to what we know about breathing during speech. As mentioned above (Fig. 16.6), the Ps tends to stay fairly constant for any given vocal effort. It is known that this result is achieved by tuning the balance between inspiratory and expiratory muscles. When the lungs are expanded so that the effect of elastic recoil creates a significant expiration force, inspiratory muscles predominate to put a brake on that force. For reduced lung volumes the situation is the opposite. The effect of elastic recoil is rather to increase lung volume. Accordingly, the muscle recruitment needed to maintain Ps is expected to be primarily expiratory. That is indeed what the data show [16.8, 66, 165].

The bite-block paradigm offers another speech example. In one set of experiments [16.166] subjects were instructed to pronounce syllables consisting only of a long vowel under two conditions: first normally, then with a bite block (BB) between the teeth. The subjects all non-phoneticians were told to try to sound as normal as possible despite the BB. No practice was allowed. The purpose of the BB was to create an abnormally large jaw opening for close vowels such as [i] and [u]. It was argued that, if no tongue compensation occurred, this large opening would drastically change the area function of the close vowels and disrupt their formant patterns. In other words, the question investigated was whether the subjects were able to sound normal despite the BB.

Acoustic recordings were made and formant pattern data were collected for comparisons between conditions, vowels and subjects. The analyses demonstrated clearly that subjects were indeed able to produce normal sounding vowels in spite of the BB. At the moment of the first glottal pulse formant patters were well within the normal ranges of variation.

In a follow-up X-ray investigation [16.167] it was found that compensatory productions of [i] and [u] were made with super-palatal and super-velar tongue shapes. In other words, tongue bodies were raised high above normal positions so as to approximate the normal cross-sectional area functions for the test vowels.

Speech and singing share a lot of features but significant differences are brought to light when we consider what the goals of the two behaviors are.

It is more than 50 years since the sound spectrograph became commercially available [16.91]. In this time we have learned from visible speech displays and other records that the relation between phonetic units and the acoustic signal is extremely complex [16.168]. The lesson taught is that invariant correlates of linguistic categories are not readily apparent in the speech wave [16.169]. In the preceding discussion we touched on some of the sources of this variability: coarticulation, reduction and elaboration, prosodic modulation, stylistic, situational and speaker-specific characteristics. From this vantage point it is astonishing to note that, even under noisy conditions, speech communication is a reliable and robust process. How is this remarkable fact to be accounted for?

In response to this problem a number of ideas and explanatory frameworks have been proposed. For instance, it has been suggested that acoustic invariants are rela-

Amazon
EX1013 Page 712

tional rather than absolute (à la tone intervals defined as frequency ratios).

Speech is rather a set of movements made audible than a set of sounds produced by movements [16.14, p. 33].

In keeping with this classical statement, many investigators have argued that speech entities are to be found at some upstream speech production level and should be defined as *gestures* [16.170–173].

At the opposite end, we find researchers (e.g., Perkell [16.163]) who endorse Roman Jakobson's view which, in the search for units gives primacy to the perceptual representation of speech sounds. To Jakobson the stages of the speech chain form an "... *operational hierarchy of levels of decreasing pertinence: perceptual, aural, acoustical and articulatory (the latter carrying no direct information to the receiver)*." [16.174]

These viewpoints appear to conflict and do indeed divide the field into those who see speech as a motoric code (the *gesturalist* camp) and those who maintain that it is primarily shaped by perceptual processes (the *listener-oriented* school).

There is a great deal of experimental evidence for both sides, suggesting that this dilemma is not an either–or issue but that both parties offer valuable complementary perspectives.

A different approach is taken by the H & H (Hyper and Hypo) theory [16.175, 176]. This account is developed from the following key observations about speaker–listener interaction:

1. Speech perception is always a product of signal information and listener knowledge;
2. Speech production is adaptively organized.

Here is an experiment that illustrates the first claim about perceptual processing. Two groups of subjects listen to a sequence of two phrases: a question followed by an answer. The subject groups hear different questions but a single physically identical reply. The subjects' task is to say how many words the reply contains.

The point made here is that Group 1 subjects hear [lɛsn̩ faɪv] as "*less than five*". Those in Group 2 interpret it as "*lesson five*". The first group's response is "*three words*", and the answer of the second is "*two words*". This is *despite the fact* that physically the [lɛsn̩ faɪv] stimulus is exactly the same. The syllabic [n̩] signals the word *than* in one case and the syllable *–on* in the other. Looking for the invariant correlates of the initial consonant *than* is doomed to failure because of the severe degree of reduction.

To proponents of H & H theory this is not an isolated case. This is the way that perception in general works. The speech percepts can never be raw records of the signal because listener knowledge will inevitably interact with the stimulus and will contribute to shaping the percept.

Furthermore, H & H theory highlights the fact that spoken messages show a non-uniform distribution of information in that predictability varies from situation to situation, from word to word and from syllable to syllable. Compare (a) and (b) below. What word is likely to be represented by the gap?

1. "*The next word is* __."
2. "*A bird in the hand is worth two in the* __."

Any word can be expected in (1) whereas in (2) the predictability of the word "*bush*" is high.

H & H theory assumes that, while learning and using their native language, speakers develop a sense of this informational dynamics. Introducing the abstraction of an *ideal speaker*, it proposes that the talker estimates the running contribution that signal-complementary information (listener knowledge) will make during the course of the utterance and then tunes his articulatory performance to the presumed short-term listener needs. Statistically, this type of behavior has the long-term consequence of distributing phonetic output forms along a continuum with clear and elaborated forms (hyperspeech) at one end and casual and reduced pronunciations (hypospeech) at the other.

The implication for the invariance issue is that the task of the speaker is not to encode linguistic units as physical invariants, but to make sure that the signal attributes (of phonemes, syllables, words and phrases) carry discriminative power *sufficient* for successful lexical access. To make the same point in slightly different terms, the task of the signal is not to embody phonetic patterns of constancy but to provide *missing information*.

It is interesting to put this account of speech processes next to what we know about singing. Experimental observations indicate that singing in tune is not simply about invariant F0 targets. F0 is affected by the singer's expressive performance [16.177] and by the tonal context in which a given note is embedded [16.178]. Certain deviations from nominal target frequencies are not unlike the coarticulation and undershoot effects ubiquitously present in speech. Accordingly, with respect to frequency control, speaking and singing are qualitatively similar. However, quantitatively, they differ drastically. Recall that in describing the dynamics of vowel reduction we noted that formant fre-

Amazon
EX1013 Page 713

quencies can be displaced by as much as 50% from target values. Clearly, a musician or singer with a comparable under-/overshoot record would be recommended to embark on an alternative career, the margins of perceptual tolerance being much narrower for singing.

What accounts for this discrepancy in target attainment? Our short answer is that singing or playing *out of tune* is a bad thing. Where does this taboo come from? From consonance and harmony constraints. In simplified terms, an arbitrary sample of tonal music can be analyzed into a sequence of chords. Its melodic line is a rhythmic and tonal elaboration of this harmonic structure. Statistically, long prominent melody tones tend to attract chord notes. Notes of less prominence typically *interpolate* along scales in smaller intervals between the metrically heavier chord notes. The notion of *consonance* goes a long way towards explaining why singing or playing out of tune is tabooed in music: hitting the right pitch is required by the consonance and harmony constraints expected by the listener and historically presumably linked with a combination of the polyphony in our Western tradition of music composition and the fact that most of our music instruments produce tones with harmonic spectra. This implies that intervals departing too much from just intonation will generate beats between simultaneously sounding and only nearly coinciding partials, particularly if the fundamental frequencies are constant, lacking vibrato and flutter.

16.9 The Expressive Power of the Human Voice

The human voice is an extremely expressive instrument both when used for speech and for singing. By means of subtle variations of timing and pitch contour speakers and singers add a substantial amount of expressiveness to the linguistic or musical content and we are quite skilled in deciphering this information. Indeed a good deal of vocal artistry seems to lie in the artist's skill in making nothing but such changes of pitch, timbre, loudness and timing that a listener can perceive as carrying some meaning.

We perceive the extra-linguistic or expressive information in speech and singing in various shapes. For example, we can interpret certain combinations of acoustic characteristics in speech in terms of a smile or a particular forming of the lips on the face of the speaker [16.179]. For example *Fónagy* [16.180] found that listeners were able to replicate rather accurately the facial expression of speakers only by listening to their voices.

The emotive transforms in speech and singing seem partly similar and sometimes identical [16.181, 182]. Final lengthening, mentioned above, uses the same code for marking the end of a structural element, such as a sentence in speech or a phrase in sung and played music performance. Emphasis by delayed arrival is another example, i. e., delaying an emphasized stressed syllable or note by lengthening the unstressed syllable/note preceding it [16.183].

The expressive potential of the human voice is indeed enormous, and would transpire from the ubiquitous use of the voice for the purpose of communication. Correct interpretation of the extra-linguistic content of a spoken utterance is certainly important in our daily life, so we are skilled in deciphering vocal signals also along those dimensions. The importance of correct encoding of the extra-linguistic implies that speakers acquire a great skill in this respect. This skill would be the basic requirement for vocal art, in singing as well as in acting.

References

16.1 T.J. Hixon: *Respiratory Function in Speech and Song* (Singular, San Diego 1991) pp. 1–54

16.2 A. L. Winkworth, P. J. Davis, E. Ellis, R. D. Adams: Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors, JSHR **37**, 535–556 (1994)

16.3 M. Thomasson: *From Air to Aria*, Ph.D. Thesis (Music Acoustics, KTH 2003)

16.4 B. Conrad, P. Schönle: Speech and respiration, Arch. Psychiat. Nervenkr. **226**, 251–268 (1979)

16.5 M.H. Draper, P. Ladefoged, D. Whitteridge: Respiratory muscles in speech, J. Speech Hear. Disord. **2**, 16–27 (1959)

16.6 R. Netsell: Speech physiology. In: *Normal aspects of speech, hearing, and language*, ed. by P.D. Minifie, T.J. Hixon, P. Hixon, P. Williams (Prentice-Hall, Englewood Cliffs 1973) pp. 211–234

Amazon
EX1013 Page 714

16.7 P. Ladefoged, M.H. Draper, D. Whitteridge: Syllables and stress, Misc. Phonetica **3**, 1–14 (1958)

16.8 P. Ladefoged: Speculations on the control of speech. In: *A Figure of Speech: A Festschrift for John Laver*, ed. by W.J. Hardcastle, J. Mackenzie Beck (Lawrence Erlbaum, Mahwah 2005) pp. 3–21

16.9 T.J. Hixon, G. Weismer: Perspectives on the Edinburgh study of speech breathing, J. Speech Hear. Disord. **38**, 42–60 (1995)

16.10 S. Nooteboom: The prosody of speech: melody and rhythm. In: *The Handbook of Phonetic Sciences*, ed. by W.J. Hardcastle, J. Laver (Blackwell, Oxford 1997) pp. 640–673

16.11 M. Rothenberg: *The breath-stream dynamics of simple-released-plosive production Bibliotheca Phonetica 6* (Karger, Basel 1968)

16.12 D.H. Klatt, K.N. Stevens, J. Mead: Studies of articulatory activity and airflow during speech, Ann. NY Acad. Sci. **155**, 42–55 (1968)

16.13 J.J. Ohala: Respiratory activity in speech. In: *Speech production and speech modeling*, ed. by W.J. Hardcastle, A. Marchal (Dordrecht, Kluwer 1990) pp. 23–53

16.14 R.H. Stetson: *Motor Phonetics: A Study of Movements in Action* (North Holland, Amsterdam 1951)

16.15 P. Ladefoged: Linguistic aspects of respiratory phenomena, Ann. NY Acad. Sci. **155**, 141–151 (1968)

16.16 L.H. Kunze: Evaluation of methods of estimating sub-glottal air pressure muscles, J. Speech Hear. Disord. **7**, 151–164 (1964)

16.17 R. Leanderson, J. Sundberg, C. von Euler: Role of the diaphragmatic activity during singing: a study of transdiaphragmatic pressures, J. Appl. Physiol. **62**, 259–270 (1987)

16.18 J. Sundberg, N. Elliot, P. Gramming, L. Nord: Short-term variation of subglottal pressure for expressive purposes in singing and stage speech. A preliminary investigation, J. Voice **7**, 227–234 (1993)

16.19 J. Sundberg: Synthesis of singing, in Musica e Technologia: Industria e Cultura per lo Sviluppo del Mezzagiorno. In: *Proceedings of a symposium in Venice*, ed. by R. Favaro (Unicopli, Milan 1987) pp. 145–162

16.20 J. Sundberg: Synthesis of singing by rule. In: *Current Directions in Computer Music Research, System Development Foundation Benchmark Series*, ed. by M. Mathews, J. Pierce (MIT, Cambridge 1989), 45–55 & 401–403

16.21 J. Molinder: *Akustiska och perceptuella skillnader mellan röstfacken lyrisk och dramatisk sopran, unpublished thesis work* (Lund Univ. Hospital, Dept of Logopedics, Lund 1997)

16.22 T. Baer: Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes, J. Acoust. Soc. Am. **65**, 1271–1275 (1979)

16.23 T. Cleveland, J. Sundberg: Acoustic analyses of three male voices of different quality. In: *SMAC 83. Proceedings of the Stockholm Internat Music Acoustics Conf*, Vol. 1, ed. by A. Askenfelt, S. Felicetti, E. Jansson, J. Sundberg (Roy. Sw. Acad. Music, Stockholm 1985) pp. 143–156, No. 46:1

16.24 J. Sundberg, C. Johansson, H. Willbrand, C. Ytterbergh: From sagittal distance to area, Phonetica **44**, 76–90 (1987)

16.25 I.R. Titze: Phonation threshold pressure: A missing link in glottal aerodynamics, J. Acoust. Soc. Am. **91**, 2926–2935 (1992)

16.26 I.R. Titze: *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs 1994)

16.27 G. Fant: *Acoustic theory of speech production* (Mouton, The Hague 1960)

16.28 K.N. Stevens: *Acoustic Phonetics* (MIT, Cambridge 1998)

16.29 M. Hirano: *Clinical Examination of Voice* (Springer, New York 1981)

16.30 M. Rothenberg: A new inversefiltering technique for deriving the glottal air flow waveform during voicing, J. Acoust. Soc. Am. **53**, 1632–1645 (1973)

16.31 G. Fant: The voice source – Acoustic modeling. In: *STL/Quart. Prog. Status Rep. 4* (Royal Inst. of Technology, Stockholm 1982) pp. 28–48

16.32 C. Gobl: *The voice source in speech communication production and perception experiments involving inverse filtering and synthesis. D.Sc. thesis* (Royal Inst. of Technology (KTH), Stockholm 2003)

16.33 G. Fant, J. Liljencrants, Q. Lin: A four-parameter model of glottal flow. In: *STL/Quart. Prog. Status Rep. 4, Speech, Music and Hearing* (Royal Inst. of Technology, Stockholm 1985) pp. 1–13

16.34 D.H. Klatt, L.C. Klatt: Analysis, synthesis and perception of voice quality variations among female and male talkers, J. Acoust. Soc. Am. **87**(2), 820–857 (1990)

16.35 M. Ljungqvist, H. Fujisaki: A comparative study of glottal waveform models. In: *Technical Report of the Institute of Electronics and Communications Engineers*, Vol. EA85-58 (Institute of Electronics and Communications Engineers, Tokyo 1985) pp. 23–29

16.36 A.E. Rosenberg: Effect of glottal pulse shape on the quality of natural vowels, J. Acoust. Soc. Am. **49**, 583–598 (1971)

16.37 M. Rothenberg, R. Carlson, B. Granström, J. Lindqvist-Gauffin: A three- parameter voice source for speech synthesis. In: *Proceedings of the Speech Communication Seminar 2*, ed. by G. Fant (Almqvist & Wiksell, Stockholm 1975) pp. 235–243

16.38 K. Ishizaka, J.L. Flanagan: Synthesis of voiced sounds from a two-mass model of the vocal cords, The Bell Syst. Tech. J. **52**, 1233–1268 (1972)

16.39 Liljencrants: Chapter A translating and rotating mass model of the vocal folds. In: *STL/Quart. Prog. Status Rep. 1, Speech, Music and Hearing* (Royal Inst. of Technology, Stockholm 1991) pp. 1–18

16.40 A. Ní Chasaide, C. Gobl: Voice source variation. In: *The Handbook of Phonetic Sciences*, ed. by

Amazon
EX1013 Page 715

W.J. Hardcastle, J. Laver (Blackwell, Oxford 1997) pp. 427–462

16.41 E.B. Holmberg, R.E. Hillman, J.S. Perkell: Glottal air flow and pressure measurements for loudness variation by male and female speakers, J. Acoust. Soc. Am. **84**, 511–529 (1988)

16.42 J.S. Perkell, R.E. Hillman, E.B. Holmberg: Group differences in measures of voice production and revised values of maximum airflow declination rate, J. Acoust. Soc. Am. **96**, 695–698 (1994)

16.43 J. Gauffin, J. Sundberg: Spectral correlates of glottal voice source waveform characteristics, J. Speech Hear. Res. **32**, 556–565 (1989)

16.44 J. Svec, H. Schutte, D. Miller: On pitch jumps between chest and falsetto registers in voice: Data on living and excised human larynges, J. Acoust. Soc. Am. **106**, 1523–1531 (1999)

16.45 J. Sundberg, M. Andersson, C. Hultqvist: Effects of subglottal pressure variation on professional baritone singers voice sources, J. Acoust. Soc. Am. **105**, 1965–1971 (1999)

16.46 J. Sundberg, E. Fahlstedt, A. Morell: Effects on the glottal voice source of vocal loudness variation in untrained female and male subjects, J. Acoust. Soc. Am. **117**, 879–885 (2005)

16.47 P. Sjölander, J. Sundberg: Spectrum effects of subglottal pressure variation in professional baritone singers, J. Acoust. Soc. Am. **115**, 1270–1273 (2004)

16.48 P. Branderud, H. Lundberg, J. Lander, H. Djamshidpey, I. Wäneland, D. Krull, B. Lindblom: *X-ray analyses of speech: Methodological aspects, Proc. of 11th Swedish Phonetics Conference* (Stockholm Univ., Stockholm 1996) pp. 168–171

16.49 B. Lindblom: A numerical model of coarticulation based on a Principal Components analysis of tongue shapes. In: *Proc. 15th Int. Congress of the Phonetic Sciences*, ed. by D. Recasens, M. Josep Solé, J. Romero (Universitat Autònoma de Barcelona, Barcelona 2003), CD-ROM

16.50 G.E. Peterson, H. Barney: Control methods used in a study of the vowels, J. Acoust. Soc. Am. **24**, 175–184 (1952)

16.51 Hillenbrand et al.: Acoustic characteristics of American English vowels, J. Acoust. Soc. Am. **97**(5), 3099–3111 (1995)

16.52 G. Fant: Analysis and synthesis of speech processes. In: *Manual of Phonetics*, ed. by B. Malmberg (North-Holland, Amsterdam 1968) pp. 173–277

16.53 G. Fant: Formant bandwidth data. In: *STL/Quart. Prog. Status Rep. 7* (Royal Inst. of Technology, Stockholm 1962) pp. 1–3

16.54 G. Fant: Vocal tract wall effects, losses, and resonance bandwidths. In: *STL/Quart. Prog. Status Rep. 2–3* (Royal Inst. of Technology, Stockholm 1972) pp. 173–277

16.55 A.S. House, K.N. Stevens: Estimation of formant bandwidths from measurements of transient response of the vocal tract, J. Speech Hear. Disord. **1**, 309–315 (1958)

16.56 O. Fujimura, J. Lindqvist: Sweep-tone measurements of vocal-tract characteristics, J. Acoust. Soc. Am. **49**, 541–558 (1971)

16.57 I. Lehiste, G.E. Peterson: Vowel amplitude and phonemic stress in American English, J. Acoust. Soc. Am. **3**, 428–435 (1959)

16.58 I. Lehiste: *Suprasegmentals* (MIT Press, Cambridge 1970)

16.59 O. Jespersen: *Lehrbuch der Phonetik* (Teubner, Leipzig 1926)

16.60 T. Bartholomew: A physical definition of good voice quality in the male voice, J. Acoust. Soc. Am. **6**, 25–33 (1934)

16.61 J. Sundberg: Production and function of the singing formant. In: *Report of the eleventh congress Copenhagen 1972 (Proceedings of the 11th international congress of musicology)*, ed. by H. Glahn, S. Sörensen, P. Ryom (Wilhelm Hansen, Copenhagen 1972) pp. 679–686

16.62 J. Sundberg: Articulatory interpretation of the 'singing formant', J. Acoust. Soc. Am. **55**, 838–844 (1974)

16.63 J. Sundberg: Level and center frequency of the singer´s formant, J. Voice. **15**, 176–186 (2001)

16.64 G. Berndtsson, J. Sundberg: Perceptual significance of the center frequency of the singers formant, Scand. J. Logopedics Phoniatrics **20**, 35–41 (1995)

16.65 L. Dmitriev, A. Kiselev: Relationship between the formant structure of different types of singing voices and the dimension of supraglottal cavities, Fol. Phoniat. **31**, 238–41 (1979)

16.66 P. Ladefoged: *Three areas of experimental phonetics* (Oxford Univ. Press, London 1967)

16.67 J. Barnes, P. Davis, J. Oates, J. Chapman: The relationship between professional operatic soprano voice and high range spectral energy, J. Acoust. Soc. Am. **116**, 530–538 (2004)

16.68 M. Nordenberg, J. Sundberg: Effect on LTAS on vocal loudness variation, Logopedics Phoniatrics Vocology **29**, 183–191 (2004)

16.69 R. Weiss, W.S. Brown, J. Morris: Singer's formant in sopranos: Fact or fiction, J. Voice **15**, 457–468 (2001)

16.70 J.M. Heinz, K.N. Stevens: On the relations between lateral cineradiographs, area functions, and acoustics of speech. In: *Proc. Fourth Int. Congress on Acoustics*, Vol. 1a (1965), paper A44

16.71 C. Johansson, J. Sundberg, H. Willbrand: X-ray study of articulation and formant frequencies in two female singers. In: *Proc. of Stockholm Music Acoustics Conference 1983 (SMAC 83)*, Vol. 46(1), ed. by A. Askenfelt, S. Felicetti, E. Jansson, J Sundberg (Kgl. Musikaliska Akad., Stockholm 1985) pp. 203–218

16.72 T. Baer, J.C. Gore, L.C. Gracco, P. Nye: Analysis of vocal tract shape and dimensions using magnetic

Amazon
EX1013 Page 716

resonance imaging: Vowels, J. Acoust. Soc. Am. **90**(2), 799–828 (1991)
16.73 D. Demolin, M. George, V. Lecuit, T. Metens, A. Soquet: Détermination par IRM de l'ouverture du velum des voyelles nasales du français. In: *Actes des XXièmes Journées d'Études sur la Parole* (1996)
16.74 A. Foldvik, K. Kristiansen, J. Kvaerness, A. Torp, H. Torp: *Three-dimensional ultrasound and magnetic resonance imaging: a new dimension in phonetic research* (Proc. Fut. Congress Phonetic Science Stockholm Univ., Stockholm 1995), Vol. 4, 46–49
16.75 B.H. Story, I.R. Titze, E.A. Hoffman: Vocal tract area functions from magnetic resonance imaging, J. Acoust. Soc. Am. **100**, 537–554 (1996)
16.76 O. Engwall: *Talking tongues, D.Sc. thesis* (Royal Institute of Technology (KTH), Stockholm 2002)
16.77 B. Lindblom, J. Sundberg: Acoustical consequences of lip, tongue, jaw and larynx movement, J. Acoust. Soc. Am. **50**, 1166–1179 (1971), also in Papers in Speech Communication: Speech Production, ed. by R.D. Kent, B.S. Atal, J.L. Miller (Acoust. Soc. Am., New York 1991) pp.329–342
16.78 J. Stark, B. Lindblom, J. Sundberg: APEX – an articulatory synthesis model for experimental and computational studies of speech production. In: *Fonetik 96: Papers presented at the Swedish Phonetics Conference TMH-QPSR 2/1996* (Royal Institute of Technology, Stockholm 1996) pp. 45–48
16.79 J. Stark, C. Ericsdotter, B. Lindblom, J. Sundberg: Using X-ray data to calibrate the APEX the synthesis. In: *Fonetik 98: Papers presented at the Swedish Phonetics Conference* (Stockholm Univ., Stockholm 1998)
16.80 J. Stark, C. Ericsdotter, P. Branderud, J. Sundberg, H.-J. Lundberg, J. Lander: The APEX model as a tool in the specification of speaker-specific articulatory behavior. In: *Proc. 14th Int. Congress of the Phonetic Sciences*, ed. by J.J. Ohala (1999)
16.81 C. Ericsdotter: Articulatory copy synthesis: Acoustic performane of an MRI and X-ray-based framework. In: *Proc. 15th Int. Congress of the Phonetic Sciences*, ed. by D. Recasens, M. Josep Solé, J. Romero (Universitat Autònoma de Barcelona, Barcelona 2003), CD-ROM
16.82 C. Ericsdotter: *Articulatory-acoustic relationships in Swedish vowel sounds, PhD thesis* (Stockholm University, Stockholm 2005)
16.83 K.N. Stevens, A.S. House: Development of a quantitative description of vowel articulation, J. Acoust. Soc. Am. **27**, 484–493 (1955)
16.84 S. Maeda: Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: *Speech Production and Speech Modeling*, ed. by W.J. Hardcastle, A. Marchal (Dordrecht, Kluwer 1990) pp. 131–150
16.85 P. Branderud, H. Lundberg, J. Lander, H. Djamshidpey, I. Wäneland, D. Krull, B. Lindblom: X-ray analyses of speech: methodological aspects. In: *Proc. XIIIth Swedish Phonetics Conf. (FONETIK 1998)* (KTH, Stockholm 1998)
16.86 C.Y. Espy-Wilson: Articulatory strategies, speech acoustics and variability. In: *From sound to Sense: 50+ Years of Discoveries in Speech Communication*, ed. by J. Slifka, S. Manuel, M. Mathies (MIT, Cambridge 2004)
16.87 J. Sundberg: Formant technique in a professional female singer, Acustica **32**, 89–96 (1975)
16.88 J. Sundberg, J. Skoog: Dependence of jaw opening on pitch and vowel in singers, J. Voice **11**, 301–306 (1997)
16.89 G. Fant: Glottal flow, models and interaction, J. Phon. **14**, 393–399 (1986)
16.90 E. Joliveau, J. Smith, J. Wolfe: Vocal tract resonances in singing: The soprano voice, J. Acoust. Soc. Am. **116**, 2434–2439 (2004)
16.91 R.K. Potter, A.G. Kopp, H.C. Green: *Visible Speech* (Van Norstrand, New York 1947)
16.92 M. Joosg: Acoustic phonetics, Language **24**, 447–460 (2003), supplement 2
16.93 C.F. Hockett: *A Manual of Phonology* (Indiana Univ. Publ., Bloomington 1955)
16.94 F.H. Guenther: Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production, Psychol. Rev. **102**, 594–621 (1995)
16.95 R.D. Kent, B.S. Atal, J.L. Miller: *Papers in Speech Communication: Speech Perception* (Acoust. Soc. Am., New York 1991)
16.96 S.D. Goldinger, D.B. Pisoni, P. Luce: Speech perception and spoken word recognition. In: *Principles of experimental phonetics*, ed. by N.J. Lass (Mosby, St Louis 1996) pp. 277–327
16.97 H.M. Sussman, D. Fruchter, J. Hilbert, J. Sirosh: Linear correlates in the speech signal: The orderly output constraint, Behav. Brain Sci. **21**, 241–299 (1998)
16.98 B. Lindblom: Economy of speech gestures. In: *The Production of Speech*, ed. by P.F. MacNeilage (Springer, New York 1983) pp. 217–245
16.99 P.A. Keating, B. Lindblom, J. Lubker, J. Kreiman: Variability in jaw height for segments in English and Swedish VCVs, J. Phonetics **22**, 407–422 (1994)
16.100 K. Rapp: A study of syllable timing. In: *STL/Quart. Prog. Status Rep. 1* (Royal Inst. of Technology, Stockholm(1971) pp. 14–19
16.101 F. Koopmans-van Beinum, J. Van der Stelt (Eds.): *Early stages in the development of speech movements* (Stockton, New York 1986)
16.102 K. Oller: Metaphonology and infant vocalizations. In: *Precursors of early speech*, ed. by B. Lindblom, R. Zetterström (Stockton, New York 1986) pp. 21–36

Amazon
EX1013 Page 717

16.103 L. Roug, L. Landberg, L. Lundberg: Phonetic development in early infancy, J. Child Language **16**, 19–40 (1989)

16.104 R. Stark: Stages of speech development in the first year of life. In: *Child Phonology: Volume 1: Production*, ed. by G. Yeni-Komshian, J. Kavanagh, C. Ferguson (Academic, New York 1980) pp. 73–90

16.105 R. Stark: Prespeech segmental feature development. In: *Language Acquisition*, ed. by P. Fletcher, M. Garman (Cambridge UP, New York 1986) pp. 149–173

16.106 D.K. Oller, R.E. Eilers: The role of audition in infant babbling, Child Devel. **59**(2), 441–449 (1988)

16.107 C. Stoel-Gammon, D. Otomo: Babbling development of hearing-impaired and normally hearing subjects, J. Speech Hear. Dis. **51**, 33–41 (1986)

16.108 R.E. Eilers, D.K. Oller: Infant vocalizations and the early diagnosis of severe hearing impairment, J. Pediatr. **124**(2), 99–203 (1994)

16.109 D. Ertmer, J. Mellon: Beginning to talk at 20 months: Early vocal development in a young cochlear implant recipient, J. Speech Lang. Hear. Res. **44**, 192–206 (2001)

16.110 R.D. Kent, M.J. Osberger, R. Netsell, C.G. Hustedde: Phonetic development in identical twins who differ in auditory function, J. Speech Hear. Dis. **52**, 64–75 (1991)

16.111 M. Lynch, D. Oller, M. Steffens: Development of speech-like vocalizations in a child with congenital absence of cochleas: The case of total deafness, Appl. Psychol. **10**, 315–333 (1989)

16.112 C. Stoel-Gammon: Prelinguistic vocalizations of hearing-impaired and normally hearing subjects: A comparison of consonantal inventories, J. Speech Hear. Dis. **53**, 302–315 (1988)

16.113 P.F. MacNeilage, B.L. Davis: Acquisition of speech production: The achievement of segmental independence. In: *Speech production and speech modeling*, ed. by W.J. Hardcastle, A. Marchal (Dordrecht, Kluwer 1990) pp. 55–68

16.114 T. Houtgast, H.J.M. Steeneken: A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria, J. Acoust. Soc. Am. **77**, 1069–1077 (1985)

16.115 T. Houtgast, H.J.M. Steeneken: *Past, Present and Future of the Speech Transmission Index*, ed. by S.J. van Wijngaarden (NTO Human Factors, Soesterberg 2002)

16.116 R. Drullman, J.M. Festen, R. Plomp: Effect of temporal envelope smearing on speech reception, J. Acoust. Soc. Am. **95**, 1053–1064 (1994)

16.117 R. Drullman, J.M. Festen, R. Plomp: Effect of reducing slow temporal modulations on speech reception, J. Acoust. Soc. Am. **95**, 2670–2680 (1994)

16.118 J. Morton, S. Marcus, C. Frankish: Perceptual centers (P-centers), Psych. Rev. **83**, 405–408 (1976)

16.119 S. Marcus: Acoustic determinants of perceptual center (P-center) location, Perception & Psychophysics **30**, 247–256 (1981)

16.120 G. Allen: The place of rhythm in a theory of language, UCLA Working Papers **10**, 60–84 (1968)

16.121 G. Allen: The location of rhythmic stress beats in English: An experimental study, UCLA Working Papers **14**, 80–132 (1970)

16.122 J. Eggermont: Location of the syllable beat in routine scansion recitations of a Dutch poem, IPO Annu. Prog. Rep. **4**, 60–69 (1969)

16.123 V.A. Kozhevnikov, L.A. Chistovich: Speech Articulation and Perception, JPRS **30**, 543 (1965)

16.124 C.E. Hoequist: The perceptual center and rhythm categories, Lang. Speech **26**, 367–376 (1983)

16.125 K.J. deJong: The correlation of P-center adjustments with articulatory and acoustic events, Perception Psychophys. **56**, 447–460 (1994)

16.126 A.D. Patel, A. Löfqvist, W. Naito: The acoustics and kinematics of regularly timed speech: a database and method for the study of the P-center problem. In: *Proc. 14th Int. Congress of the Phonetic Sciences*, ed. by J.J. Ohala (1999)

16.127 P. Howell: Prediction of P-centre location from the distribution of energy in the amplitude envelope: I & II, Perception Psychophys. **43**, 90–93, 99 (1988)

16.128 B. Pompino-Marschall: On the psychoacoustic nature of the Pcenter phenomenon, J. Phonetics **17**, 175–192 (1989)

16.129 C.A. Harsin: Perceptual-center modeling is affected by including acoustic rate-of-change modulations, Perception Psychophys. **59**, 243–251 (1997)

16.130 C.A. Fowler: Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet, J. Exp. Psychol. Gen. **112**, 386–412 (1983)

16.131 H. Fujisaki: Dynamic characteristics of voice fundamental frequency in speech and singing. In: *The Production of Speech*, ed. by P.F. MacNeilage (Springer, New York 1983) pp. 39–55

16.132 J. Frid: *Lexical and acoustic modelling of Swedish prosody, Dissertation* (Lund University, Lund 2003)

16.133 S. Öhman: Numerical model of coarticulation, J. Acoust. Soc. Am. **41**, 310–320 (1967)

16.134 J. t'Hart: F0 stylization in speech: Straight lines versus parabolas, J. Acoust. Soc. Am. **90**(6), 3368–3370 (1991)

16.135 D. Abercrombie: *Elements of General Phonetics* (Edinburgh Univ. Press, Edinburgh 1967)

16.136 K.L. Pike: *The intonation of America English* (Univ. of Michigan Press, Ann Arbor 1945)

16.137 G. Fant, A. Kruckenberg: Notes on stress and word accent in Swedish. In: *STL/Quart. Prog. Status Rep. 2-3* (Royal Inst. of Technology, Stockholm 1994) pp. 125–144

16.138 R. Dauer: Stress timing and syllable-timing reanalyzed, J. Phonetics **11**, 51–62 (1983)

Amazon
EX1013 Page 718

16.139 A. Eriksson: *Aspects of Swedish rhythm, PhD thesis, Gothenburg Monographs in Linguistics* (Gothenburg University, Gothenburg 1991)
16.140 O. Engstrand, D. Krull: Duration of syllable-sized units in casual and elaborated speech: cross-language observations on Swedish and Spanish, TMH-QPSR **44**, 69–72 (2002)
16.141 A.D. Patel, J.R. Daniele: An empirical comparison of rhythm in language and music, Cognition **87**, B35–B45 (2003)
16.142 D. Huron, J. Ollen: Agogic contrast in French and English themes: Further support for Patel and Daniele (2003), Music Perception **21**, 267–271 (2003)
16.143 D.H. Klatt: Synthesis by rule of segmental durations in English sentences. In: *Frontiers of speech communication research*, ed. by B. Lindblom, S. Öhman (Academic, London 1979) pp. 287–299
16.144 B. Lindblom: Final lengthening in speech and music. In: *Nordic Prosody*, ed. by E. Gårding, R. Bannert (Department of Linguistics Lund University, Lund 1978) pp. 85–102
16.145 A. Friberg, U Battel: Structural communication. In: *The Science and Psychology of Music Performance*, ed. by R. Parncutt, GE McPherson (Oxford Univ., Oxford 2001) pp. 199–218
16.146 J. Sundberg: Emotive transforms, Phonetica **57**, 95–112 (2000)
16.147 Brownlee: *The role of sentence stress in vowel reduction and formant undershoot: A study of lab speech and informal spontaneous speech, PhD thesis* (University of Texas, Austin 1996)
16.148 S.-J. Moon: *An acoustic and perceptual study of undershoot in clear and citation- form speech, PhD dissertation* (Univ. of Texas, Austin 1991)
16.149 K.N. Stevens, A.S. House: Perturbation of vowel articulations by consonantal context. An acoustical study, JSHR **6**, 111–128 (1963)
16.150 B. Lindblom: Spectrographic study of vowel reduction, J. Acoust. Soc. Am. **35**, 1773–1781 (1963)
16.151 P. Delattre: An acoustic and articulatory study of vowel reduction in four languages, IRAL-Int. Ref. Appl. VII/ **4**, 295–325 (1969)
16.152 D.P. Kuehn, K.L. Moll: A cineradiographic study of VC and CV articulatory velocities, J. Phonetics **4**, 303–320 (1976)
16.153 J.E. Flege: Effects of speaking rate on tongue position and velocity of movement in vowel production, J. Acoust. Soc. Am. **84**(3), 901–916 (1988)
16.154 R.J.J.H. van Son, L.C.W. Pols: "Formant movements of Dutch vowels in a text, read at normal and fast rate, J. Acoust. Soc. Am. **92**(1), 121–127 (1992)
16.155 D. van Bergem: *Acoustic and Lexical Vowel Reduction, Doctoral Dissertation* (University of Amsterdam, Amsterdam 1995)
16.156 W.L. Nelson, J.S. Perkell, J.R. Westbury: Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations, J. Acoust. Soc. Am. **75**(3), 945–951 (1984)
16.157 S.-J. Moon, B. Lindblom: Interaction between duration, context and speaking style in English stressed vowels, J. Acoust. Soc. Am. **96**(1), 40–55 (1994)
16.158 C.S. Sherrington: *Man on his nature* (MacMillan, London 1986)
16.159 R. Granit: *The Purposive Brain* (MIT, Cambridge 1979)
16.160 N. Bernstein: *The coordination and regulation of movements* (Pergamon, Oxford 1967)
16.161 P.F. MacNeilage: Motor control of serial ordering of speech, Psychol. Rev. **77**, 182–196 (1970)
16.162 A. Löfqvist: Theories and Models of Speech Production. In: *The Handbook of Phonetic Sciences*, ed. by W.J. Hardcastle, J. Laver (Blackwell, Oxford 1997) pp. 405–426
16.163 J.S. Perkell: Articulatory processes. In: *The Handbook of Phonetic Sciences. 5*, ed. by W.J. Hardcastle, J. Laver. (Blackwell, Oxford 1997) pp. 333–370
16.164 J. Sundberg, R. Leandersson, C. von Euler, E. Knutsson: Influence of body posture and lung volume on subglottal pressure control during singing, J. Voice **5**, 283–291 (1991)
16.165 T. Sears, J. Newsom Davis: The control of respiratory muscles during voluntary breathing. In: *Sound production in man*, ed. by A. Bouhuys et al. (Annals of the New York Academy of Science, New York 1968) pp. 183–190
16.166 B. Lindblom, J. Lubker, T. Gay: Formant frequencies of some fixed-mandible vowels and a model of motor programming by predictive simulation, J. Phonetics **7**, 147–161 (1979)
16.167 T. Gay, B. Lindblom, J. Lubker: Production of bite-block vowels: Acoustic equivalence by selective compensation, J. Acoust. Soc. Am. **69**(3), 802–810 (1981)
16.168 W.J. Hardcastle, J. Laver (Eds.): *The Handbook of Phonetic Sciences* (Blackwell, Oxford 1997)
16.169 J. S. Perkell, D. H. Klatt: *Invariance and variability in speech processes* (LEA, Hillsdale 1986)
16.170 A. Liberman, I. Mattingly: The motor theory of speech perception revised, Cognition **21**, 1–36 (1985)
16.171 C.A. Fowler: An event approach to the study of speech perception from a direct- realist perspective, J. Phon. **14**(1), 3–28 (1986)
16.172 E.L. Saltzman, K.G. Munhall: A dynamical approach to gestural patterning in speech production, Ecol. Psychol. **1**, 91–163 (1989)
16.173 M. Studdert-Kennedy: How did language go discrete?. In: *Evolutionary Prerequisites of Language*, ed. by M. Tallerman (Oxford Univ., Oxford 2005) pp. 47–68
16.174 R. Jakobson, G. Fant, M. Halle: *Preliminaries to Speech Analysis, Acoustics Laboratory, MIT Tech. Rep. No. 13* (MIT, Cambridge 1952)
16.175 B. Lindblom: Explaining phonetic variation: A sketch of the H&H theory. In: *Speech Produc-*

Amazon
EX1013 Page 719

tion and Speech Modeling, ed. by W.J. Hardcastle, A. Marchal (Dordrecht, Kluwer 1990) pp. 403–439

16.176 B. Lindblom: Role of articulation in speech perception: Clues from production, J. Acoust. Soc. Am. **99**(3), 1683–1692 (1996)

16.177 E. Rapoport: Emotional expression code in opera and lied singing, J. New Music Res. **25**, 109–149 (1996)

16.178 J. Sundberg, E. Prame, J. Iwarsson: Replicability and accuracy of pitch patterns in professional singers. In: *Vocal Fold Physiology, Controlling Complexity and Chaos*, ed. by P. Davis, N. Fletcher (Singular, San Diego 1996) pp. 291–306, Chap. 20

16.179 J.J. Ohala: An ethological perspective on common cross-language utilization of F0 of voice, Phonetica **41**, 1–16 (1984)

16.180 I. Fónagy: Hörbare Mimik, Phonetica **1**, 25–35 (1967)

16.181 K. Scherer: Expression of emotion in voice and music, J. Voice **9**, 235–248 (1995)

16.182 P. Juslin, P. Laukka: Communication of emotions in vocal expression and music performance: Different channels, same code?, Psychol. Rev. **129**, 770–814 (2003)

16.183 J. Sundberg, J. Iwarsson, H. Hagegård: A singers expression of emotions in sung performance,. In: *Vocal Fold Physiology: Voice Quality Control*, ed. by O. Fujimura, M. Hirano (Singular, San Diego 1995) pp. 217–229

Amazon
EX1013 Page 720

17. Computer Music

This chapter covers algorithms, technologies, computer languages, and systems for computer music. Computer music involves the application of computers and other digital/electronic technologies to music composition, performance, theory, history, and perception. The field combines digital signal processing, computational algorithms, computer languages, hardware and software systems, acoustics, psychoacoustics (low-level perception of sounds from the raw acoustic signal), and music cognition (higher-level perception of musical style, form, emotion, etc.). Although most people would think that analog synthesizers and electronic music substantially predate the use of computers in music, many experiments and complete computer music systems were being constructed and used as early as the 1950s.

Because of this rich legacy, and the large number of researchers working on digital audio (primarily in speech research laboratories), there are a large number of algorithms for synthesizing sound using computers. Thus, a significant emphasis in this chapter will be placed on digital sound synthesis and processing, first providing an overview of the representation of audio in digital systems, then covering most of the popular algorithms for digital analysis and synthesis of sound.

17.1 Computer Audio Basics 714
17.2 Pulse Code Modulation Synthesis 717
17.3 Additive (Fourier, Sinusoidal) Synthesis . 719
17.4 Modal (Damped Sinusoidal) Synthesis 722
17.5 Subtractive (Source-Filter) Synthesis..... 724
17.6 Frequency Modulation (FM) Synthesis 727
17.7 FOFs, Wavelets, and Grains..... 728
17.8 Physical Modeling (The Wave Equation) . 730
17.9 Music Description and Control..... 735
17.10 Composition..... 737
17.11 Controllers and Performance Systems 737
17.12 Music Understanding and Modeling by Computer 738
17.13 Conclusions, and the Future 740
References 740

Pulse code modulation (PCM) is the means for sampling and retrieval of audio in computers, and PCM synthesis uses combinations of prerecorded waveforms to reconstruct speech, sound effects, and music instrument sounds, based largely on the work of Joseph Fourier, who gave us a technique for formulating any waveform from sine waves. Additive synthesis uses fundamental waveforms (often sine waves), to construct more complicated waves. PCM and sinusoidal (Fourier) additive synthesis are often called *nonparametric* techniques. The word parametric in this context means that an algorithm with a few (but hopefully expressive) parameters can be used to generate a large variety of waveforms and spectral properties by varying the parameters.

Since PCM involves recording and playing back waveforms, no actual expressive parameters are available in PCM systems. Similarly, Fourier's theorem says that we can represent any waveform with a sum of sine waves, but in fact it might take a huge number of sine waves to represent a particular waveform accurately. So, without further work to *parameterize* PCM samples or large groups of additive sine waves, these techniques are nonparametric.

Modal synthesis recognizes the fact that many vibrating systems exhibit a relatively few strong sinusoidal modes (natural frequencies) of vibration. Rigid bars and plates, plucked strings, and other structures are good candidates for modal synthesis, where after being excited (struck, plucked, etc.) the vibration/sound generation is restricted to a limited set of exponentially decaying sine waves. Modal synthesis is considered by some to be parametric, since the number of sine waves needed is often quite limited. Others claim that modal

Amazon
EX1013 Page 721

synthesis is nonparametric, since there is no small set of meaningful and expressive parameters that control variation in the sound.

Subtractive synthesis begins with spectrally complex waveforms (such as pulses or noise), and uses filters to shape the spectrum and waveform to the desired result. Subtractive synthesis works particularly well for certain classes of sounds, such as human speech, where a spectrally rich source (pulses of the vocal folds) is filtered by a time-varying filter (the acoustic tube of the vocal tract). Subtractive synthesis is parametric, because only a few descriptive numbers control such a model. For example, a subtractive speech model might be controlled by numbers expressing how pitched versus noisy the voice is, the voice pitch, the loudness, and 8–10 locations of the resonances of the vocal tract. Using only a dozen or so variable numbers (parameters), a rich variety of vowel and consonant sounds can be synthesized.

Frequency modulation (FM) synthesis exploits properties of nonlinearity to generate complex spectra and waveforms from much simpler ones. Modulating the frequency or phase of a sine wave (the carrier) by another sine wave (the modulator) causes multiple *sideband* sinusoids to appear at frequencies related to the ratio of the carrier and modulator frequencies. Thus an important parameter in FM is the carrier:modulator (C:M) ratio, and another is the amount (index) of modulation, which when increased causes the number of sidebands to increase.

Wavelets are waveforms that are usually compact in time and frequency. They are viewed by some as related to the fundamental unit of sound, by others as convenient building blocks for constructing sonic textures, and by others as related to important small events in nature and acoustics. *Formes d'onde formantiques* (FOFs, French for formant wave functions) are special wavelets that model the individual resonances of the vocal tract, and thus have been successfully used for modeling speech and singing.

Physical models solve the wave equation in space and time to generate waveforms in the same ways that physical acoustics does. Efficient techniques combining discrete simulation of physics with modern digital filtering techniques allow models of plucked strings, wind instruments, percussion, the vocal tract, and many other musical systems to be solved efficiently by computers.

After the presentation and examples of the main synthesis algorithms, some industry standards for representing and manipulating musical data in computerized systems will be discussed. The musical instrument digital interface (MIDI), downloadable sounds (DLS), open sound control (OSC), structured audio orchestra language (SAOL), and other score and control standards will be briefly covered.

Some languages and systems for computer music composition are discussed, including the historical MusicX languages, and many modern languages such as Cmusic and MAX/MSP, followed by a few real-time performance and interaction systems. Finally, some aspects of computer modeling of human listening and systems for machine-assisted and automatic composition are covered.

17.1 Computer Audio Basics

Typically, digital audio signals are formed by *sampling* analog (continuous in time and amplitude) signals at regular intervals in time, and then *quantizing* the amplitudes to discrete values. The process of sampling a waveform, holding the value, and quantizing the value to the nearest number that can be digitally represented (as a specific integer on a finite range of integers) is called analog-to-digital (A to D, or A/D) conversion [17.1]. A device that does A/D conversion is called an analog-to-digital converter (ADC). Coding and representing waveforms in sampled digital form is called pulse code modulation (PCM), and digital audio signals are often called PCM audio. The process of converting a sampled signal back into an analog signal is called digital-to-analog conversion (D to A, or D/A), and the device is called a digital-to-analog converter (DAC). Low-pass filtering (smoothing the samples to remove unwanted high frequencies) is necessary to reconstruct the sampled signal back into a smooth continuous-time analog signal. This filtering is usually contained in the DAC hardware.

The time between successive samples is usually denoted by T. Sampling an analog signal first requires filtering it to remove unwanted high frequencies (more on this shortly), holding the value steady for a period while a stable measurement can be made, then associating the analog value with a digital number (coding). Analog signals can have any of the infinity of real-numbered amplitude values. Since computers work with fixed word sizes (8 bit bytes, 16 bit words, etc.), digital signals can only have a finite number of amplitude val-

Amazon
EX1013 Page 722

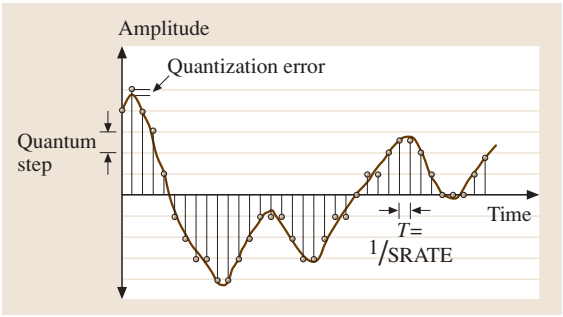


Fig. 17.1 Linear sampling and quantization

ues. In converting from analog to digital, rounding takes place and a given analog value must be quantized to the nearest digital value. The difference between quantization steps is called the quantum (not as in quantum physics or leaps, but the Latin word for a fixed-sized jump in value or magnitude). Sampling and quantization is shown in Fig. 17.1; note the errors introduced in some sample values due to the quantization process.

A fundamental law of digital signal processing states that, if an analog signal is bandlimited with bandwidth B Hz, (Hz = Hertz = samples per second), the signal can be periodically sampled at a rate of $2B$ Hz or greater, and exactly reconstructed from the samples. Bandlimited with bandwidth B means that no frequencies above B exist in the signal. The rate $2B$ is called the sampling rate, and B is called the Nyquist frequency. Intuitively, a sine wave at the highest frequency B present in a bandlimited signal can be represented using two samples per period (one sample at each of the positive and negative peaks), corresponding to a sampling frequency of $2B$. All signal components at lower frequencies can be uniquely represented and distinguished from each other using this same sampling frequency of $2B$. If there are components present in a signal at frequencies greater than $1/2$ the sampling rate, these components will not be represented properly, and will *alias* (i. e.,, show up as frequencies different from their original values).

To avoid aliasing, ADC hardware often includes filters that limit the bandwidth of the incoming signal before sampling takes place, automatically changing as a function of the selected sampling rate. Figure 17.2 shows aliasing in complex and simple (sinusoidal) waveforms. Note the loss of detail in the complex waveform. Also note that samples at less than two times the frequency of a sine wave could also have arisen from a sine wave of much lower frequency. This is the fundamental nature of aliasing, because frequencies higher than $1/2$ sample rate *alias* as lower frequencies.

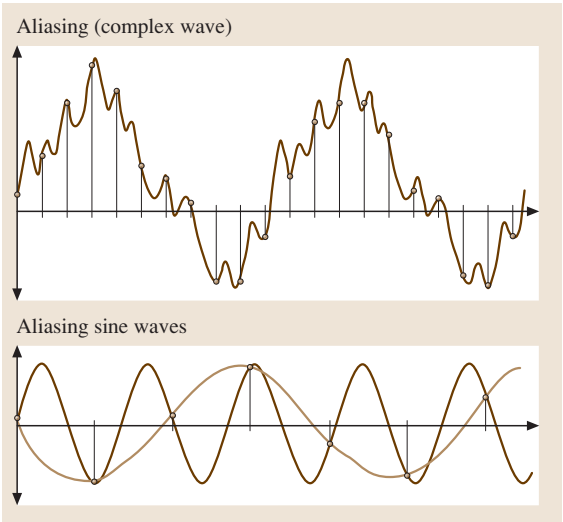


Fig. 17.2 Because of inadequate sampling rate, aliasing causes important features to be lost

Humans can perceive frequencies from roughly 20 Hz to 20 kHz, thus requiring a minimum sampling rate of at least 40 kHz. Speech signals are often sampled at 8 kHz (*telephone quality*) or 11.025 kHz, while music is usually sampled at 22.05 kHz, 44.1 kHz (the sampling rate used on audio compact discs), or 48 kHz. Some new formats allow for sampling rates of 96 kHz, and even 192 kHz. This is because some engineers believe that we can actually hear things, or the effects of things, higher than 20 kHz. Dolphins, dogs, and some other animals can actually hear higher than us, so these new formats could be considered as catering to those potential markets.

In a digital system, a fixed number of binary digits (bits) are used to sample the analog waveform, by quantizing it to the closest number that can be represented. This quantization is accomplished either by rounding to the quantum value nearest the actual analog value, or by truncation to the nearest quantum value less than or equal to the actual analog value. With uniform sampling in time, a properly bandlimited signal can be exactly recovered provided that the sampling rate is twice the bandwidth or greater, but only if there is no quantization. When the signal values are rounded or truncated, the amplitude difference between the original signal and the quantized signal is lost forever. This can be viewed as an additive noise component upon reconstruction. Using the additive-noise assumption gives an approximate best-case signal-to-quantization-noise ratio (SNR) of approximately $6N$ dB, where N is the number of

Amazon
EX1013 Page 723

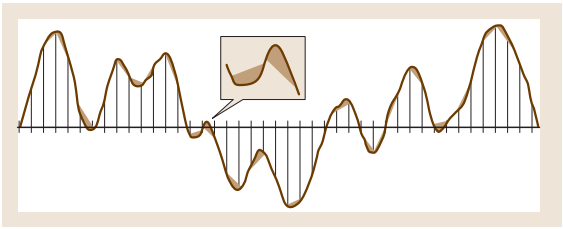


Fig. 17.3 Resampling by linear interpolation. Grey regions show the errors due to linear interpolation

bits. Using this approximation implies that a 16 bit linear quantization system will exhibit an SNR of approximately 96 dB. 8 bit quantization exhibits a signal-to-quantization-noise ratio of approximately 48 dB. Each extra bit improves the signal-to-noise ratio by about 6 dB.

Most computer audio systems use two or three types of audio data words. 16 bit (per channel) data is quite common, and this is the data format used in compact disc systems. Many recent (high-definition) formats allow for 24 bit samples. 8 bit data is common for speech data in personal computer (PC) and telephone systems, usually using methods of quantization that are nonlinear. In such systems the quantum is smaller for small amplitudes, and larger for large amplitudes.

Changing the playback sample rate on sampled sound results in a pitch shift. Many systems for recording, playback, processing, and synthesis of music, speech, or other sounds allow or require flexible control of pitch (sample rate). The most accurate pitch control is necessary for music synthesis. In sampling synthesis, this is accomplished by dynamic sample-rate conversion (interpolation). In order to convert a sampled data signal from one sampling rate to another, three steps are required: bandlimiting, interpolation, and resampling.

Bandlimiting: First, it must be assured that the original signal was properly bandlimited to less than half the new sampling rate. This is always true when *upsampling* (i. e., converting from a lower sampling rate to a higher one). When *downsampling*, however, the signal must be first low-pass filtered to exclude frequencies greater than half the new target rate.

Interpolation: When resampling by any but simple integer factors such as 2, 4, 1/2, etc., the task requires the computation of the values of the original bandlimited analog signal at the correct points between the existing sampled points. This is called interpolation. The most common form of interpolation used is linear interpolation, where the fractional time samples needed are computed as a linear combination of the two surrounding samples. Many assume that linear interpolation is the correct means, or at least an adequate means for accomplishing audio sample rate conversion. Figure 17.3 shows resampling by linear interpolation; note the errors shown by the shaded regions.

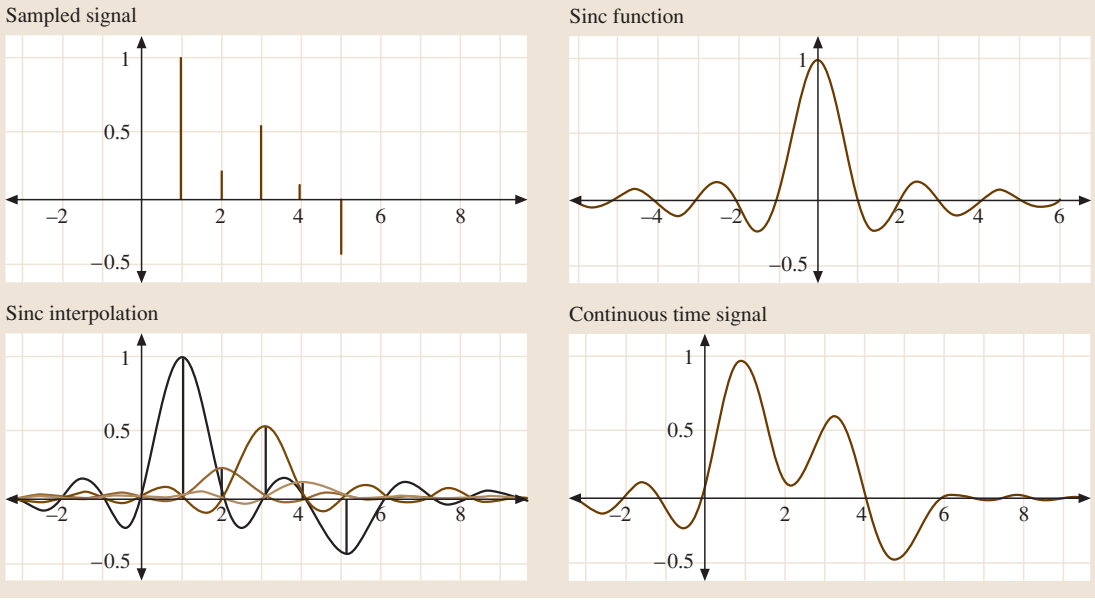


Fig. 17.4 Sinc interpolation for reconstruction and resampling

Amazon
EX1013 Page 724

Some might notice that linear interpolation is not quite adequate, so they might assume that the correct solution must lie in more elaborate curve-fitting techniques using quadratics, cubics, or higher-order splines, and indeed these types of interpolation can be adequate for some applications. To arrive at the correct answer (provably correct from theory) the interpolation task should be viewed and accomplished as a filtering problem, with the filter designed to meet some appropriate error criterion. Linear time-invariant filtering is accomplished by convolution with a filter function. If the resampling filter is defined appropriately, we can exactly reconstruct the original analog waveform from the samples.

The correct (ideal in a provable digital signal processing sense) way to perform interpolation is convolution with the sinc function, defined as:

$$\mathrm{sinc}(t/T) = \sin(\pi t/T)/(\pi t/T)\,,$$

where

$$T = 1/\mathrm{SRATE}\,.$$

The sinc function is the ideal low-pass filter with a cutoff of SRATE/2, where SRATE is the sampling rate. Figure 17.4 shows reconstruction of a continuous waveform by convolution of a digital signal with the sinc function. Each sample is multiplied by a corresponding continuous sinc function, and those are added up to arrive at the continuous reconstructed signal [17.2].

Resampling: This is usually accomplished at the same time as interpolation, because it is not necessary to reconstruct the entire continuous waveform in order to acquire new discrete samples. The resampling ratio can be time varying, making the problem a little more difficult. However, viewing the problem as a filter-design and implementation issue allows for guaranteed tradeoffs of quality and computational complexity.

17.2 Pulse Code Modulation Synthesis

The majority of digital sound and music synthesis today is accomplished via the playback of stored pulse code modulation (PCM) waveforms. Single-shot playback of entire segments of stored sounds is common for sound effects, narrations, prompts, segments of music, etc. Most high-quality modern electronic music synthesizers, speech synthesis systems, and PC software systems for sound synthesis use pre-stored PCM as the basic data. This data is sometimes manipulated to yield the final output sound(s).

There are a number of different ways to look at sound for computer music, with PCM being only one. We can look at the physics that produce the sound and try to model those. We could also look at the spectrum of the sound and other characteristics having to do with the perception of those sounds. Indeed, much of the legacy of computer music has revolved around parametric (using mathematical algorithms, controlled by a few well-chosen/-designed control parameters) analysis and synthesis algorithms. We will discuss most of the commonly used algorithms later, but first we should look at PCM in more depth.

For speech, the most common synthesis technique is *concatenative* synthesis [17.3]. Concatenative phoneme synthesis relies on the concatenation of roughly 40 pre-stored phonemes (for English). Examples of vowel phonemes are /i/ as in beet, /I/ as in bit, /a/ as in father, etc. Examples of nasals are /m/ as in mom, /n/ as in none, /ng/ as in sing, etc. Examples of fricative consonant phonemes are /s/ as in sit, /sh/ as in ship, /f/ as in fifty, etc. Examples of voiced fricative consonants are /z/, /v/ (visualize), etc. Examples of plosive consonants are /t/ as in tat, /p/ as in pop, /k/ as in kick, etc. Examples of voiced plosives include /d/, /b/, /g/ (dude, bob, gag) etc. Vowels and nasals are essentially periodic pitched sounds, so the minimal required stored waveform is only one single period of each. Consonants require more storage because of their noisy (non-pitched, aperiodic) nature.

The quality of concatenative phoneme synthesis is generally considered quite low, due to the simplistic assumption that all of the pitched sounds (vowels, etc.) are purely periodic. Also, simply *gluing* /s/ /I/ and /ng/ together does not make for a high-quality realistic synthesis of the word "sing". In actual speech, phonemes gradually blend into each other as the jaw, tongue, and other articulators move with time.

Accurately capturing the transitions between phonemes with PCM requires recording transitions from phoneme to phoneme, called *diphones*. A concatenative diphone synthesizer blends together stored diphones. Examples of diphones include see, she, thee, and most of the roughly 40x40 possible combinations of phonemes. Much more storage is necessary for a diphone synthesizer, but the resulting increase in quality is significant. PCM speech synthesis can be improved further by stor-

Amazon
EX1013 Page 725

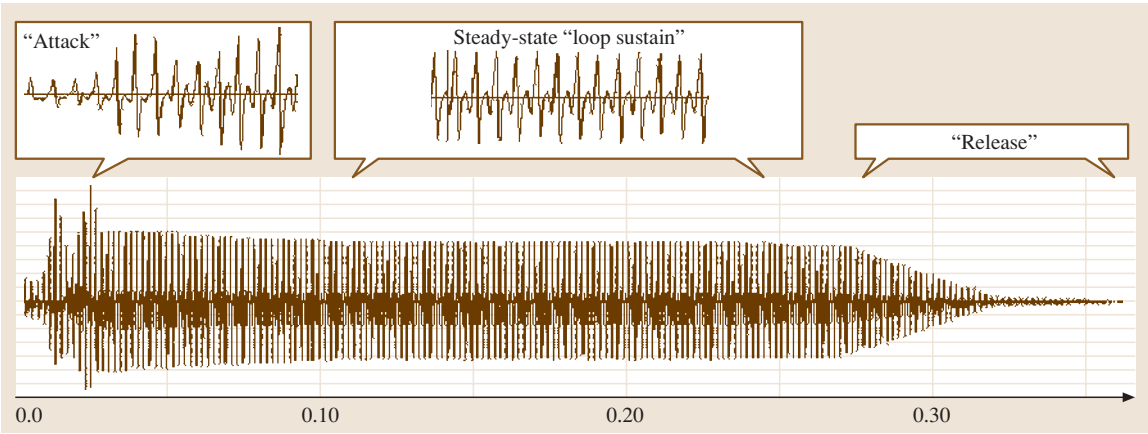


Fig. 17.5 Wave-table synthesis of a trumpet tone

ing multiple samples, for different pitch ranges, genders, voice qualities, etc.

For musical sounds, it is common to store only a loop, or table, of the periodic component of a recorded sound waveform and play that loop back repeatedly. This is called "wave-table synthesis" [17.4]. For more realism, the attack or beginning portion of the recorded sound can be stored in addition to the periodic steady-state part. Figure 17.5 shows the synthesis of a trumpet tone starting with an attack segment, followed by repetition of a periodic loop, ending with an enveloped decay (or release). The *envelope* is a synthesizer/computer music term for a time-varying change applied to a waveform amplitude, or other parameter. Envelopes are often described by four components: the *attack time*, the *decay time* (decay here means the initial decay down to the steady state segment), the *sustain level*, and the *release time* (final decay). Hence, envelopes are sometimes called ADSRs.

Originally called *sampling synthesis* in the music industry, any synthesis using stored PCM waveforms has now become commonly known as "wave-table synthesis". Filters are usually added to high-quality wave-table synthesis, allowing control of spectral brightness as a function of intensity, and to get more variety of sounds from a given set of samples.

Pitch shifting of a PCM sample or wave table is accomplished via interpolation as discussed in the previous section. A given sample can be pitch-shifted only so far in either direction before it begins to sound unnatural. This can be dealt with by storing multiple recordings of the sound at different pitches, and switching or interpolating between these upon resynthesis. This is called *multi-sampling*. Multi-sampling might also include the storage of separate samples for loud and soft sounds. Linear or other interpolation is used to blend the loudness multi-samples as a function of the desired synthesized volume. This adds realism, because loudness is not simply a matter of amplitude or power, and most sound sources exhibit spectral variations as a function of loudness. For example, there is usually more high-frequency energy (*brightness*) in loud sounds than in soft sounds. Filters can also be used to add spectral variation.

A common tool used to describe the various components and steps of signal processing in performing digital music synthesis is the *synthesizer patch* (historically named from hooking various electrical components together using patch cords). In a patch, a set of fairly commonly agreed building blocks, called *unit generators* (also called modules, plug-ins, operators, op-codes, and other terms) are hooked together in a signal flow diagram. This historical [17.5] graphical method of describing signal generation and processing affords a visual representation that is easily printed in pa-

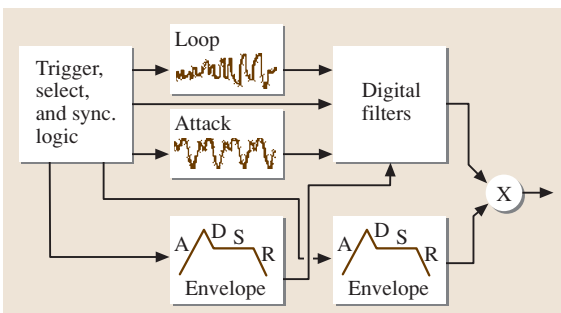


Fig. 17.6 A synthesis *patch* showing interconnection of *unit generators*

Amazon
EX1013 Page 726

pers, textbooks, patents, etc. Further, graphical patching systems and languages have been important to the development and popularization of certain algorithms, and computer music in general. Figure 17.6 shows a PCM synthesizer patch with attack and loop wave tables whose amplitudes are controlled by an envelope generator, and a time-varying filter (also controlled by another envelope generator).

17.3 Additive (Fourier, Sinusoidal) Synthesis

Lots of sound-producing objects and systems exhibit sinusoidal modes. A plucked string might exhibit many modes, with the strength of each mode determined by the boundary conditions of the terminations, and the shape of the excitation pluck. Striking a metal plate with a hammer excites many of the vibrational modes of the plate, determined by the shape of the plate, and by where it is struck. A singing voice, struck drum head, bowed violin string, struck bell, or blown trumpet exhibit oscillations characterized by a sum of sinusoids. The recognition of the fundamental nature of the sinusoid gives rise to a powerful model of sound synthesis based on summing sinusoidal modes.

These modes have a very special relationship in the case of the plucked string, a singing voice, and some other limited systems, in that their frequencies are all integer multiples (at least approximately) of one basic sinusoid, called the *fundamental*. This special series of sinusoids is called a *harmonic series*, and lies at the basis of the Fourier-series representation of shapes, waveforms, oscillations, etc. The Fourier series [17.6] solves many types of problems, including physical problems with boundary constraints, but is also applicable to any shape or function. Any periodic waveform (repeating over and over again) F_{per} can be transformed into a Fourier series, written as:

$$F_{\text{per}}(t) = a_0 + \Sigma_m [b_m \cos(2\pi f_0 m t) + c_m \sin(2\pi f_0 m t)] \,. \tag{17.1}$$

The limits of the summation are technically infinite, but we know that we can cut off our frequencies at the Nyquist frequency for digital signals. The a_0 term is a constant offset, or the average of the waveform. The b_m and c_m coefficients are the weights of the mth harmonic cosine and sine terms. If the function $F_{\text{per}}(t)$ is purely even about $t = 0$ $[(F(-t) = F(t))]$, only cosines are required to represent it, and only the b_m terms would be nonzero. Similarly, if the function $F_{\text{per}}(t)$ is odd $[(F(-t) = -F(t))]$, only the c_m terms would be required. An arbitrary function $F_{\text{per}}(t)$ will require sinusoidal harmonics of arbitrary (but specific) amplitudes and phases. The magnitude and phase of the mth harmonic in the Fourier series can be found by:

$$A_m = \sqrt{b_m^2 + c_m^2} \,, \tag{17.2a}$$

$$\theta_m = \arctan\,(c_m / b_m) \,. \tag{17.2b}$$

Phase is defined relative to the cosine, so if c_m is zero, θ_m is zero. As a brief example, Fig. 17.7 shows the first few sinusoidal harmonics required to build up an approximation of a square wave. Note that due to symmetries (boundary conditions), only odd sine harmonics are required. Using more sines improves the approximation.

The process of solving for the sine and cosine components of a signal or waveform is called Fourier analysis, or the Fourier transform. If the frequency variable is sampled (as is the case in the Fourier series, represented by m), and the time variable t is sampled as well (as it is in PCM waveform data, represented by n), then the Fourier transform is called the *discrete Fourier transform* (DFT). The DFT is given by

$$F(m) = \sum_{n=0}^{N-1} f(n)[\cos(2\pi mn/N) - \mathrm{i}\sin(2\pi mn/N)] \,. \tag{17.3}$$

Where N is the length of the signal being analyzed. The inverse DFT (IDFT) is very similar to the Fourier

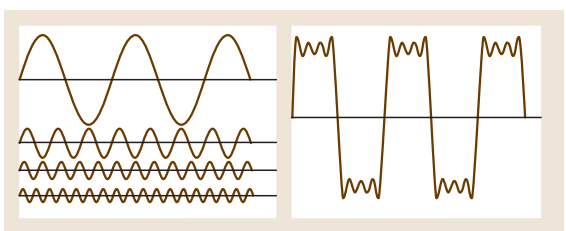


Fig. 17.7 A sum of odd harmonics approximates a square wave

Amazon
EX1013 Page 727

series

$$f(n) = 1/N \sum_{m=0}^{N-1} F(m)[\cos(2\pi mn/N) + \mathrm{i}\sin(2\pi mn/N)] \,. \quad (17.4)$$

The imaginary number $\mathrm{i}(=\sqrt{-1})$ is used to place the cosine and sine components in a unique mathematical arrangement, where odd $[(x(-n) = -x(n))]$ terms of the waveform are represented as imaginary components, and even $[(x(-n) = x(n))]$ terms are represented as real components. This gives us a means to talk about the magnitude and phase in terms of the magnitude and phase of $F(m)$ (a complex number).

There is a near-mystical expression of equality in mathematics known as Euler's identity, which links trigonometry, exponential functions, and complex numbers in a single equation

$$\mathrm{e}^{\mathrm{i}\theta} = \cos(\theta) + \mathrm{i}\sin(\theta) \,.$$

We can use Euler's identity to write the DFT and IDFT in shorthand

$$F(m) = \sum_{n=0}^{N-1} f(n) e^{-\mathrm{i}2\pi mn/N} \,, \quad (17.5)$$

$$f(n) = 1/N \sum_{m=0}^{N-1} F(m)\,\mathrm{e}^{\mathrm{i}2\pi mn/N} \,. \quad (17.6)$$

Converting the cosine/sine form to the complex exponential form allows lots of manipulations that would be difficult otherwise. But we can also write the DFT in real-number terms as a form of the Fourier series

$$f(n) = 1/N \sum_{m=0}^{N-1} F_b(n)\cos(2\pi mn/N) + F_c(n)\sin(2\pi mn/N) \,, \quad (17.7)$$

where

$$F_b(m) = \sum_{n=0}^{N-1} f(n)\cos(2\pi mn/N) \,, \quad (17.8)$$

$$F_c(m) = \sum_{n=0}^{N-1} -f(n)\sin(2\pi mn/N) \,. \quad (17.9)$$

The fast Fourier transform (FFT) is a cheap way of calculating the DFT. There are thousands of references on the FFT [17.6], and scores of implementations of it, so for our purposes we will just say that it is much more efficient than trying to compute the DFT directly from the definition. A well-crafted FFT algorithm for real input data takes on the order of $N\log_2(N)$ multiply–adds to compute. Comparing this to the N^2 multiplies of the DFT, N does not have to be very big before the FFT is a winner. There are some down sides, such as the fact that FFTs can only be computed for signals whose lengths are exactly powers of 2, but the advantages of using it often outweigh the annoying power-of-two problems. Practically speaking, users of FFTs usually carve up signals into small chunks (powers of two), or *zero pad* a signal out to the next biggest power of two.

The short-time Fourier transform (STFT) breaks up the signal and applies the Fourier transform to each segment individually [17.7]. By selecting the window size (length of the segments), and hop size (how far the window is advanced along the signal) to be perceptually relevant, the STFT can be used as an approximation of human audio perception. Figure 17.8 shows the waveform of the utterance of the word "synthesize", and some STFT spectra corresponding to windows at particular points in time.

If we inspect the various spectra in Fig. 17.8, we can note that the vowels exhibit harmonic spectra (clear, evenly spaced peaks corresponding to the harmonics of the pitched voice), while the consonants exhibit noisy spectra (no clear sinusoidal peaks). Recognizing that some sounds are well approximated/modeled by additive sine waves [17.8], while other sounds are essentially noisy, spectral modeling [17.7] breaks the sound into deterministic (sines) and stochastic (noise) components. Figure 17.9 shows a general sines+noise additive synthesis model, allowing us to control the amplitudes and frequencies of a number of sinusoidal oscillators, and model the noisy component with a noise source and a spectral shaping filter.

The beauty of this type of model is that it recognizes the dominant sinusoidal nature of many sounds, while still recognizing the noisy components that might be also present. More efficient and parametric representations, and many interesting modifications can be made to the signal on resynthesis. For example, removing the harmonics from voiced speech, followed by resynthesizing with a scaled version of the noise residual, can result in the synthesis of whispered speech.

One further improvement to spectral modeling is the recognition [17.9] that there are often brief (impulsive) moments in sounds that are really too short in time to be adequately analyzed by spectrum analysis. Further, such moments in the signal usually corrupt the sinusoidal/noise analysis process. Such

Amazon
EX1013 Page 728

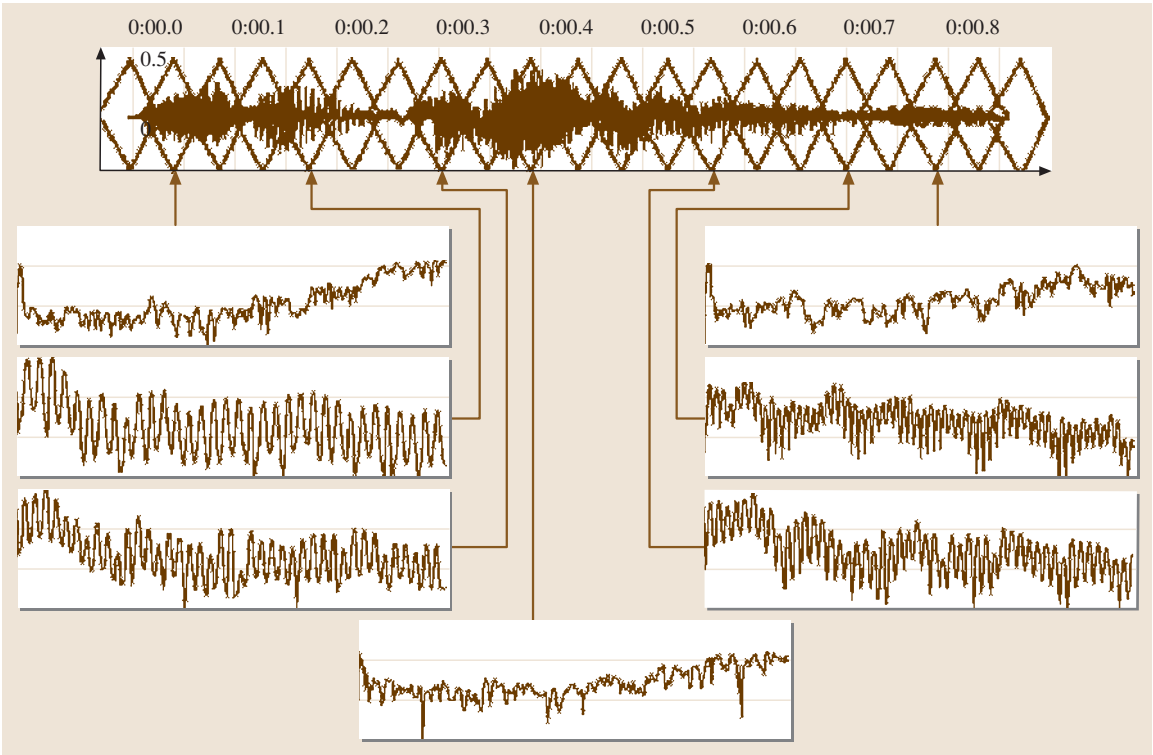


Fig. 17.8 Some STFT frames of the word 'synthesize'

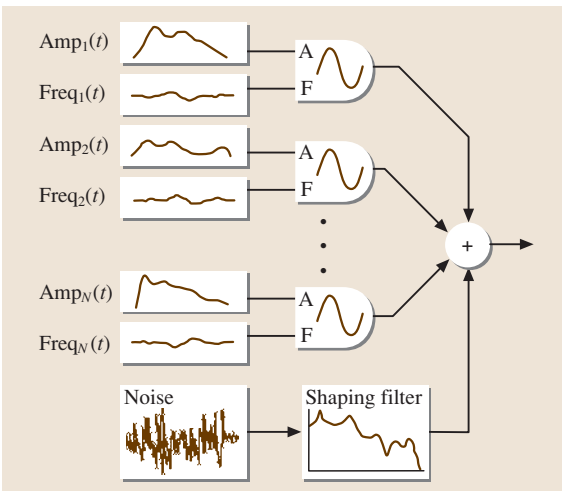


Fig. 17.9 Sinusoidal additive model with filtered noise added for spectral modeling synthesis

events, called transients, can be modeled in other ways (often by simply keeping the stored PCM for that segment).

Using the short-time Fourier transform, the *phase vocoder* (VoiceCoder) [17.10, 11] processes sound by calculating and maintaining both magnitude and phase. The frequency bins (basis sinusoids) of the DFT can be viewed as narrow-band filters, so the Fourier transform of an input signal can be viewed as passing it through a bank of narrow band-pass filters. This means that on the order of hundreds to thousands of sub-bands are used.

The phase vocoder has found extensive use in computer music composition. Many interesting practical and artistic transformations can be accomplished using the phase vocoder, including nearly artifact-free and independent time and pitch shifting. A technique called *cross synthesis* assumes that one signal is the analysis signal. The time-varying magnitude spectrum of the analysis signal (usually smoothed in the frequency domain) is multiplied by the spectral frames of another input (or filtered) signal (sometimes brightened first by high-frequency emphasis pre-filtering), yielding a composite signal that has the attributes of both. Cross-synthesis has produced the sounds of talking cows, *morphs* between people and cats, trumpet/flute hybrids, etc.

Amazon
EX1013 Page 729

17.4 Modal (Damped Sinusoidal) Synthesis

The simplest physical system that does something acoustically (and musically) interesting is the mass/spring/damper (Fig. 17.10 [17.12]. The differential equation describing that system has a solution that is a single exponentially decaying cosine wave. The Helmholtz resonator (large, contained air cavity with a small long-necked opening, like a pop bottle, Fig. 17.11 behaves like a mass/spring/damper system, with the same exponentially damped cosine behavior. The equations describing the behavior of these systems is:

$$d^2y/dt^2 + (r/m)\,dy/dt + (k/m)y = 0\,, \qquad (17.10)$$

$$y(t) = y_0\,e^{(-rt/2m)}\cos\left(t\sqrt{\left[k/m - (r/2m)^2\right]}\right). \qquad (17.11)$$

Of course, most systems that produce sound are more complex than the ideal mass/spring/damper system, or a pop bottle. And of course most sounds are more complex than a simple damped exponential sinusoid. Mathematical expressions of the physical forces (thus the accelerations) can be written for nearly any system, but solving such equations is often difficult or impossible. Some systems have simple enough properties and geometries to allow an exact solution to be written out for their vibrational behavior. An ideal string under tension is one such system.

Here we will resort to some graphical arguments and our prior discussion of the Fourier transform to motivate further the notion of sinusoids in real physical systems. The top of Fig. 17.12 shows a string, lifted from a point in the center (halfway along its length). Below that is shown a set of sinusoidal mode that the center-plucked

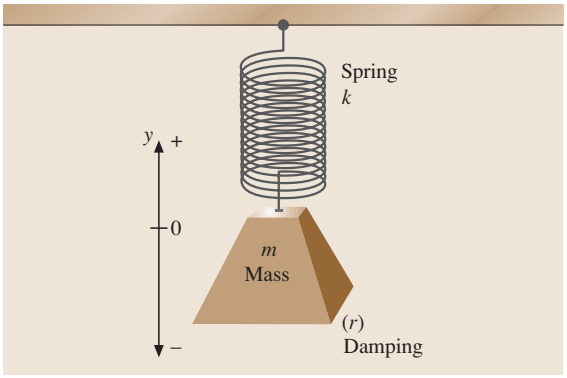


Fig. 17.10 Mass/spring/damper system

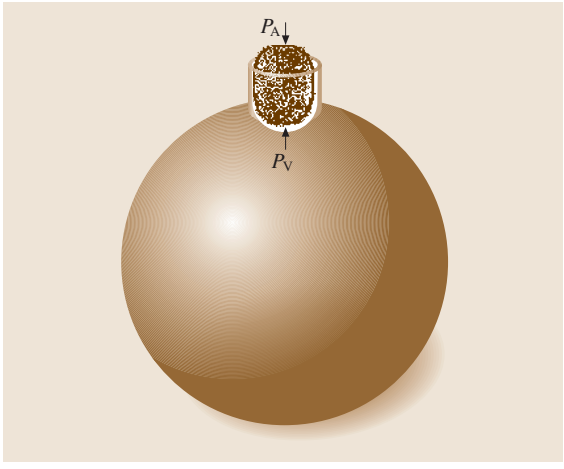


Fig. 17.11 Helmholtz resonator

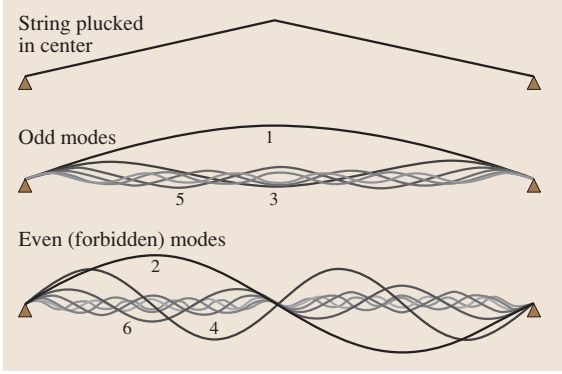


Fig. 17.12 Plucked string (*top*). The *center* shows sinusoidal *modes* of vibration of a center-plucked string. The *bottom* shows the even modes, which would not be excited by the center-plucked condition

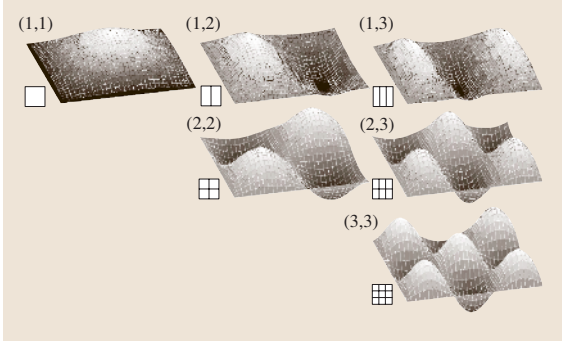


Fig. 17.13 Square-membrane modes

Amazon
EX1013 Page 730

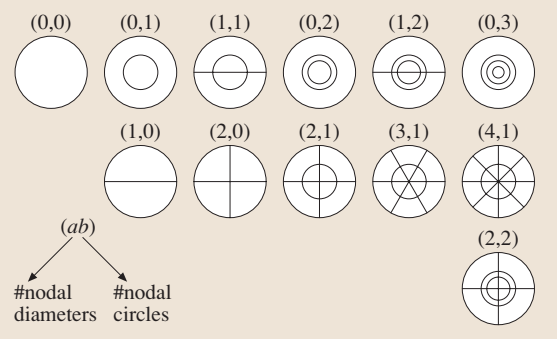


Fig. 17.14 Circular-membrane nodal lines

string vibration would have. These are spatial functions (sine as function of position along the string), but they also correspond to natural frequencies of vibration of the string. At the bottom of Fig. 17.12 is another set of modes that would not be possible with the center-plucked condition, because all of these even modes are restricted to have no vibration in the center of the string, and thus they could not contribute to the triangular shape of the string. These conditions of no displacement, corresponding to the zero crossings of the sine functions, are called *nodes*. Note that the end points are forced nodes of the plucked string system for all possible conditions of excitation.

Physical constraints on a system, such as the pinned ends of a string, and the center plucked initial shape, are known as *boundary conditions*. Spatial sinusoidal solutions like those shown in Fig. 17.12 are called *boundary solutions* (the legal sinusoidal modes of displacement and vibration) [17.13].

Just as one can use Fourier boundary methods to solve the one-dimensional (1-D) string, we can also extend boundary methods to two dimensions. Figures 17.13 and 17.14 show the first few vibration modes of uniform square and circular membranes. The small boxes at the lower left corners of each square modal-shape diagram depict the modes in a purely two-dimensional way, showing lines corresponding to the

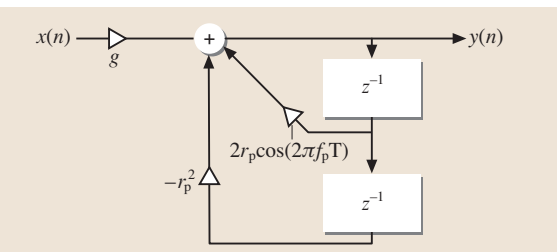


Fig. 17.15 Flexible parametric modal synthesis algorithm

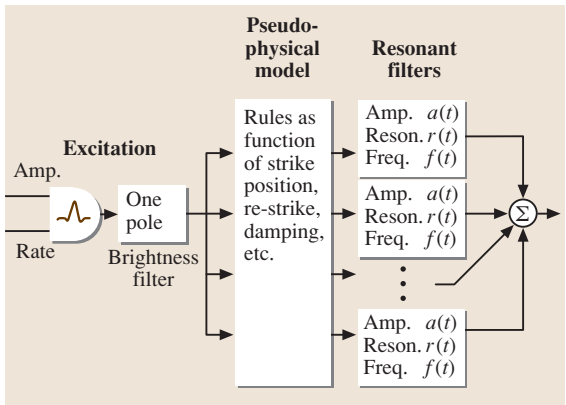


Fig. 17.16 Flexible parametric modal synthesis algorithm

spatial sinusoidal nodes (regions of no displacement vibration). This is how the circular modes are presented. The natural modes must obey the two-dimensional boundary conditions at the edges, but unlike the string, the square membrane modes are not integer-related harmonic frequencies. In fact they obey the relationship

$$f_{mn} = f_{11}\sqrt{\left[(m^2+n^2)/2\right]}\,, \qquad (17.12)$$

where m and n range from 1 to (potentially) infinity, and f_{11} is $c/2L$ (the speed of sound on the membrane divided by the square edge lengths). The circular modes are predictable from mathematics, but have a much more complex form than the square modes.

Unfortunately, circles, rectangles, and other simple geometries turn out to be the only ones for which the boundary conditions yield a closed-form solution in terms of spatial and temporal sinusoidal terms. However, we can measure and model the modes of any system by using the Fourier transform of the sound it produces, and looking for exponentially decaying sinusoidal components.

We can approximate the differential equation describing the mass/spring/damper system of (17.11) by replacing the derivatives (velocity as the derivative of position, and acceleration as the second derivative of position) with sampled time differences (normalized by the sampling interval T seconds). In doing so we would arrive at an equation that is a recursion in past values of $y(n)$, the position variable

$$[y(n)-2y(n-1)+y(n-2)]/T^2 + \frac{r/m[y(n)-y(n-1)]}{T+k/m\,y(n)} = 0\,. \qquad (17.13)$$

Note that if the values of mass, damping, spring constant, and sampling rate are constant, then the coeffi-

Amazon
EX1013 Page 731

cients $(2m+Tr)/(m+Tr+T^2k)$ for the single delay, and $m/(m+Tr+T^2k)$ for the twice-delayed signal applied to past y values are constant. Digital signal processing (DSP) engineers would note that a standard infinite impulse response (IIR) recursive filter as shown in Fig. 17.15 can be used to implement (17.13) (the Z^{-1} represents a single sample of delay). In fact, the second order two-pole feedback filter can be used to generate an exponentially decaying sinusoid, called a *phasor* in DSP literature [17.14]. The connection between the second order digital filter and the physical notion of a mode of vibration forms the basis for modal sound synthesis [17.15]. Figure 17.16 shows a general model for modal synthesis of struck/plucked objects, in which an impulsive excitation function is used to excite a number of filters that model the modes. Rules for controlling the modes as a function of strike position, striking object, changes in damping, and other physical constraints are included in the model.

17.5 Subtractive (Source–Filter) Synthesis

Subtractive synthesis uses a complex source wave, such as an impulse, a periodic train of impulses, or white noise, to excite spectral shaping filters. One of the earliest uses of electronic subtractive synthesis dates back to the 1920s and 1930s, with the invention of the *channel vocoder* (or VOiceCODER) [17.16]. In the this device, the spectrum is broken into sections called sub-bands, and the information in each sub-band is converted to a signal representing (generally slowly varying) power. The analyzed parameters are then stored or transmitted (potentially compressed) for reconstruction at another time or physical site. The parametric data representing the information in each sub-band can be manipulated in various ways, yielding transformations such as pitch or time shifting, spectral shaping, cross-synthesis, and other effects. Figure 17.17 shows a block diagram of a channel vocoder. The detected envelopes serve as control signals for a bank of bandpass synthesis filters (identical to the analysis filters used to extract the sub-band envelopes). The synthesis filters have gain inputs that are fed by the analysis control signals.

When used to encode and process speech, the channel vocoder explicitly makes an assumption that the signal being modeled is a single human voice. The *source analysis* block extracts parameters related to finer spectral details, such as whether the sound is pitched (vowel) or noisy (consonant or whispered). If the sound is pitched, the pitch is estimated. The overall energy in the signal is also estimated. These parameters become additional low-bandwidth control signals for the synthesizer. Intelligible speech can be synthesized using only a few hundred numbers per second. An example coding scheme might use eight channel gains + pitch + power, per frame, at 40 frames per second, yielding a total of only 400 numbers per second. The channel vocoder, as designed for speech coding, does not generalize to arbitrary sounds, and fails horribly when the source parameters deviate from expected harmonicity, reasonable pitch range, etc. But the ideas of sub-band decomposition, envelope detection, and driving a synthesis filter bank with control signals give rise to many other interesting applications and implementations of the vocoder concept.

A number of analog hardware devices were produced and sold as musical instrument processing devices in the 1970s and 1980s. These were called simply *vocoders*, but had a different purpose than speech coding. Figure 17.18 shows the block diagram of a cross-

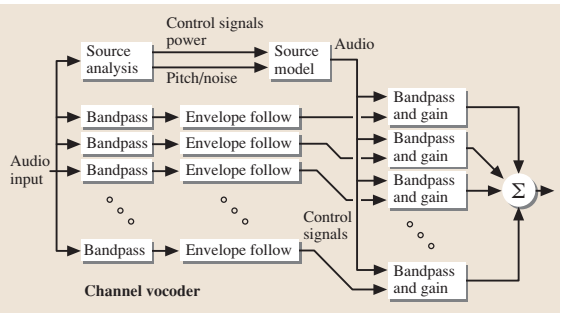


Fig. 17.17 Channel vocoder block diagram

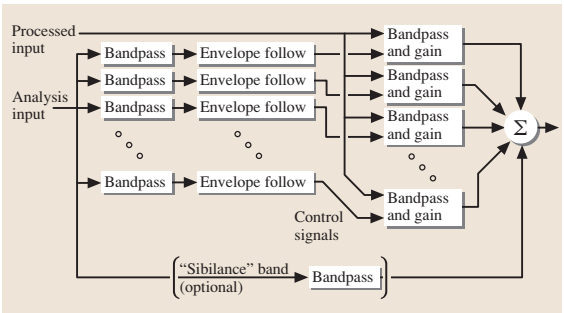


Fig. 17.18 Cross-synthesizing vocoder block diagram

Amazon
EX1013 Page 732

synthesizing vocoder. The main difference is that the parametric source has been replaced by an arbitrary audio signal. The cross-synthesizer can be used to make non-speech sounds talk or sing. A typical example would be feeding a voice into the analysis filter bank, and an electric guitar sound into the synthesis bank audio inputs to make a talking guitar. To be effective, the synthesis audio input should have suitably rich spectral content (for instance, distorted electric guitar works better than undistorted). Some vocoders allowed an additional direct *sibilance band* (like the highest band shown in Fig. 17.18) pass-through of the analyzed sound. This would allow consonants to pass through directly, creating a more intelligible speech-like effect.

Modal synthesis, as discussed before, is a form of subtractive synthesis, but the spectral characteristics of modes are sinusoidal and thus exhibit very narrow spectral peaks. For modeling the gross peaks in a spectrum, which could correspond to weaker resonances, we can exploit the same two-pole resonance filters. This type of source-filter synthesis has been very popular for voice synthesis.

Having its origins and applications in many different disciplines, time-series prediction is the task of estimating future sample values from prior samples. Linear prediction is the task of estimating a future sample (usually the next in the time series) by forming a linear combination of some number of prior samples. Linear predictive coding (LPC) is a technique that can automatically extract the gross spectral features and design filters to match those, and give us a source that we can use to drive the filters [17.17, 18]. Figure 17.19 shows linear prediction in block diagram form. The difference equation for a linear predictor is

$$y(n) = \hat{x}(n+1) = \sum_{i=0}^{m} a_i x(n-i)\,. \tag{17.14}$$

The task of linear prediction is to select the vector of predictor coefficients

$$A = (a_0, a_1, a_2, a_2, \ldots, a_m)$$

such that $\hat{x}(n+1)$ (the estimate) is as close as possible to $x(n+1)$ (the real sample) over a set of samples (often called a frame) $x(0)$ to $x(N-1)$. Usually, "close as possible" is defined by minimizing the mean square error (MSE):

$$MSE = 1/N \sum_{i=0}^{N-1} [\hat{x}(n+1) - x(n+1)]^2\,. \tag{17.15}$$

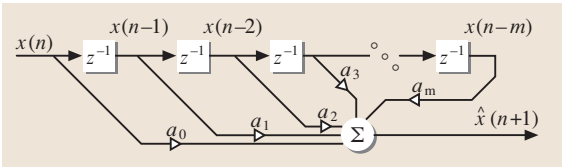


Fig. 17.19 A linear prediction filter

Many methods exist to arrive at the set of predictor coefficients a_i that yield a minimum MSE. The most common uses correlation or covariance data from each frame of samples to be predicted. The *autocorrelation* function, defined as

$$x \diamond x(n) = \sum_{i=0}^{N-1} x(i)x(i+n)\,. \tag{17.16}$$

Autocorrelation computes a time (lag) domain function that expresses the similarity of a signal to delayed versions of itself. It can also be viewed as the convolution of a signal with the time-reversed version of itself. Purely periodic signals exhibit periodic peaks in the autocorrelation function. Autocorrelations of white-noise sequences exhibit only one clear peak at the zero-lag position, and are small (zero for infinite-length sequences) for all other lags. Since the autocorrelation function contains useful information about a signal's similarity to itself at various delays, it can be used to compute the optimal set of predictor coefficients by forming the autocorrelation matrix

$$R = \begin{pmatrix} x\blacklozenge x(0) & x\blacklozenge x(1) & x\blacklozenge x(2) & \ldots & x\blacklozenge x(m) \\ x\blacklozenge x(1) & x\blacklozenge x(0) & x\blacklozenge x(1) & \ldots & x\blacklozenge x(m-1) \\ x\blacklozenge x(2) & x\blacklozenge x(1) & x\blacklozenge x(0) & \ldots & x\blacklozenge x(m-2) \\ x\blacklozenge x(3) & x\blacklozenge x(2) & x\blacklozenge x(1) & \ldots & x\blacklozenge x(m-3) \\ . & . & . & . & \\ x\blacklozenge x(m) & x\blacklozenge x(m-1) & x\blacklozenge x(m-2) & \ldots & x\blacklozenge x(0) \end{pmatrix} \tag{17.17}$$

We can get the least-squares predictor coefficients by forming

$$A = (a_0 a_1 a_2 a_3 \ldots a_m) = PR^{-1}\,, \tag{17.18}$$

where P is the vector of prediction correlation coefficients

$$P = (x\blacklozenge x(1) \;\; x\blacklozenge x(2) \;\; x\blacklozenge x(3) \;\; \ldots \;\; x\blacklozenge x(m+1))\,. \tag{17.19}$$

For low-order LPC (delay order of 6–20 or so), the filter fits itself to the coarse spectral features, and the

Amazon
EX1013 Page 733

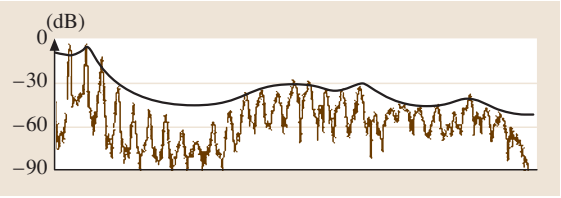


Fig. 17.20 Tenth order LPC filter fit to a voiced 'ooo' spectrum

residue contains the remaining part of the sound that cannot be linearly predicted. A common and popular use of LPC is for speech analysis, synthesis, and compression. The reason for this is that the voice can be viewed as a source-filter model, where a spectrally rich input (pulses from the vocal folds or noise from turbulence) excites a filter (the resonances of the vocal tract). LPC is another form of spectral vocoder (voice coder) as discussed previously, but since LPC filters are not fixed in frequency or shape, fewer bands are needed to model the changing speech spectral shape dynamically.

LPC speech analysis/coding involves processing the signal in blocks and computing a set of filter coefficients for each block. Based on the slowly varying nature of speech sounds (the speech articulators can only move so fast), the coefficients are relatively stable for milliseconds at a time (typically 5–20 ms is used in speech coders). If we store the coefficients and information about the residual signal for each block, we will have captured many of the essential aspects of the signal. Figure 17.20 shows an LPC fit to a speech spectrum. Note that the fit is better at the peak locations than at the valleys. This is due to the least-squares minimization criterion. Missing the mark on low-amplitude parts of the spectrum is not as important as missing it on high-amplitude parts. This is fortunate for audio signal modeling, in that the human auditory system is more sensitive to spectral peaks (formants, poles, resonances) than valleys (zeroes, anti-resonances).

Once we have performed LPC on speech, if we inspect the residual we might note that it is often a stream

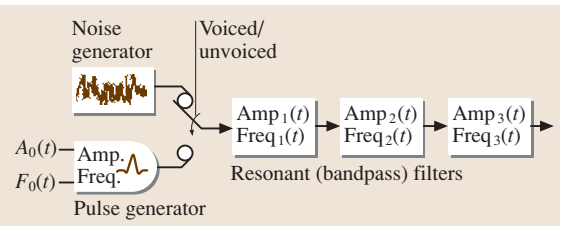


Fig. 17.21 Cascade factored formant subtractive synthesizer

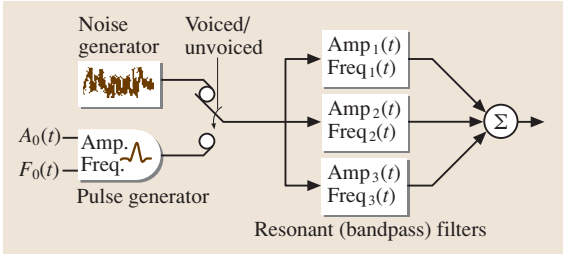


Fig. 17.22 Parallel factored formant subtractive synthesizer

of pulses for voiced speech, or white noise for unvoiced speech. Thus, if we store parameters about the residual, such as whether it is periodic pulses or noise, the frequency of the pulses, and the energy in the residual, then we can get back a signal that is very close to the original. This is the basis of much of modern speech compression. If a signal is entirely predictable using a linear combination of prior samples, and if the predictor filter is doing its job perfectly, we should be able to hook the output back to the input and let the filter predict the rest of the signal automatically. This form of filter, with feedback from output to input, is called *recursive*. The recursive LPC reconstruction is sometimes called *all pole*, to reflect the high-gain poles corresponding to the primary resonances of the vocal tract. The poles do not capture all of the acoustic effects going on in speech, however, such as zeroes that are introduced in nasalization, aerodynamic effects, etc. However, as mentioned before, since our auditory systems are most sensitive to peaks (poles), the LPC representation does quite a good job of capturing the most important aspects of speech spectra.

In recursive resynthesis form, any deviation of the predicted signal from the actual original signal will show up in the error signal, so if we excite the recursive LPC reconstruction filter with the residual signal itself, we can get back the original signal exactly. This is a form of identity analysis/resynthesis, performing deconvolution to separate the source from the filter, and using the residue to excite the filter to arrive at the original signal, by convolution, or using the residue as the input to the filter.

For computer music composition, LPC can be used for cross-synthesis, by keeping the time-varying filter coefficients, and replacing the residual with some other sound source. Using the parametric source model also allows for flexible time and pitch shifting, without modifying the basic timbre. The voiced pulse period can be modified, or the frame rate update of the filter coefficients can be modified, independently. So it is easy to speed up a speech sound while making the pitch lower

Amazon
EX1013 Page 734

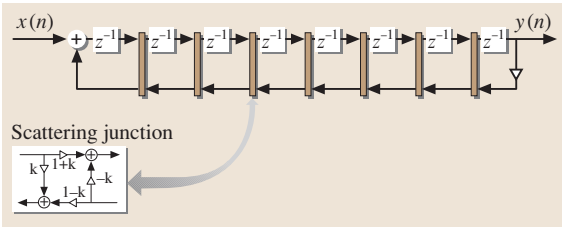


Fig. 17.23 Ladder filter implementation of an all-pole LPC filter

and still retaining the basic spectral shapes of all vowels and consonants.

In decomposing signals into a source and a filter, LPC can be a marvelous aid in analyzing and understanding some sound-producing systems. The recursive LPC reconstruction filter can be implemented in a variety of ways. Three different filter forms are commonly used to perform subtractive voice synthesis [17.19]. The filter can be implemented in series (cascade) as shown in Fig. 17.21, factoring each resonance into a separate filter block with control over center frequency, width, and amplitude. The filter can also be implemented in parallel (separate sub-band sections of the spectrum added together), as shown in Fig. 17.22.

One additional implementation of the resonant filter is the ladder filter structure, which carries with it a notion of one-dimensional spatial propagation as well [17.20]. Figure 17.23 shows a ladder filter realization of a 10th-order IIR filter. We will refer to this type of filter later when we investigate some 1-D waveguide physical models.

17.6 Frequency Modulation (FM) Synthesis

Wave-shaping synthesis involves warping a simple (usually a sine or sawtooth wave) waveform with a nonlinear function or lookup table. One popular form of wave-shaping synthesis, called frequency modulation (FM), uses sine waves for both input and warping waveforms [17.21–23]. Frequency modulation relies on modulating the frequency of a simple periodic waveform with another simple periodic waveform. When the frequency of a sine wave of average frequency f_c (called the carrier), is modulated by another sine wave of frequency f_m (called the modulator), sinusoidal sidebands are created at frequencies equal to the carrier frequency plus and minus integer multiples of the modulator frequency. Figure 17.24 shows a block diagram for simple FM synthesis (one sinusoidal carrier and one sinusoidal modulator). Mathematically, FM is expressed as

$$y(t) = \sin[2\pi t f_c + \Delta f_c \cos(2\pi t f_m)] \,. \qquad (17.20)$$

The index of modulation I is defined as $\Delta f_c / f_c$. The equation as shown is actually phase modulation, but FM and PM can be related by a differentiation/integration of the modulation, and thus do not differ for cosine modulation.

Carson's rule (a rule of thumb) states that the number of significant bands on each side of the carrier frequency (sidebands) is roughly equal to $I + 2$. For example, a carrier sinusoid of frequency 600 Hz, a modulator sinusoid of frequency 100 Hz, and a modulation index of 3 would produce sinusoidal components of frequencies 600, 700, 500, 800, 400, 900, 300, 1000, 200, 1100, and 100 Hz. Inspecting these components reveals that a harmonic spectrum with 11 significant harmonics, based on a fundamental frequency of 100 Hz can be produced by using only two sinusoidal generating functions. Figure 17.25 shows the spectrum of this synthesis.

Selecting carrier and modulator frequencies that are not related by simple integer ratios yields an inharmonic spectrum. For example, a carrier of 500 Hz, modula-

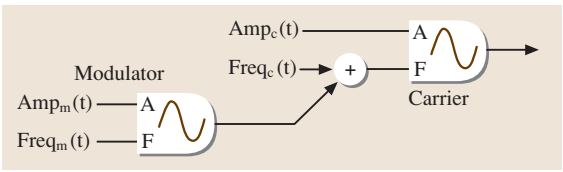


Fig. 17.24 Simple FM (one carrier and one modulator sine wave) synthesis

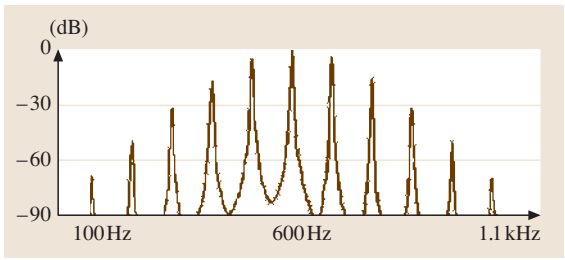


Fig. 17.25 Simple FM with 600 Hz carrier, 100 Hz modulator, and index of modulation of 3

Amazon
EX1013 Page 735

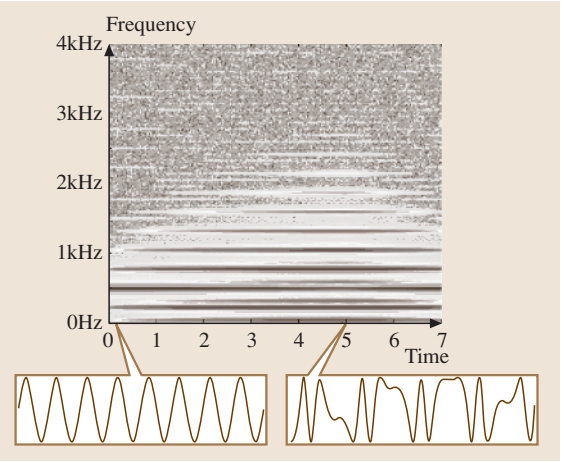


Fig. 17.26 Inharmonic simple FM with 500 Hz carrier and 273 Hz modulator. The index of modulation is ramped from zero to 5 then back to zero

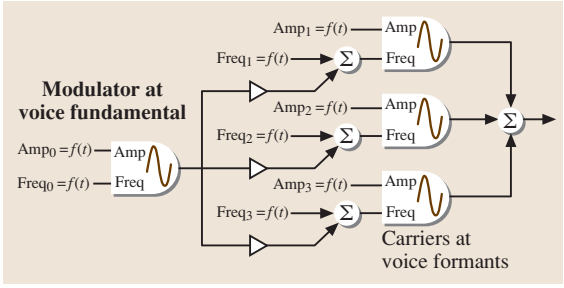


Fig. 17.27 Frequency-modulation voice synthesis diagram

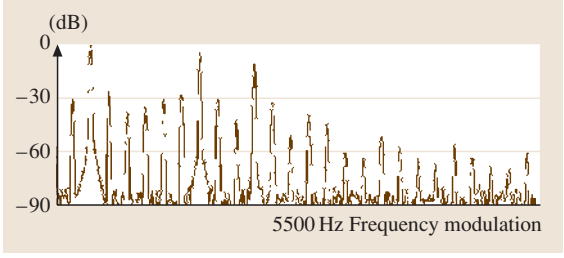


Fig. 17.28 Frequency-modulation voice spectrum

tor of 273 Hz, and an index of 5 yields frequencies of 500 (carrier), 227, 46, 319, 592, 865, 1138, 1411 (negative sidebands), and 773, 1046, 1319, 1592, 1865, 2138, 2411 (positive sidebands). Figure 17.26 shows a spectrogram of this FM tone, as the index of modulation is ramped from zero to 5. The synthesized waveforms at $I = 0$ and $I = 5$ are shown as well.

By setting the modulation index high enough, huge numbers of sidebands are generated, and the aliasing and addition of these (in most cases) results in noise. By careful selection of the component frequencies and index of modulation, and combining multiple carrier/modulator pairs, many spectra can be approximated using FM. However, the amplitudes and phases (described by Bessel functions) of the individual components cannot be independently controlled, so FM is not a truly generic waveform or spectral synthesis method. Using multiple carriers and modulators, connection topologies (algorithms) have been designed for the synthesis of complex sounds such as human voices, violins, brass instruments, percussion, etc. Figures 17.27 and 17.28 show an FM connection topology for human voice synthesis, and a typical resulting spectrum [17.24]. Because of the extreme efficiency of FM, it became popular in the 1980s as a music synthesizer algorithm. Later FM found wide use in early PC sound cards.

17.7 FOFs, Wavelets, and Grains

In a source/filter vocal model such as LPC or parallel/cascade formant subtractive synthesis, periodic impulses are used to excite resonant filters to produce vowels. We could construct a simple alternative model using three, four, or more tables storing the impulse responses of the individual vowel formants. Note that it is not necessary for the tables to contain pure exponentially decaying sinusoids. We could include any aspects of the voice source, etc., as long as those effects are periodic. Formes d'onde formantiques (FOFs, French for formant wave functions) were created for voice synthesis using such tables, overlapped and added at the repetition period of the voice source [17.25]. Figure 17.29 depicts FOF synthesis of a vowel. FOFs are composed of a sinusoid at the formant center frequency, with an amplitude that rises rapidly upon excitation, then decays exponentially. The control parameters define the center frequency and bandwidth of the formant being modeled, and the rate at which the FOFs are generated and added determines the fundamental frequency of the voice.

Note that each individual FOF is a simple wavelet (local and compact wave both in frequency and time).

Amazon
EX1013 Page 736

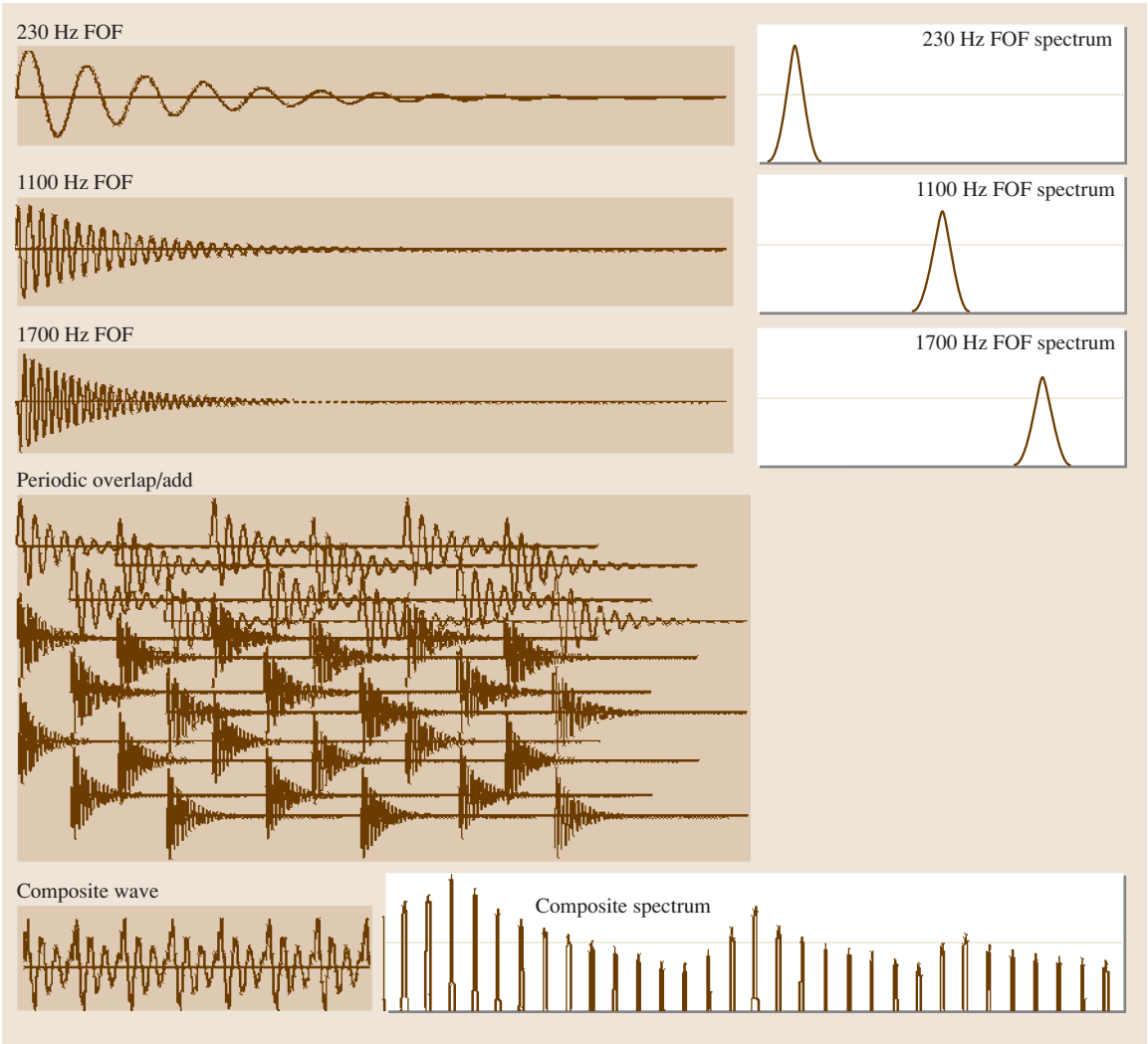


Fig. 17.29 FOF synthesis of a vowel

A family of filter-based frequency transforms known as *wavelet transforms* have been used for the analysis and synthesis of sound. Instead of being based on steady sinusoids like the Fourier transform, wavelet transforms are based on the decomposition of signals into fairly arbitrary wavelets [17.26]. These wavelets can be thought of as mathematical sound grains. Some benefits of wavelet transforms over Fourier transforms are that they can be implemented using fairly arbitrary filter criteria, on a logarithmic frequency scale rather than a linear scale as in the DFT, and that time resolution can be a function of the frequency range of interest. This latter point means that we can say accurate things about high frequencies as well as low. This contrasts to the Fourier transform, which requires the analysis window width to be the same for all frequencies. This means that we must either average out lots of the interesting high-frequency time information in favor of being able to resolve low-frequency information (large window), or opt for good time resolution (small window) at the expense of low-frequency resolution, or perform multiple transforms with different sized windows to catch both time and frequency details. There are a number of fast wavelet transform techniques that allow the subband decomposition to be accomplished in essentially $N \log_2(N)$ time just like the FFT.

Amazon
EX1013 Page 737

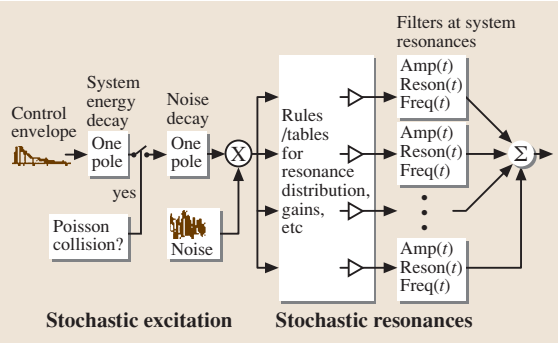


Fig. 17.30 Complete PhISEM model showing stochastic resonances

Most of classical physics can be modeled as objects interacting with each other. Lots of small objects are usually called particles. Granular synthesis involves cutting sound into grains and reassembling them by mixing [17.27]. The grains usually range in length from 5 to 100 ms. The reassembly can be systematic, but often granular synthesis involves randomized grain sizes, locations, and amplitudes. The transformed result usually bears some characteristics of the original sound, just as a mildly blended mixture of fruits still bears some attributes of the original fruits, as well as taking on new attributes due to the mixture. Granular synthesis is mostly used as a compositional and musical-effect type of signal processing, but some also take a physically motivated look at sound grains.

The physically informed stochastic event modeling (PhISEM) algorithm is based on pseudo-random overlapping and adding of parametrically synthesized sound grains [17.28]. At the heart of PhISEM algorithms are particle models, characterized by basic Newtonian equations governing the motion and collisions of point masses as can be found in any introductory physics textbook. By reducing the physical interactions of many particles to their statistical behavior, exhaustive calculation of the position, and velocity of each individual particle can be avoided. By factoring out the resonances of the system, the wavelets can be shortened to impulses or short bursts of exponentially decaying noise. The main PhISEM assumption is that the sound-producing particle collisions follow a Poisson process. Another assumption is that the system energy decays exponentially (the decay of the sound of a maraca after being shaken once, for example). Figure 17.30 shows the PhISEM algorithm block diagram.

The PhISEM maraca synthesis algorithm requires only two random number calculations, two exponential decays, and one resonant filter calculation per sample. Other musical instruments that are quite similar to the maraca include the sekere and cabasa (afuche). Outside the realm of multicultural musical instruments, there are many real-world particle systems that exhibit one or two fixed resonances like the maraca. A bag or box of candy or gum, a salt shaker, a box of wooden matches, and gravel or leaves under walking feet all fit well within this modeling technique.

In contrast to the maraca and guiro-like gourd resonator instruments, which exhibit one or two weak resonances, instruments such as the tamborine (timbrel) and sleigh bells use metal cymbals, coins, or bells suspended on a frame or stick. The interactions of the metal objects produce much more pronounced resonances than the maraca-type instruments, but the Poisson event and exponential system energy statistics are similar enough to justify the use of the PhISEM algorithm for synthesis. To implement these in PhISEM, more filters are used to model the individual partials, and at each collision, the resonant frequencies of the filters are randomly set to frequencies around the main resonances. Other sounds that can be modeled using stochastic filter resonances include bamboo wind chimes (related to a musical instrument as well in the Javanese anklung) [17.29].

17.8 Physical Modeling (The Wave Equation)

One way to model a vibrating string physically would be to build it up as a chain of coupled masses and springs, as shown in Fig. 17.31. The masses slide on frictionless guide rods to restrict their motion to transverse (at right angles to the length), as is the case in the ideal string. In fact, some synthesis projects have used mass/spring/damper systems to represent arbitrary objects [17.30].

If we force the masses to move only up and down, and restrict them to small displacements, we do not actually have to solve the acceleration of each mass as a function of the spring forces pushing them up and down. There is a simple differential equation that completely describes the motions of the ideal string:

$$d^2y/dx^2 = (1/c^2)d^2y/dt^2 \,. \qquad (17.21)$$

Amazon
EX1013 Page 738

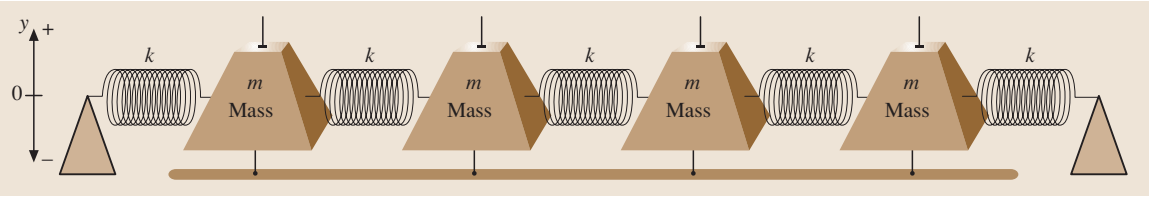


Fig. 17.31 Mass/spring string network, with frictionless *guide rods* to restrict motion to one dimension

This equation (called the wave equation) means that the acceleration (up and down) of any point on the string is equal to a constant times the curvature of the string at that point. The constant c is the speed of wave motion on the string, and is proportional to the square root of the string tension, and inversely proportional to the square root of the mass per unit length. This equation could be solved numerically, by sampling it in both time and space, and using the difference approximations for acceleration and curvature (much like we did with the mass/spring/damper system earlier). With boundary conditions (like rigid terminations at each end), we could express the solution of this equation as a Fourier series, as we did earlier in graphical form (Fig. 17.12). However, there is one more wonderfully simple solution to (17.21), given by

$$y(x,t) = y_{\mathrm{L}}(t + x/c) + y_{\mathrm{r}}(t - x/c)\,. \qquad (17.22)$$

This equation says that any vibration of the string can be expressed as a combination of two separate traveling waves, one traveling left (y_{L}) and one traveling right (y_{r}). They move at the rate c, which is the speed of sound propagation on the string. For an ideal (no damping) string and ideally rigid boundaries at the ends, the wave reflects with an inversion at each end, and would travel back and forth indefinitely. This view of two traveling waves summing to make a displacement wave gives rise to the waveguide filter technique of modeling the vibrating string [17.31, 32]. Figure 17.32 shows a waveguide filter model of the ideal string. The two delay lines model the propagation of left- and right-going traveling waves. The conditions at the ends model the reflection of the traveling waves at the ends. The −1 in the left reflection models the reflection with inversion of a displacement wave when it hits an ideally rigid termination (like a fret on a guitar neck). The −0.99 on the right reflection models the slight loss that happens when the wave hits a termination that yields slightly (like the bridge of the guitar which couples the string motion to the body), and models all other losses the wave might experience (internal damping in the string, viscous losses as the string cuts the air, etc.) in making its round-trip path around the string.

Figure 17.33 shows the waveguide string as a digital filter block diagram. The $Z^{-P/2}$ blocks represent a delay equal to the time required for a wave to travel down the string. Thus a wave completes a round trip each P samples (down and back), which is the fundamental period of oscillation of the string, expressed in samples. Initial conditions can be injected into the string via the input $x(n)$. The output $y(n)$ would yield the right-going traveling wave component. Of course, neither of these conditions is actually physical in terms of the way a real

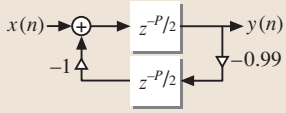


Fig. 17.33 Filter view of a waveguide string

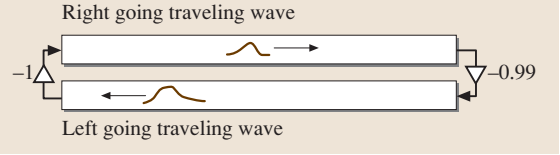


Fig. 17.32 Waveguide string modeled as two delay lines

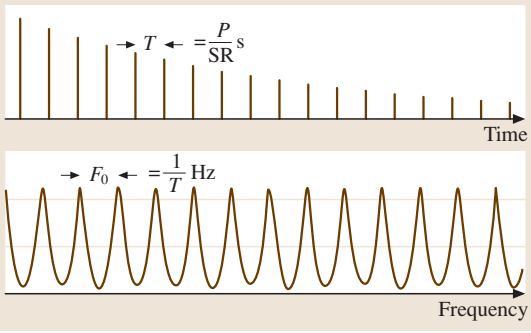


Fig. 17.34 Impulse response and spectrum of a comb (string) filter

Amazon
EX1013 Page 739

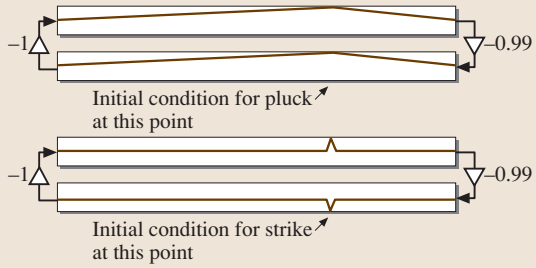


Fig. 17.35 Waveguide pluck and strike initial conditions

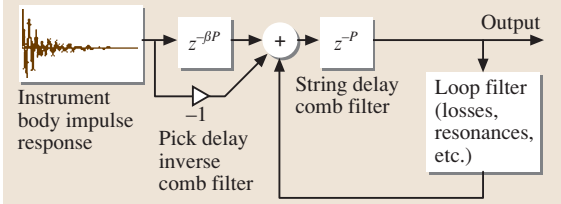


Fig. 17.37 Commuted plucked string model

string is plucked and listened to, but feeding the correct signal into x is identical to loading the delay lines with a predetermined shape.

The impulse response and spectrum of the filter shown in Fig. 17.32 is shown in Fig. 17.34. As would be expected, the impulse response is an exponentially decaying train of pulses spaced $T = P/\text{SRATE}$ seconds apart, and the spectrum is a set of harmonics spaced $F_0 = 1/T$ Hz apart. This type of filter response and spectrum is called a *comb filter*, so named because of the comb-like appearance of the time-domain impulse response, and of the frequency-domain harmonics.

The two delay lines taken together are called a *waveguide filter*. The sum of the contents of the two delay lines is the displacement of the string, and the difference of the contents of the two delay lines is the velocity of the string. If we wish to pluck the string, we simply need to load 1/2 of the initial-shape magnitude into each of the upper and lower delay lines. If we wish to strike the string, we would load in an initial velocity by entering a positive pulse into one delay line and a negative pulse into the other (difference = initial velocity, sum = initial position = 0). These conditions are shown in Fig. 17.35.

Figure 17.36 shows a relatively complete model of a plucked string simulation system using digital filters. The inverse comb filters model the nodal effects of picking, and the output of an electrical pickup, emphasizing certain harmonics and forbidding others based on the pick (pickup) position [17.33]. Output channels for the pickup position and body radiation are provided separately. A solid-body electric guitar would have no direct radiation and only pickup output(s), while a purely acoustic guitar would have no pickup output, but possibly a family of directional filters to model body radiation in different directions [17.34].

We should note that everything in the block diagram of Fig. 17.35 is linear and time invariant (LTI). Linearity means that for input x and output y, if $x_1(t) \rightarrow y_1(t)$ and $x_1(t) \rightarrow y_2(t)$, then $x_2(t)+x_2(t) \rightarrow y_1(t)+y_2(t)$, for any functions $x_{1,2}$ (signals add and scale linearly). Time invariant means that if $x(t) \rightarrow y(t)$, then $x(t-T) \rightarrow y(t-T)$ for any $x(t)$ and any T (the system doesn't change with time). Given that the LTI condition is satisfied, we can then happily (legally) commute (swap the order) of any of the blocks. For example, we could put the pick position filter at the end, or move the body radiation filter to the beginning. Since the body radiation filter is just the impulse response of the instrument body, we can actually record the impulse response with the strings damped, and use that as the input to the string model. This is called *commuted synthesis* [17.35]. Figure 17.37 shows a commuted synthesis plucked string model.

Adding a model of bowing friction allows the string model to be used for the violin and other bowed strings. This focused nonlinearity is what is responsible for turning the steady linear motion of a bow into an oscillation of the string [17.36, 37]. The bow sticks to the string

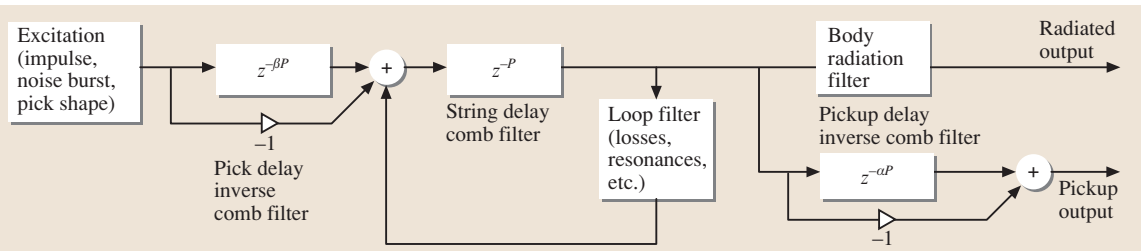


Fig. 17.36 Fairly complete digital filter simulation of the plucked string system

Amazon
EX1013 Page 740

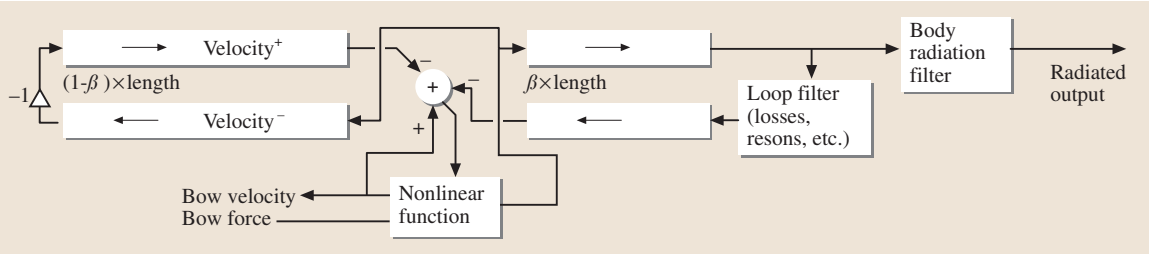


Fig. 17.38 Bowed string model

for a while, pulling it along, then the forces become too great and the string breaks away, flying back toward rest position. This process repeats, yielding a periodic oscillation. Figure 17.38 shows a simple bowed string model, in which string velocity is compared to bow velocity, then put through a nonlinear friction function controlled by bow force. The output of the nonlinear function is the velocity input back into the string.

In mathematically describing the air within a cylindrical acoustic tube (like a trombone slide, clarinet bore, or human vocal tract), the defining equation is:

$$\mathrm{d}^2P/\mathrm{d}x^2 = \left(1/c^2\right)\left(\mathrm{d}^2P/\mathrm{d}t^2\right) \qquad (17.23)$$

which we note has exactly the same form as (17.21), except that the displacement y is replaced by the pressure P. A very important paper in the history of physical modeling [17.36] noted that many musical instruments can be characterized as a linear resonator, modeled by filters such as all-pole resonators or waveguides, and a single nonlinear oscillator like the reed of the clarinet, the lips of the brass player, the jet of the flute, or the bow–string friction of the violin. Since the wave equation says that we can model a simple tube as a pair of bidirectional delay lines (waveguides), then we can build models using this simple structure. If we would like to do something interesting with a tube, we could use it to build a flute or clarinet. Our simple clarinet model might look like the block diagram shown in Fig. 17.39.

To model the reed, we assume that the mass of the reed is so small that the only thing that must be considered is the instantaneous force on the reed (spring). The pressure inside the bore P_b is the calculated pressure in our waveguide model, the mouth pressure P_m is an external control parameter representing the breath pressure inside the mouth of the player (Fig. 17.40). The net force acting on the reed/spring can be calculated as the difference between the internal and external pressures, multiplied by the area of the reed (pressure is force per unit area). This can be used to calculate a reed opening position from the spring constant of the reed. From the reed opening, we can compute the amount of pressure that is allowed to leak into the bore from the player's mouth. If bore pressure is much greater than mouth pressure, the reed opens far. If the mouth pressure is much greater than the bore pressure, the reed slams shut. These two extreme conditions represent an asym-

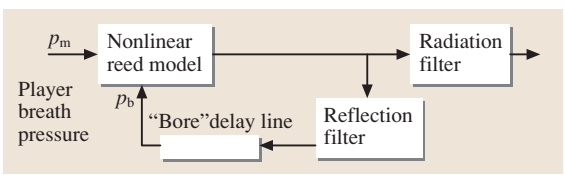


Fig. 17.39 Simple clarinet model

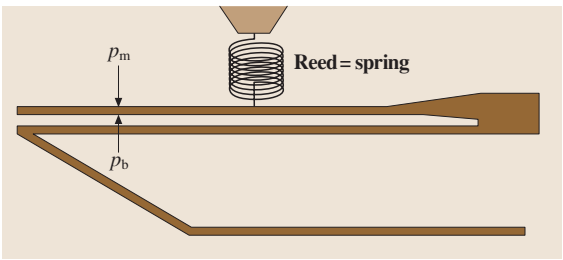


Fig. 17.40 Reed model

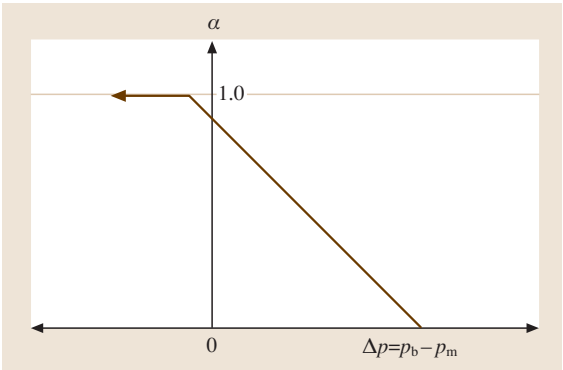


Fig. 17.41 Reed reflection table

Amazon
EX1013 Page 741

metric nonlinearity in the reed response. Even a grossly simplified model of this nonlinear spring action results in a pretty good model of a clarinet [17.37]. Figure 17.41 shows a plot of a simple reed reflection function (as seen from within the bore) as a function of differential pressure. Once this nonlinear signal-dependent reflection coefficient is calculated (or looked up in a table), the right-going pressure injected into the bore can be calculated as $P_b^+ = \alpha P_b^- + (1-\alpha)P_m$.

The clarinet is open at the bell end, and essentially closed at the reed end. This results in a reflection with inversion at the bell and a reflection without inversion (plus any added pressure from the mouth through the reed opening) at the reed end. These odd boundary conditions cause odd harmonics to dominate in the clarinet spectrum.

We noted that the ideal string equation and the ideal acoustic tube equation are essentially identical. Just as there are many refinements possible to the plucked string model to make it more realistic, there are many possible improvements for the clarinet model. Replacing the simple reed model with a variable mass/spring/damper allows the modeling of a lip reed as is found in brass instruments. Replacing the reed model with an air-jet model allows the modeling of flute and recorder-like instruments. With all wind or friction (bowed) excited resonant systems, adding a little noise in the reed/jet/bow region adds greatly to the quality (and behavior) of the synthesized sound.

In an ideal string or membrane, the only restoring force is assumed to be the tension under which they are stretched. We can further refine solid systems such as strings and membranes to model more-rigid objects, such as bars and plates, by noting that the more-rigid objects exhibit internal restoring forces due to their stiffness. We know that if we bend a stiff string, it wants to return back to straightness even when there is no tension on the string. Cloth string or thread has almost no stiffness. Nylon and gut strings have some stiffness, but not as much as steel strings. Larger-diameter strings have more stiffness than thin ones. In the musical world, piano strings exhibit the most stiffness. Stiffness results in the restoring force being higher (thus the speed of sound propagation as well) for high frequencies than for low. So the traveling wave solution is still true in stiff systems, but a frequency-dependent propagation speed is needed

$$y(x,t) = y_l\left[t + x/c(f)\right] + y_r\left[t - x/c(f)\right] . \quad (17.24)$$

and the waveguide filter must be modified to simulate frequency-dependent delay, as shown in Fig. 17.42.

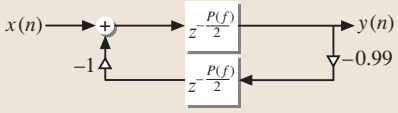


Fig. 17.42 Stiffness-modified waveguide string filter

For basic stiff strings, a function that predicts the frequencies of the partials has this form:

$$f_n = n\left[f_0\left(1 + Bn^2\right)\right] , \quad (17.25)$$

where B is a number slightly greater than 0, equal to zero for perfect harmonicity (no stiffness), and increasing for increasing stiffness. Typical values of B are 0.00001 for guitar strings, and 0.004 or so for piano strings. This means that $P(f)$ should modify P by the inverse of the $(1 + Bn^2)^{1/2}$ factor.

Unfortunately, implementing the $Z^{-P(f)/2}$ frequency-dependent delay function is not simple, especially for arbitrary functions of frequency. One way to implement the $P(f)$ function is by replacing each of the Z^{-1} with a first-order all-pass (phase) filter, as shown in Fig. 17.43 [17.33]. The first-order all-pass filter has one pole and one zero, controlled by the same coefficient. The all-pass implements a frequency-dependent phase delay, but exhibits a gain of 1.0 for all frequencies. The coefficient α can take on values between -1.0 and 1.0. For $\alpha = 0$, the filter behaves as a standard unit delay. For $\alpha > 0.0$, the filter exhibits delays longer than one sample, increasingly long for higher frequencies. For $\alpha < 0.0$ the filter exhibits delays shorter than one sample, decreasingly so for high frequencies.

It is much less efficient to implement a chain of all-pass filters than a simple delay line. But for weak stiffness it is possible that only a few all-pass sections will provide a good frequency-dependent delay. Another option is to implement a higher-order all-pass filter, designed to give the correct stretching of the upper frequencies, added to simple delay lines to give the correct longest bulk delay required.

For very stiff systems such as rigid bars, a single waveguide with all-pass filters is not adequate to give enough delay, or far too inefficient to calculate. A technique called *banded waveguides* employs

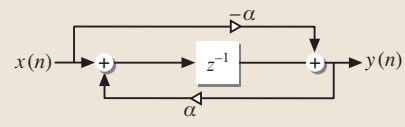


Fig. 17.43 First-order all-pass filter

Amazon
EX1013 Page 742

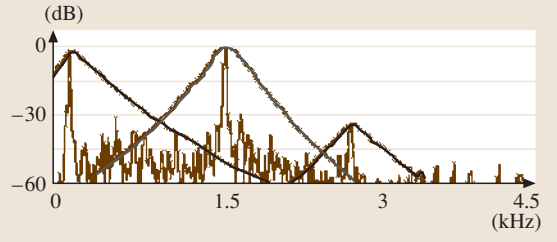


Fig. 17.44 Banded decomposition of a struck-bar spectrum

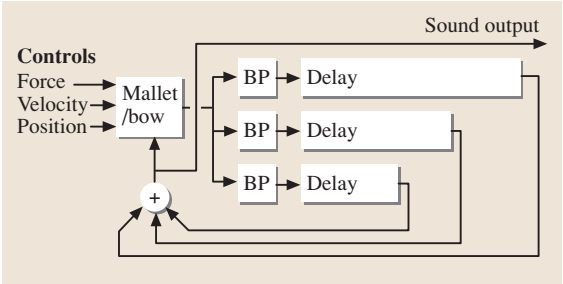


Fig. 17.45 Banded waveguide model

sampling in time, space, and frequency to model stiff one-dimensional systems [17.38]. This can be viewed as a hybrid of modal and waveguide synthesis, in that each waveguide models the speed of sound in the region around each significant mode of the system. As an example, Fig. 17.44 shows the spectrum of a struck marimba bar, with additional bandpass filters superimposed on the spectrum, centered at the three main modes. In the banded waveguide technique, each mode is modeled by a bandpass filter, plus a delay line to impose the correct round-trip delay, as shown in Fig. 17.45.

17.9 Music Description and Control

The musical instrument digital interface (MIDI) standard, adopted in 1984, revolutionized electronic music, and also profoundly affected the computer industry [17.39]. A simple two-wire serial electrical connection standard allows interconnection of musical devices and computers over cable distances of up to 15 m (longer over networks and extensions to the basic MIDI standard). The MIDI software protocol is best described as musical keyboard gestural, meaning that the messages carried over MIDI are the gestures that a pianist or organist uses to control a traditional keyboard instrument. There is no time information contained in the basic MIDI messages; they are intended to take effect as soon as they come over the wire. As such it is a real-time gestural protocol, and can be adapted for real-time non-musical sound synthesis applications. There are complaints about MIDI, however, mostly related to limited control bandwidth (approximately 1000–1500 continuous messages per second maximum), and the keyboard-centric bias of the messages.

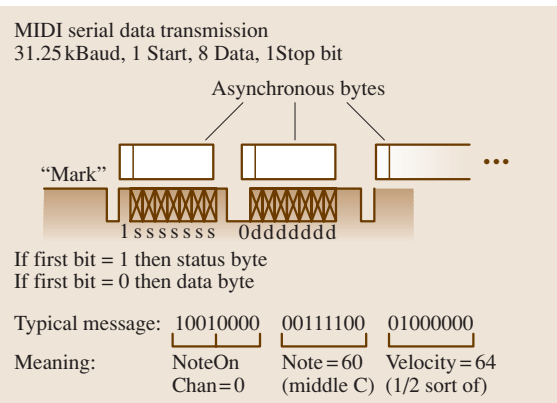


Fig. 17.46 MIDI software transmission protocol

Basic MIDI message types include NoteOn and NoteOff, sustain pedal up and down, amount of modulation (vibrato), pitch bend, key pressure (also called AfterTouch), breath pressure, volume, pan, balance, reverberation amount, and others. NoteOn and NoteOff messages carry a note number corresponding to a particular piano key, and a velocity corresponding to how hard that key is hit. Figure 17.46 shows the software serial format of MIDI, and gives an example of a NoteOn message.

Another MIDI message is program change, which is used to select the particular sound being controlled in the synthesizer. MIDI provides for 16 channels, and the channel number is encoded into each message. Each channel is capable of one type of sound, but possibly many simultaneous *voices* (a voice is an individual sound) of that same sound. Channels allow instruments and devices to all listen on the same network, and choose to respond to the messages sent on particular channels.

General MIDI (1991), and the standard MIDI file specifications serve to extend MIDI and made it even more popular [17.40]. General MIDI helps to assure

Amazon
EX1013 Page 743

the performance of MIDI on different synthesizers, by specifying that a particular program (the algorithm for producing the sound of an instrument) number must call up a program the approximates the same instrument sound on all general MIDI-compliant synthesizers. There are 128 such defined instrument sounds available on MIDI channels 1–9 and 11–16. For example, MIDI program 1 is grand piano, and 57 is trumpet. Some of the basic instrument sounds are sound effects. Program #125 is a telephone ring, #126 is helicopter, #127 is applause, and #128 is a gunshot sound. On general MIDI channel 10, each note is mapped to a different percussion sound. For example, the bass drum is note number 35, and the cowbell is note number 56.

The MIDI file formats provide a means for the standardized exchange of musical/sonic information. A MIDI level 0 file carries the basic information that would be carried over a MIDI serial connection, which is the basic bytes in order, with added time stamps. Time stamps allow a very simple program to play back MIDI files. A MIDI level 1 file is more suited to manipulation by a notation program or MIDI sequencer (a form of multi-track recorder program that records and manipulates MIDI events). Data is arranged by *tracks*, (each track transmits on only one MIDI channel) which are the individual instruments in the virtual synthesized orchestra. Meta messages allow for information which is not actually required for a real-time MIDI playback, but might carry information related to score markings, lyrics, composer and copyright information, etc.

Recognizing the limitations of fixed PCM samples in music synthesizers, the downloadable sounds (DLS) specification was added to MIDI in 1999. This provides a means to load PCM into a synthesizer, then use MIDI commands to play it back. Essentially any instrument/percussion sound can be replaced with arbitrary PCM, making it possible to control large amounts of recorded sounds. The emergence of software synthesizers, on PC platforms has made the use of DLS more feasible than in earlier days when the samples were contained in a static ROM within a hardware synthesizer.

The simplicity and ubiquity of MIDI, with thousands of supporting devices and software systems, leads many to use it for all sorts of projects and systems. While not being directly compatible with popular music synthesizers, using MIDI for the control of synthesis still allows the use of the large variety of MIDI control signal generators and processors, all of which emit standard and user-programmable MIDI messages. Examples include MIDI sequencers and other programs for general-purpose computers, as well as breath controllers, drum controllers, and fader boxes with rows of sliders and buttons.

Sadly, one thing that is still missing from MIDI at the time of writing is the ability to download arbitrary synthesis algorithms. Beyond MIDI, there are a number of emerging standards for sound and multimedia control. As mentioned in the last section, one of the things MIDI lacks is high-bandwidth message capability for handling many streams of gestural/model control data. Another complaint about MIDI relates to the small word sizes, such as semitone quantized pitches (regular musical notes), and only 128 levels of velocity, volume, pan, and other controllers. MIDI does support some extended precision messages, allowing 14 bits rather than 7. Other specifications have proposed floating-point representations, or at least larger integer word sizes. A number of new specifications and systems have arisen recently.

One system for describing sounds themselves is called the sound description interchange format (SDIF 1999) [17.41], originally called the spectral description interchange file (SDIF) format. It is largely intended to encapsulate the parameters arising from spectral analysis, such as spectral modeling sinusoidal/noise tracks, or source/filter model parameters using LPC.

Related to SDIF, but on the control side, is open sound control (OSC, 2002) [17.41]. Open sound control allows for networked, extended-precision, object-oriented control of synthesis and processing algorithms over high-bandwidth transmission control protocol/internet protocol (TCP/IP), FireWire, and other high-bandwidth protocols.

As part of the moving-picture experts group 4 (MPEG4) standard, the structured audio orchestra language (SAOL, 1999) allows for parametric descriptions of sound algorithms, controls, effects processing, mixing, and essentially all layers of the audio synthesis and processing chain [17.42]. One fundamental notion of structured audio is that the more understanding we have about how a sound is made, the more flexible we are in compressing, manipulating, and controlling it.

Amazon
EX1013 Page 744

17.10 Composition

The history of computer languages for music dates back well into the 1950s. Beginning with Max Mathew's Music I, many features of these languages have remained constant, such as the notion of unit generators as discussed before, and the idea of instruments within an orchestra, controlled by a score. Early popular languages include Music V (a direct descendent of Music I) [17.5], Csound, CMix, and CMusic (written in C) [17.43]. More recent languages developed in the last decade include SuperCollider [17.44], Jsyn (in Java) [17.45], STK in C++ [17.46], Aura [17.47], Nyquist [17.48], and ChucK [17.49].

One class of languages that has advanced education and participation in computer music are graphical systems such as MAX/MSP [17.50] and Pure Data (PD) [17.51]. These software systems allow novice (and expert) users to construct patches by dragging unit generator icons around on the desktop, connecting them with virtual patch cords (Fig. 17.47).

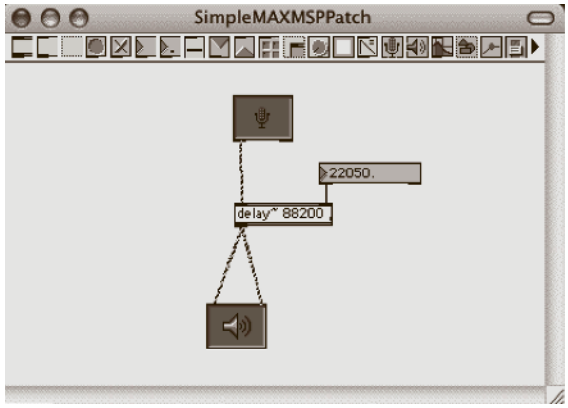


Fig. 17.47 A simple MAX/MSP patch connects a microphone through a 1/2 s (22050 samples at 44.1 k sample rate) delay to the speaker output

17.11 Controllers and Performance Systems

Music performance is structured real-time control of sound, to achieve an aesthetic goal (hopefully for both the performers and audience). So one obvious application area for real-time computer sound synthesis and processing is in creating new forms of musical expression. Exciting developments in the last few years have emerged in the area of new controllers and systems for using parametric sound in real-time music performance, or in interactive installation art.

Figure 17.48 shows three recent real-time music performance systems, all using parametric digital sound synthesis and a variety of sensors mounted on the instruments and players. The author's DigitalDoo is a sensor/speaker enhanced/augmented didgeridoo. Dan Trueman's bowed sensor speaker array (BoSSA) resulted from work in modeling, refining, and redefining the violin player's interface [17.52]. Curtis Bahn's 'r!g' consists of a sensor enhanced upright (stick) bass, pedal boards, and other controllers. All three of these new instruments use spherical speakers to allow flexible control of the sound radiation into the performance space [17.53]. These instruments have been used with computers to control percussion sounds, animal noises, string models, flute models, and a large va-



Fig. 17.48 Cook's DigitalDoo, Trueman's BoSSA, Bahn's 'r!g'

Amazon
EX1013 Page 745



Fig. 17.49 Gigapop networked concert between Princeton (shown on screen to *left*) and Montreal (players on stage). The right screen shows real-time graphics generated from gestural data captured from sensors on the instruments

riety of other sample-based and parametric synthesis algorithms.

Various conducting systems, including Max Mathew's radio baton (Stanford's Center for Computer Research in Music and Acoustics) [17.54], and Theresa Marrin Nakra's conducting jacket (MIT Media Lab) [17.55] allow novices and experts to conduct music performed by the computer. Also, the growth and development of computer music has helped to advance greatly the technological capabilities of the familiar karaoke industry. Some modern karaoke systems now can deduce the song that is being sung and retrieve the appropriate accompaniment within just a few notes. Also, these systems can automatically correct the pitch and vocal qualities of a bad singer, making them sound more expert and professional. The next section will describe some of the algorithms being applied now in karaoke and other automatic accompaniment systems.

Some researchers and performers feel that there is significant potential in networked computer music performances and interactions. Projects range from MIT's Brain Opera [17.56], which allowed observers to participate and contribute content to the performance via the web, through the Auricle website, which is a kind of audio analysis/synthesis enhanced chat room [17.57]. Stanford's SoundWire Project [17.58] and Princeton's GigaPop Project [17.59] (Fig. 17.49) are aimed at synchronized multi-site live performance through high-bandwidth low-latency streaming of uncompressed audio and video.

17.12 Music Understanding and Modeling by Computer

Over the years, new research and applications have arisen in the areas of content-based audio analysis, retrieval, and modeling. Also called music information retrieval (MIR), machine listening, and other terms, the main tasks involve segmenting, separating, classifying, and modeling the textures and structures present in pieces of music or audio.

Audio segmentation is the process of breaking up an audio stream into sections that are perceptually different from adjacent sections. The audio texture within a given segment is relatively stable. Examples of segment boundaries could be a transition from background sound texture to a human beginning to speak over that background. Another segment boundary might occur when the scene changes, such as leaving an office building lobby and going outside onto a busy street. In movie production, the audio scene often changes just prior to the visual scene, allowing a noncausal mode of detecting a visual scene change.

Segmentation [17.60] can be accomplished in two primary ways: blind segmentation based on sudden changes in extracted audio features, and classification-based segmentation based on comparing audio features to a set of trained target feature values. The blind method works well and is preferred when the segment textures are varied and unpredictable, but requires the setting of thresholds to yield the best results. The classification-based method works well on a corpus of pre-labeled data, such as speech/music, musical genres, indoor/outdoor scenes, etc. databases. Both methods require the extraction of audio features.

Audio feature extraction is the process of computing a compact numerical representation of a sound segment. A variety of audio features exist, and have been used in systems for speech recognition, music/speech discrimination, musical genre (rock, pop, country, classical, jazz, etc.) labeling, and other audio classification tasks. Most features are extracted from short moving windows (5–100 ms in length, moving along at a rate of 5–20 windows per second) by using the short-time Fourier transform, wavelets, or compressed data coefficients (such as MP3). Each feature described below can be computed at different time resolutions, and the value of each feature, along with the mean

Amazon
EX1013 Page 746

and variance of the features can be used as features themselves.

- One common feature for segmentation is the gross power in each window. If the audio stream suddenly gets louder or softer, then there is a high likelihood that something different is occurring. In speech recognition and some other tasks, however, we would like the classification to be loudness invariant (over some threshold used to determine if anyone is speaking).
- Another feature is the spectral centroid, which is closely related to the brightness of the sound, or the relative amounts of high- and low-frequency energy. The spectral roll-off is another important feature that captures more information about the brightness of an audio signal.
- Spectral flux is the amount of frame-to-frame variance in the spectral shape. A steady sound or texture will exhibit little spectral flux, while a modulating sound will exhibit more flux.
- One popular set of features used for speech and speaker recognition are mel-frequency cepstral coefficients, which are a compact (between 4 and 10 numbers) representation of spectral shape.
- For multi-time-scale analysis systems, the low-energy feature is often used. This feature is defined as the percentage of small analysis windows that contain less power than the average over the larger window that includes the smaller windows. This is a coarser-time-scale version of flux, but computed only for energy.
- One other commonly used feature is zero-crossing rate, which is a simple measure of high-frequency energy.
- For music query/recognition, features such as the parametric pitch histogram, and beat/periodicity histogram can be calculated and used.

Selection of the correct feature set for a given task has proven to be an important part of building successful systems for machine audio understanding. For a fixed corpus, computing many features (40 dimensions or more), then using principal-components analysis (reduction of dimensionality through regression) has proven successful for reducing the dimensionality of the feature/search space. As an example, mapping the first three principal components onto color (red, green, blue) allows the automatic coloration of sound plots, as shown in Fig. 17.50. The three *timbregrams* on the left are speech, and the three on the right are classical music. The top two right timbergrams are orchestral recordings, while the lower right one is opera singing (hence the color similarity to speech).

One holy grail in computer music research is automatic transcription, where in the ideal case a computer could generate a perfect musical score by analyzing only a recording of the music. This problem has been around for many decades, and has generated significant funding for computer music researchers. While research still continues on this difficult and unsolved problem, the side effects of the research have given us a large number of developments in signal

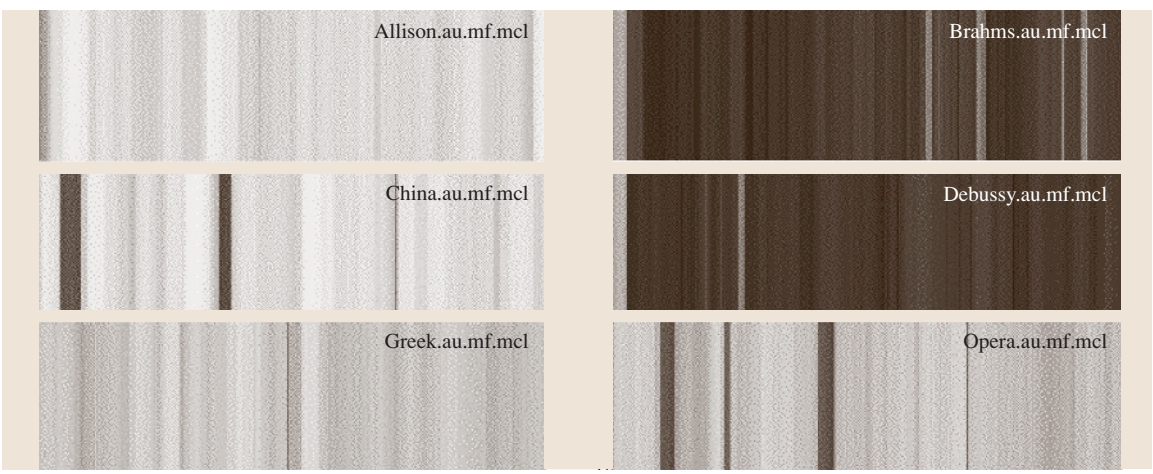


Fig. 17.50 *Timbregrams* of speech (*left*) and music (*right*). The top two sounds on the right are orchestral music, while the lower right is opera

Amazon
EX1013 Page 747

processing, analysis/synthesis algorithms, and human perception and cognition.

One other cognition-related research area in computer music is musical style modeling, also called automatic composition. One of the leading researchers in this area is David Cope, whose experiments in musical intelligence (EMI) projects have been developed over many years now. Through musical rules and data entered by Cope in the LISP computer language, the simple analytic recombinant algorithm (SARA) program has generated imitative music in styles as varied as those of Palestrina, Bach, Brahms, Chopin, Mozart, Prokofiev, Stravinsky, Gershwin, and Scott Joplin [17.61].

Another interesting project in this area is The Brain, created by Tom Hadju and Andy Milburn of the production company "tomandandy". The Brain is partially responsible for a large amount of commercial and movie music created over the last few years, including music for *The Mothman Prophesies*, and *Mean Creek*.

17.13 Conclusions, and the Future

With the advent of cheap and powerful computers, especially laptops, the last few years have greatly democratized the tools and techniques of computer music. Many free or inexpensive sound-editing tools now include features and algorithms that were the sole domain of academic researchers only ten years ago. Further, the World Wide Web has made computer tools and music available to more people. Computer-augmented performers and performance spaces have become more common, and more commonly accepted by the public. It seems certain that the future of computer music is likely to include more developments in the areas of modeling and understanding of human perception, preference, and creativity.

References

17.1 K. Pohlmann: *Principles of Digital Audio* (McGraw Hill, New York 2000)

17.2 J.O. Smith, P. Gossett: A flexible sampling-rate conversion method. In: *Proc. Int. Conf. Acoustics Speech and Signal Processing*, Vol. 2 (IEEE, Piscataway 1984) pp. 19.4.1–19.4.2

17.3 T. Dutoit: *An Introduction to Text-To-Speech Synthesis* (Kluwer Academic, Dordrecht 1996)

17.4 R. Bristow-Johnson: *Wavetable synthesis 101: a fundamental perspective AES 101 Conference* (AES, New York 1996), Music-DSPSource Archive, (http://www.musicdsp.org)

17.5 M.V. Mathews: *The Technology of Computer Music* (MIT Press, Cambridge 1969)

17.6 R. Bracewell: *The Fourier Transform and It's Applications* (McGraw Hill, New York 1999)

17.7 X. Serra, J.O. Smith: Spectral modeling synthesis: Sound analysis/synthesis based on a deterministic plus stochastic decomposition, Comp. Music J. **14**(4), 1–2 (1990)

17.8 R.J. McAulay, T. Quatieri: Speech analysis/synthesis based on a sinusoidal representation, IEEE Trans. ASSP **34**, 744–754 (1986)

17.9 T. Verma, T. Meng: *An analysis/synthesis tool for transient signals that allows a flexible sines+transients+noise model for audio 1998 IEEE ICASSP-98* (IEEE, Piscataway 1998)

17.10 A. Moorer: The use of the phase vocoder in computer music applications, J. Audio Eng. Soc. **26**(1/2), 42–45 (1978)

17.11 M. Dolson: The phase vocoder: A tutorial, CMJ **10**(4), 14–27 (1986)

17.12 P. Cook: *Real Sound Synthesis for Interactive Applications* (AK Peters, Wellesley 2002)

17.13 P. Morse: *Vibration and Sound* (for the Acoustical Society of America, New York 1986)

17.14 K. Steiglitz: *A Digital Signal Processing Primer* (Addison Wesley, Boston 1996)

17.15 J.M. Adrien: The missing link: Modal synthesis Chap. 8. In: *Representations of Musical Signals*, ed. by De Poli, Piccialli, Roads (MIT Press, Cambridge 1991) pp. 269–297,)

17.16 H. Dudley: *The Vocoder* (Bell Laboratories Record, Mirray Hill 1939), Reprinted in IEEE Trans. Acoust. Speech Signal Process. ASSP-29(3):347-351 (1981)

17.17 B. Atal: Speech analysis and synthesis by linear prediction of the speech wave, J. Acoust. Soc. Am. **A47**(65), 65 (1970)

17.18 A. Moorer: The use of linear prediction of speech in computer music applications, J. Audio Eng. Soc. **27**(3), 134–140 (1979)

17.19 D. Klatt: Software for a cascade/parallel formant synthesizer, J. Acoust. Soc. Am. **67**(3), 971–995 (1980)

Amazon
EX1013 Page 748

17.20 J.L. Kelly Jr., C.C. Lochbaum: *Speech Synthesis, in Proceedings of the Fourth ICA* (Paper G42, Copenhagen 1962)

17.21 J.M. Chowning: The synthesis of complex audio spectra by means of frequency modulation, J. Audio Eng. Soc. **21**(7), 526–534 (1973)

17.22 J. Chowning: The synthesis of complex audio spectra by means of frequency modulation, Comp. Music J. **1**(2) (1977), reprinted

17.23 J. Chowning: The synthesis of complex audio spectra by means of frequency modulation. In: *Foundations of Computer Music*, ed. by C. Roads, J. Strawn (MIT Press, Cambridge 1985), reprinted

17.24 J. Chowning: Frequency modulation synthesis of the singing voice. In: *Current Directions in Computer Music Research*, ed. by M.V. Mathews, J.R. Pierce (MIT Press, Cambridge 1980)

17.25 X. Rodet: Time-domain formant-wave-function synthesis, Comp. Music J. **8**(3), 9–14 (1984)

17.26 I. Daubechies: Orthonormal bases of compactly supported wavelets, Commun. Pure Appl. Math. **41**, 909–996 (1988)

17.27 C. Roads: Asynchronous granular synthesis. In: *Representations of Musical Signals*, ed. by G. De Poli, A. Piccialli, C. Roads (MIT Press, Cambridge 1991) pp. 143–185

17.28 P. Cook: *Physically Informed Sonic Modeling (PhISM): Percussion Instruments, in Proceedings of the ICMC* (International Computer Music Association, San Francisco 1996)

17.29 P. Cook: Physically informed sonic modeling (PhISM): Synthesis of percussive sounds, Comp. Music J. **21**, 3 (1997)

17.30 C. Cadoz, A. Luciani, J.L. Florens: CORDIS-ANIMA: A modeling and simulation system for sound image synthesis. The general formalization , Comp. Music J. **17**(1), 23–29 (1993)

17.31 K. Karplus, Strong: Digital synthesis of plucked-string and drum timbres, Comp. Music J. **7**(2), 43–55 (1983)

17.32 J.O. Smith: Acoustic modeling using digital waveguides. In: *Musical Signal Processing*, ed. by Roads et.al. (Swets, Zeitlinger, Netherlands 1997)

17.33 D.A. Jaffe, Smith: Extensions of the Karplus-Strong plucked string algorithm, Comp. Music J. **7**(2), 56–69 (1983)

17.34 P.R. Cook, D. Trueman: Spherical radiation from stringed instruments: Measured, modeled, and reproduced, J. Catgut Acoust. Soc. **1**, 1 (1999)

17.35 J.O. Smith, S.A. Van Duyne: Overview of the commuted piano synthesis technique. In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, ed. by E. Wenzel (IEEE, New York 1995)

17.36 M.E. McIntyre, R.T. Schumacher, J. Woodhouse: On the oscillations of musical instruments, J. Acoust. Soc. Am. **74**(5), 1325–1345 (1983)

17.37 J.O. Smith: Efficient simulation of the reed-bore and bow-string mechanisms. In: *Proceedings of the International Computer Music Conference* (International Computer Music Association, San Francisco 1986) pp. 275–280

17.38 G. Essl, P. Cook: *Banded waveguides: Towards physical modeling of bar percussion instruments, in Proc. 1999 Int. Computer Music Conf., Beijing, China* (International Computer Music Association, San Francisco 1999)

17.39 MMA: *The Complete MIDI 1.0 Detailed Specification* (MIDI Manufacturers Association, La Habra 1996)

17.40 S. Jungleib: *General MIDI, A-R Editions* (A-R Editions, Middleton 1995)

17.41 M. Wright, A. Chaudhary, A. Freed, D. Wessel, X. Rodet, D. Virolle, R. Woehrmann, X. Serra: *New Applications for the Sound Description Interchange Format, Proc. Of the ICMC 1998* (International Computer Music Association, San Francisco 1998)

17.42 M. Wright, A. Freed: Open sound control: A new protocol for communicating with sound synthesizers. In: *Proceedings of the International Computer Music Conference, 1997* (International Computer Music Association, San Francisco 1997),

17.43 S.T. Pope: Machine Tongues XV: Three packages for software sound synthesis, Comp. Music J. **17**(2), 23–54 (1993), Note: these are cmix, csound, cmusic

17.44 J. McCartney: SuperCollider: A new real-time synthesis language. In: *Proc. Int. Computer Music Conference* (International Computer Music Association, San Francisco 1996) pp. 257–258

17.45 P. Burk: JSyn – A real-time synthesis API for java. In: *Proc. Int. Computer Music Conference* (International Computer Music Association, San Francisco 1998) pp. 252–255

17.46 Cook, G. Scavone: The synthesis toolkit (STK). In: *Proceedings of the International Computer Music Conference* (International Computer Music Association, San Francisco 1999) pp. 164–166

17.47 R.B. Dannenberg, E. Brandt: A flexible realtime software synthesis system. In: *Proceedings of the International Computer Music Conference* (International Computer Music Association, San Francisco 1996) pp. 270–273

17.48 R.B Dannenberg: Machine Tongues XIX: Nyquist, a language for composition and sound synthesis, Comp. Music J. **21**(3), 50–60 (1997)

17.49 G. Wang, P.R. Cook: ChucK: A concurrent. In: *On-the-fly Audio Programming Language, in Proceedings of International Computer Music Conference* (International Computer Music Association, San Francisco 2003) pp. 219–226

17.50 M. Puckette: Combining event and signal processing in the MAX graphical programming environment, Comp. Music J. **15**(3), 68–77 (1991)

Amazon
EX1013 Page 749

17.51 M. Puckett: Pure data. In: *Proceedings of International Computer Music Conference* (International Computer Music Association, San Francisco 1996) pp. 269–272
17.52 D. Trueman, P. Cook: BoSSA: The deconstructed violin reconstructed, J. New Music Res. **29**(2), 120–130 (2000)
17.53 D. Trueman, C. Bahn, P. Cook: Alternative voices for electronic sound. In: *Spherical speakers and sensor-speaker arrays (SenSAs), International Computer Music Conference, Berlin, Aug. 2000* (International Computer Music Association, San Francisco 2000)
17.54 M.V. Mathews: The radio baton and conductor program, or: Pitch, the most important and least expressive part of music, Comp. Music J. **15**(4), 37–46 (1991)
17.55 T. Marin Nakra: *Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture*, Ph.D. Thesis (MIT Media Lab, Cambridge 2000)
17.56 J.A. Paradiso: The brain opera technology: New instruments and gestural sensors for musical interaction and performance, J. New Music Res. **28**, 2 (1999)
17.57 J. Freeman, S. Ramakrishnan, K. Varnik, M. Neuhaus, P. Burk, D. Birchfeld: *Adaptive High-level Classification of Vocal Gestures Within a Networked Sound Instrument, Proc. Of the International Computer Music Conference* (International Computer Music Association, San Francisco 2004)
17.58 C. Chafe, S. Wilson, R. Leistikow, D. Chisholm, G. Scavone: A simplified approach to high quality music and sound over IP. In: *Proceedings of the Conference on Digital Audio Effects (DAFX)*, ed. by D. Rocchesso (University of Verona, Italy 2000)
17.59 A. Kapur, G. Wang, P. Davidson, P.R. Cook, D. Trueman, T.H. Park, M. Bhargava: *The Gigapop Ritual: A Live Networked Performance Piece for Two Electronic Dholaks, Digital Spoon, DigitalDoo, 6 String Electric Violin, Rbow, Sitar, Tabla, and Bass Guitar, New Interfaces for Musical Expression (NIME)* (McGill University, Montreal 2003)
17.60 G. Tzanetakis, P. Cook: Multi-feature audio segmentation for browsing and annotation. In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (IEEE, Piscataway 1999)
17.61 D. Cope: *Experiments in Musical Intelligence*, Vol. 12 (A-R Editions, Madison 1996)

Amazon
EX1013 Page 750

18. Audio and Electroacoustics

This chapter surveys devices and systems associated with audio and electroacoustics: the acquisition, transmission, storage, and reproduction of audio. The chapter provides an historical overview of the field since before the days of Edison and Bell to the present day, and analyzes performance of audio transducers, components and systems from basic psychoacoustic principles, to arrive at an assessment of the perceptual performance of such elements and an indication of possible directions for future progress.

The first, introductory section is an overall historical review of audio reproduction and spatial audio to establish the context of events. The next section surveys relevant psychoacoustic principles, including performance related to frequency response, amplitude, timing, and spatial acuity. Section 3 examines common audio specifications, with reference to the psychoacoustic limitations discussed in Sect. 2. The specifications include frequency and phase response, distortion, noise, dynamic range and speed accuracy. Section 4 examines some of the common audio components in light of the psychoacoustics and specifications established in the preceding sections. The components in question include microphones, loudspeakers, record players, amplifiers, magnetic recorders, radio, and optical media. Section 5 is concerned with digital audio, including the basics of sampling, digital signal processing, and audio coding. Section 6 is devoted to an examination of complete audio systems and their ability to reproduce an arbitrary acoustic environment. The specific systems include monaural, stereo, binaural, Ambisonics, and 5.1-channel surround sound. The final section provides an overall appraisal of the current state of audio and electroacoustics, and speculates on possible future directions for research and development.

18.1 **Historical Review** 744
18.1.1 Spatial Audio History 746

18.2 **The Psychoacoustics of Audio and Electroacoustics** 747
18.2.1 Frequency Response 747
18.2.2 Amplitude (Loudness) 748
18.2.3 Timing 749
18.2.4 Spatial Acuity 750

18.3 **Audio Specifications** 751
18.3.1 Bandwidth 752
18.3.2 Amplitude Response Variation 753
18.3.3 Phase Response 753
18.3.4 Harmonic Distortion 754
18.3.5 Intermodulation Distortion 755
18.3.6 Speed Accuracy 755
18.3.7 Noise 756
18.3.8 Dynamic Range 756

18.4 **Audio Components** 757
18.4.1 Microphones 757
18.4.2 Records and Phonograph Cartridges 761
18.4.3 Loudspeakers 763
18.4.4 Amplifiers 766
18.4.5 Magnetic and Optical Media 767
18.4.6 Radio 768

18.5 **Digital Audio** 768
18.5.1 Digital Signal Processing 770
18.5.2 Audio Coding 771

18.6 **Complete Audio Systems** 775
18.6.1 Monaural 776
18.6.2 Stereo 776
18.6.3 Binaural 777
18.6.4 Ambisonics 777
18.6.5 5.1-Channel Surround 777

18.7 **Appraisal and Speculation** 778

References 778

Considering that human records extend back to cave drawings thousands of years old [18.1], it is notable that the technology to record, transmit, and reproduce sound has only existed for the last century or so. One can only

Amazon
EX1013 Page 751

wonder at the effect it might have had on the course of history, not to mention the effect it may yet have [18.2].

Sound is hardly the easiest quantity to work with. As a vibration in air, it is invisible, dynamic, and three-dimensional (3-D), not counting the dimensions of time and frequency. Human interaction with a sound field is often dynamic: natural movements of the head and body alter the perceived sounds in a way that is difficult to predict or reproduce. Despite over a century of progress, the transducers available for recording and reproducing sound, microphones and loudspeakers, are still limited to being point-in-space devices, and are not necessarily well suited to interacting with a 3-D physical phenomenon. And yet, human beings make do with just two ears; so how hard can it be to capture and reproduce all the characteristics and nuances of what the ear perceives in nature?

The intent of this chapter is to review and evaluate the performance of audio systems based on the known performance of the human auditory system. Starting with a brief historical overview, the salient aspects of human auditory performance will be described. These in turn will be used to derive a set of general audio specifications that may be expected to apply to a broad range of audio systems. Then, some of the more common audio components will be examined in light of these specifications. Finally, the performance of complete systems will be considered, in the hope of gaining some overall perspective of progress to date and possibly fertile directions for further investigation.

18.1 Historical Review

The first golden age of audio research arguably started in the late 1800s with the inventions of the telephone by A. G. Bell and the phonograph by T. A. Edison [18.3], and ended in 1982 with the introduction of digital audio as a practical reality as embodied in the audio compact disc (CD). The inventions that inaugurated this age established that an audio channel could indeed be transmitted and recorded, as a continuous mechanical, magnetic, optical, or electrical analog of the sound vibrations, but brought with them a raft of problems, including inadequate recording sensitivity and playback volume, noise, distortion, speed irregularities such as wow and flutter, nonlinear and limited frequency response, and media wear. Although steady incremental progress was made on these problems in the decades that followed, they were all largely resolved once and for all with digital audio. Efforts since then have focused on reduction of the required data rate (audio coding) and on extending the realm of audio from single isolated channels to full three-dimensional systems. Ongoing parallel efforts to understand the physiology and psychoacoustics of the human ear more fully have continued to extend the knowledge base, and in some cases have yet to be exploited by commercial audio systems.

Why wasn't sound recording invented sooner? It is hard to say. Bell's telephone may have required a steady electric current, which presumably was not available in ancient times, but Edison's phonograph was entirely mechanical. The key to the latter invention seems to have been the recognition that sound is a vibration and that the vibration is mechanically transferable, but this notion that is readily demonstrable by putting your fingers to your lips or throat, and humming. The inventions seem to have been part of a general flourishing of physical sciences and mathematics, one that has continued to the present day. Prior to that, however, there is certainly evidence of some understanding of the acoustic processes involved, as seen in the creation of musical instruments and the design of concert halls.

Actually, both Edison and Bell were driven in part by a desire to improve telegraph systems. The telegraph was popularized by Samuel Morse, starting around 1837, although the notion of an electrical telegraph harks back at least to a demonstration of sending a simple signal over a long wire by William Watson in 1747 [18.4]. The cause was greatly helped by the invention of the electric battery by Alessandro Volta in 1799, and the storage battery in 1859 by Gaston Plante [18.5].

Many of the theoretical and mathematical foundations of early audio technologies were established in the 1800s, with work in the electrical sciences by Ohm, Faraday, Henry and others, together with research in acoustics and hearing physiology by Helmholtz, Lissajous and others. Although Edison is credited with the invention of the first device to reproduce sound, the first audio recording device is generally credited to Leon Scott, who on March 25, 1857, patented the phonautograph [18.6]. This device combined a diaphragm and a stylus to inscribe a trace of a sound on a visual medium, such as a smoke-blackened glass or cylinder, but it had no way to reproduce the recording audibly. John Tyndall seems to have followed Scott's work on visualizing sound vibrations, using a variety of different devices.

Amazon
EX1013 Page 752

Alexander Graham Bell filed his patent for the telephone on February 14, 1876, just two hours before Elisha Gray filed for a similar device. Aside from its intrinsic value, the invention was significant for employing acoustical to electrical and electrical to acoustical transduction, principles that are still almost universally employed in telephony, broadcast, and audio recording.

Transduction between acoustical and electrical energy was initially less important to Edison's invention of the phonograph on December 8, 1877, since a physical medium to store the recording was necessary in any case, and there neither was nor is a ready way to directly store a dynamic electrical signal directly on such a physical medium. So it is perhaps not too surprising that the original invention employed purely mechanical transduction for both recording and playback.

With these two inventions, the foundations for analog audio were established, and the race was on to refine and commercialize the systems. In 1887, Emile Berliner patented a recording format that used lateral groove modulation on a flat disk [18.8], in contrast to Edison's use of vertical modulation on a cylinder. The underlying principles may have been much the same, but Berliner's arrangement qualified as a separate patent, and the practical aspects of mass-producing, playing, and storing disks instead of cylinders led to the grooved disk becoming the commercial recording medium of choice for over half a century.

Lacking the means to record an electrical signal directly, but seeking to avoid the wear accompanying the playback of grooved recordings with a stylus, other media were soon being explored, such as the use of magnetic recording by Valdemar Poulson in 1898 and optical recording, by Leon Gaumont in 1901. Meanwhile, the eventual foundation of the broadcast industry was begun in 1894, with the invention of the wireless telegraph by Guglielmo Marconi.

These systems were initially handicapped by a lack of viable amplification means, and although there were some imaginative attempts at mechanical amplification, the key breakthrough occurred in 1906, with Lee de Forest's invention of the *audion* vacuum-tube triode amplifier. Vacuum tubes could and would be used to amplify all manner of electrical signals, but de Forest's choice of the name *audion* seems derived directly from the word audio. It is therefore a matter of some speculation as to whether the tube amplifier would have come about when it did had it not been for the formative invention of sound recording just 29 years earlier.

Aside from resolving the problems of providing better recording sensitivity and adequate playback volume,

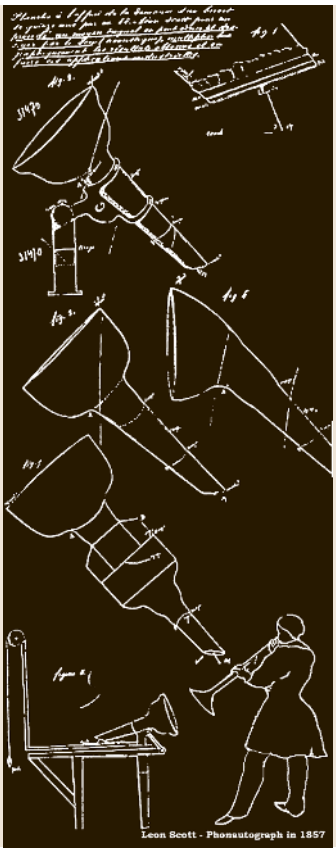


Fig. 18.1 Leon Scott and the phonautograph (After [18.7])

the vacuum tube also enabled the wireless transmission of audio, giving rise to the radio receiver and broadcast industries.

So by 1910, the basic principles, physics, and devices of audio and electroacoustics were in place, resulting in widespread continuing efforts to develop, patent, and refine components and systems, establish standards where required, and commercialize a growing variety of formats and services. Amidst competitive struggles and numerous patent fights, the following years saw a dramatic series of audio innovations, including [18.9, 10]:

- 1913: An early *talking* movie by Edison, using a cylinder synchronized with the picture.
- 1916: A superheterodyne radio circuit patented by Edwin F. Armstrong.
- 1917: A condenser microphone developed by E. C. Wente at Bell Laboratories.
- 1921: The first commercial amplitude-modulated (AM) broadcast, KDKA, Pittsburgh, PA.

Amazon
EX1013 Page 753

- 1925: An electrical record cutter developed at Bell Laboratories, followed by the commercial release of 78 revolutions per minute (RPM) records.
- 1926: Iron-oxide-coated paper tape for magnetic recoding developed by O'Neill.
- 1927: *The Jazz Singer*, the first commercial sync-sound film, using audio on disk synchronized to the picture.
- 1929: The sampling theorem, the basis of digital signal processing, developed by Harry Nyquist.
- 1935: Plastic-based magnetic recording tape, from BASF.
- 1936: The first tape recording of a symphony concert, by Sir Thomas Beecham.
- 1936: A cardioid condenser microphone, by Von Braunmuhl and Weber.
- 1941: Commercial frequency-modulated (FM) broadcasting inaugurated in the US.
- 1947: The Williamson high-fidelity power amplifier developed.
- 1948: The 33 1/3-RPM long-play record, developed by CBS laboratories.
- 1951: The transistor developed at Bell Laboratories.
- 1954: The transistor radio, Sony.
- 1963: Tape cassette format developed by Philips.
- 1965: Dolby *A* noise-reduction system developed.
- 1969: Dr. Thomas Stockham experiments with digital recording.
- 1976: Stockham produces the first 16-bit digital recording, of the Santa Fe Opera.
- 1982: The digital compact disc is released commercially.

Much of the thrust of these and countless other developments in this period was to improve the quality of a given audio channel, by extending the frequency response, reducing noise, etc. As noted earlier, once digital recording came of age, these concerns were rendered somewhat moot, as adequate quality became a matter of using a sufficient number of bits per sample and adequately sampling rate to avoid audible degradation. Of course, certain parts of the audio chain, notably transducers such as microphones and loudspeakers, and associated interface electronics, have remained analog, as is likely to continue to be true at least until a direct link is established to the human nervous system, bypassing the ears altogether.

18.1.1 Spatial Audio History

Perhaps the one area of audio that was not immediately improved with the adoption of digital recording techniques was spatial reproduction. It is unlikely that the importance of the spatial aspect of audio was lost on early audio developers, nor the fact that most humans do in fact, possess two ears. It was probably just a case of priorities. Spatial audio systems would almost certainly require more channels and/or other information, which was hard to accommodate with early audio media, and the emphasis was, quite reasonably, on refining the quality of the single-channel monophonic format.

Still, there were some promising early experiments, notably a demonstration in 1881 by Clement Ader, who set up a series of microphones in front of the stage of the Paris Opera, and ran their outputs to earphones set up in a nearby room [18.9, 10]. Whether by accident or design, some of the listeners elected to use two headphones, one on each ear, and were rewarded with a crude but effective binaural spatial presentation.

Despite isolated experiments such as Ader's, there was little organized investigation of spatial audio until 1931, when Alan Blumlein single-handedly developed a range of devices and protocols to support stereo sound, including techniques to upgrade the entire audio broadcast and recording infrastructure of the time. Blumlein filed a far-ranging patent covering the bulk of his work; unfortunately, it seems to have been about a quarter century ahead of his time, as it would be the late 1950s before stereophonic sound became a commercial reality.

In the 1930s, Bell Laboratories also became interested in spatial audio, albeit with more of an emphasis on audience presentation. They concluded that stereo was ineffective for such uses, as the virtual imaging of just two channels would not work for listeners removed from the centerline. They further concluded that adding a center channel produced a far more satisfactory result, as demonstrated by Fletcher, Steinberg, and Snow in 1933.

With consumer media still limited to single-channel operation, it fell for a time to film studios to engage in occasional commercial forays into spatial audio. A notable experiment in the venue was the film *Fantasia* (1940), the road show productions of which used a synchronized three-channel soundtrack, plus control track, to feed as many as 10 loudspeakers [18.9, 10].

The use of multiple film channels accelerated somewhat in the 1950s, as film studios sought ways to differentiate their product from television. One popular approach was to use wider-than-standard 70 mm film stock, instead of the normal 35 mm, which provided enough room along the edge for several audio channels

Amazon
EX1013 Page 754

on magnetic stripes. A common arrangement was three front channels plus a surround channel, the latter usually fed to an array of loudspeakers arranged around the sides and rear of the theater.

In the 1970s, the blockbuster movie genre was kicked off with the releases of *Star Wars* and *Close Encounters of the Third Kind*, along with the synergistic Dolby matrix surround system, which encodes control signals in the audio to pan two optical film channels to four output channels (again, three in front, plus surround).

In the 1990s, multichannel digital soundtracks came into use, first in movie theaters and then in home systems, generally supporting three front channels, a pair of surround channels, and a low-frequency channel, otherwise referred to as a 5.1-channel system.

18.2 The Psychoacoustics of Audio and Electroacoustics

In order to evaluate the requirements and performance of audio systems, it is necessary to establish their relation to the capabilities and limitations of the human hearing. People perceive sound on the basis of frequency, amplitude, time and, somewhat indirectly, space. The perceptual resolution limits of these quantities are fairly well established, even if the underlying mechanisms are still being explored.

18.2.1 Frequency Response

1. Range: The commonly quoted frequency range of human hearing is 20–20 000 Hz, representing the range of frequencies that can be heard as sounds. In many cases, this is a somewhat optimistic specification, and in any case certainly varies with the individual. Only young people are likely to hear out to 20 kHz. With age generally comes a reduction in the upper limit (along with other reductions in hearing acuity, a process known as presbycusis [18.11]), with typical values of the upper limit of hearing in the range of 8–15 kHz common for middle-aged adults. There has been considerable debate about the ability to hear above 20 kHz, with some claiming some sort of perception out to frequencies of 40–50 kHz. Hearing limits above 20 kHz are known to exist in some small animals. At the low-frequency end of the spectrum, a frequency of perhaps 32 Hz is typically the minimum value at which a pitch can be discerned, below which the sensation becomes more of a series of discrete vibrations. Although it is difficult to construct loudspeakers with flat response below 20 Hz, humans can still perceive such vibrations if they are sufficiently large in amplitude.
2. Sensitivity: The ear is not uniformly sensitive to all frequencies, due in part to resonances and mechanical limitations of structures in the outer and middle ear, and in part to neural effects. The sensitivity also tends to be a function of absolute loudness (defined later). Contours of perceived equal loudness were first measured by Fletcher and Munson at Bell Laboratories in the 1930s [18.12]. The figure below shows typical equal loudness curves as a function of absolute level [dB sound pressure level (SPL)].
3. Resolution: The ear exhibits a curious duality with respect to the degree it can differentiate frequencies. On the one hand, certain frequency-dependent processes, such as masking, appear to indicate

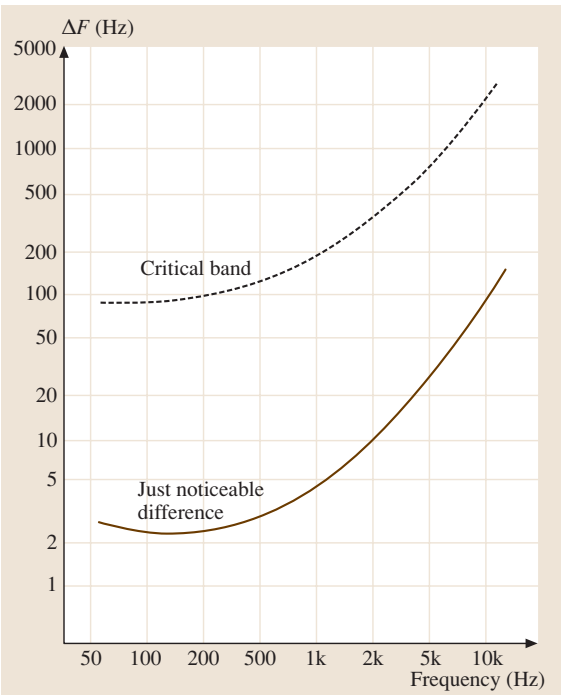


Fig. 18.2 Critical bandwidths (*upper curve*) and sine-wave pitch just noticeable difference (JND) as a function of frequency

Amazon
EX1013 Page 755

a grouping of the frequency range into somewhat coarse sub-bands about 1/4 to 1/3 octave wide, called *critical bands*. On the other hand, frequency discrimination of individual sine waves on the basis of their perceived pitch has a typical resolution limit on the order of 1/100th of an octave, corresponding to a frequency change of about 0.7%. This is one of the cases where the underlying mechanisms are uncertain, and a mixture of temporal and spectral discrimination appears to be necessary to explain the various observed phenomena.

4. Pitch: In general terms, pitch is the perceptual response to differences in the frequency content of a source, but it is far more complex than a simple mapping of frequency to a perceptual scale, and can be affected by parameters such as the loudness of the signal, which ear it is played to, and the timing of the signal. The perception of pitch of spectrally complex signals is especially difficult to model, in part because a given signal may be perceived as having more than one pitch, and the signal itself may or may not actually contain spectral components corresponding to the perceived pitch [18.13].

18.2.2 Amplitude (Loudness)

Absolute Sensitivity

The total dynamic range of the human ear, that is, the range from loudest to softest sound one can perceive, is generally given to be about 120 dB a pressure range of about a million to one, or a power range of about a trillion to one. The term *sound pressure level* (SPL) is used to indicate where in this range a given sound falls, with 0 dB SPL taken as the approximate threshold of hearing at 1 kHz, corresponding to a level of about 20 μPa [18.14]. The mathematical relation between a sound pressure p and corresponding SPL value is

$$L_\mathrm{p} = \mathrm{SPL} = 20\log(p/p_0)\,,$$

where p_0 corresponds to the 0 dB SPL quoted above.

The upper limit of loudness perception is not well defined, but corresponds in principle to the onset of nonaudible phenomena such as tingling and outright pain. Depending on the listener, the upper limit is likely to be in the range of 115–140 dB SPL. (Note that exposure to such levels, especially for extended periods, can lead to hearing loss [18.11]).

As noted, sensitivity is a function of frequency and absolute level, with the greatest sensitivity observed in the range 3500–4000 Hz, mainly due to resonance of the ear canals, where the hearing threshold may actually be slightly lower in level than 0 dB SPL. The greatest variation as a function of frequency is at the lowest levels, where the threshold of hearing may rise to in excess of 20 dB SPL at high frequencies, and fully 60 dB SPL at low frequencies.

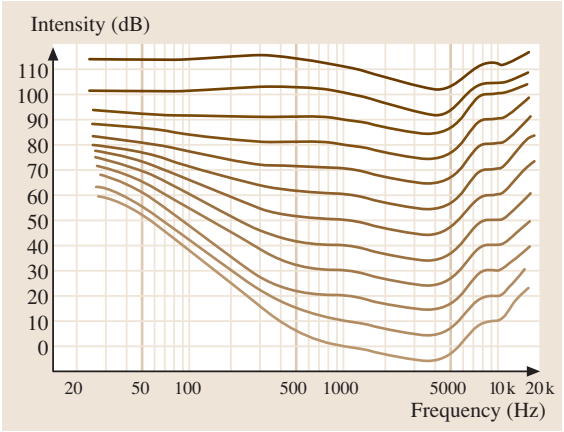


Fig. 18.3 Equal-loudness curves as a function of frequency and absolute level (After [18.9])

The variation of perceived loudness with actual sound pressure, while monotonic, is certainly nonlinear, as can be seen from the differences in spacing between the equal loudness curves at different frequencies. The actual perception of a difference in loudness will generally be less than the actual difference in SPL. According to the Sone scale of loudness perception, a difference in SPL of 10 dB at 1 kHz is roughly equivalent to a doubling of perceived loudness Sones $= 2^{[(\mathrm{Phons}-40)/10]}$ [18.15].

Because of the variation in perceived loudness with the actual SPL and frequency of a signal, it is a nontrivial matter to calculate the perceived loudness of an arbitrary, spectrally complex signal. Recent efforts using multiband analysis and power summation have improved the predictability to within close to a loudness just noticeable difference (JND) for most signals [18.16].

Resolution

The amplitude resolution of the ear is generally taken to be about 0.25 dB under best-case conditions, although for some situations it is considered to be slightly larger, on the order of 0.5–1.0 dB.

Masking

The audibility of a sine wave in isolation may be modified in the presence of other spectral components. In particular, the addition of spectral components is likely

Amazon
EX1013 Page 756

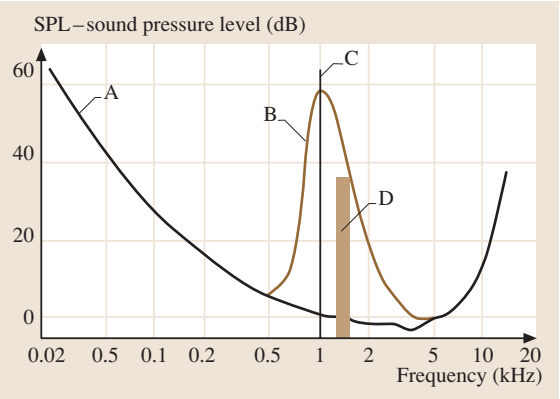


Fig. 18.4 Frequency Domain Masking: Curve A is hearing threshold with no signal present. Curve B is revised threshold in presence of 1 kHz tone C. Signal D, rising above curve A but below curve B, will be audible by itself but masked by signal C

to render an existing component less audible, a process referred to as masking.

The basic notion of masking is that a louder sound will tend to render as less audible or inaudible a softer sound that would otherwise be audible in isolation. The degree of masking will tend to depend in part on the frequency spacing of the spectral components in question. This gives rise to the notion of a prototype masking curve, delineating the threshold of audibility in the presence of a sine wave of some frequency and amplitude, as a function of frequency, as shown in Fig. 18.4.

To at least first-order approximation, the masking effect of two or more sine waves can be calculated by summing their respective individual masking curves. Since, from Fourier theory, any signal can be decomposed into a sum of sinusoids, a composite masking curve from an arbitrary signal can be calculated as the sum of the masking curves of each spectral component. This constitutes a convolution (filtering operation) of a prototype masking curve with the signal spectrum. In practice, the prototype masking curve used for the convolution is likely to vary as a function of frequency, amplitude, or signal type (noise versus sine wave). Further, nonlinearities in the ear may result in significant deviations from the predicted curve at times.

Although the masking effect of a signal is, quite logically, maximized while the signal is present, some residual of the masking effect will extend beyond the termination of the signal, and even slightly before its onset. This process is referred to as temporal masking, and typical behavior is illustrated in Fig. 18.5) [18.18].

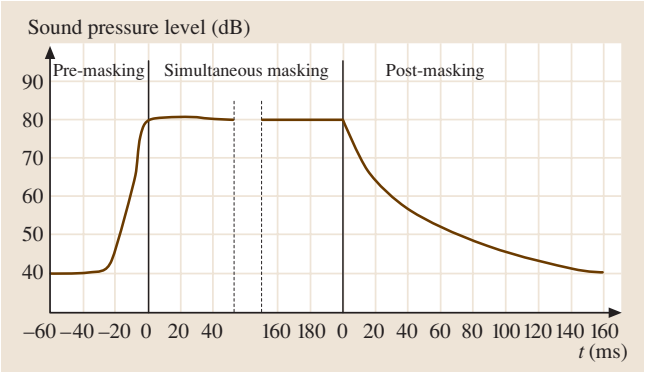


Fig. 18.5 Temporal masking characteristic

18.2.3 Timing

Because the transduction structure of the ear – the basilar membrane and associated hair cells – functionally resembles a filter bank, questions of timing must be considered as applying to the filtered output signals of that

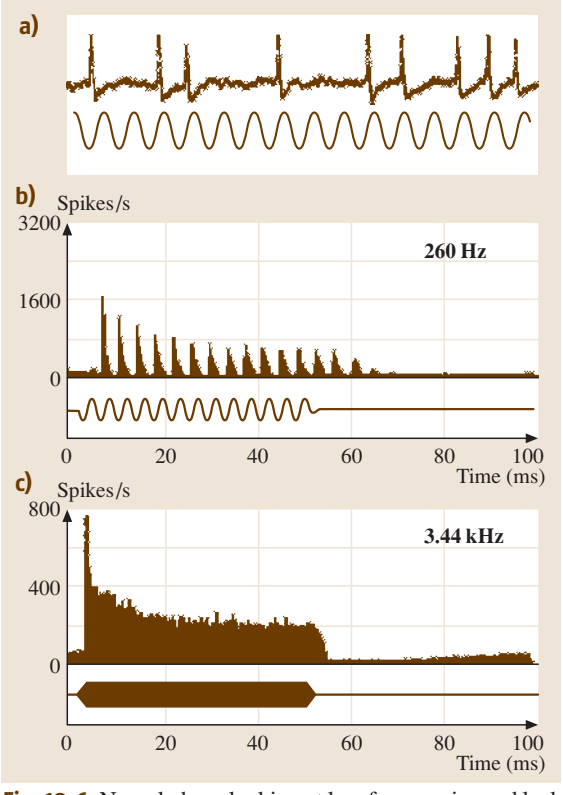


Fig. 18.6 Neural phase locking at low frequencies and lack of phase locking at higher frequencies [18.17]

Amazon
EX1013 Page 757

filter bank. The brain does not *see* the wide-band audio signal arriving at the ear, as it might appear on an oscilloscope, but instead processes the outputs of the basilar membrane filter bank, subject to effective half-wave rectification and loss of high-frequency synchrony by the neurons (Fig. 18.6). Output neural signals from each ear are ultimately processed by multiple areas of the brain. So, it is not surprising that issues of timing are a little ill-defined, and that there are a number of time constants associated with different aspects of hearing.

For starters, there is the basic ability of the neurons in the auditory nerve to follow the individual cycles of audio impinging on the basilar membrane; i.e. to exhibit phase locking. Current research seems to put the upper frequency of this activity at about 5 kHz, although this seems somewhat at odds with lateralization of sine waves on the basic of interaural time difference, which only extends up to about 1500 Hz.

The ability to follow individual cycles of audio may aid in pitch perception, as the effective shapes of the basilar membrane filters appear too broad to account for the observed pitch resolution. However, the physiological mechanisms for how this might be accomplished are still the subject of some debate.

Other time constants appear to apply to the aggregate audio, regardless of the presence of a filter bank at the input. Already noted are the masking time constants, which substantially extend only 1–2 ms prior to the onset of a loud sound, but continue for several tens of milliseconds after its cessation.

Each of these may be related to a more fundamental time constant. The short time constant of about 1–2 ms may represent the shortest time one can perceive without relying on spectral cues, while the post-masking interval may be related to the fusion time. The fusion time in particular seems to represent a kind of acoustic integration time, or the limit of primary acoustic memory, typically on the order of 30–50 ms, and appears to be associated with the lower frequency limit of the audio band (20–30 Hz). It also seems to explain why echoes are only heard as such in large cathedral-like rooms, as they are integrated with the direct arrivals in normal-size rooms.

Table 18.1 Summary of approximate perceptual time/amplitude limits

Parameter	Range	JND
Amplitude	120 dB	0.25 dB
Premask	N/A	2 ms
Postmask	N/A	30–50 ms
Binaural timing	700 µs	10 µs

On the other hand, binaural timing clearly exhibits higher resolution, as differences of interaural timing on the order of 10 ms may be audible.

18.2.4 Spatial Acuity

Strictly speaking, the spatiality of a sound field is not perceived directly, but is instead derived from analysis of the physical attributes already described. However, given its importance to audio and electroacoustics, it is nonetheless useful to review the processes involved.

Specification of the position of a sound source relative to a listener in three-space requires three coordinates. From the mechanisms used by the human ear, the natural selection is to use a combination of azimuth, elevation, and distance. Of these, distance is generally considered to be an inferred metric, based on the amplitude, spectral balance, and reverberation content of a source relative to other sources, and is not especially accurate in any absolute sense.

This leaves directional estimation as the primary localization task. It appears that the ear uses separate processes to determine direction in a left/right sense and a front/top/back sense. Given the physical placement of the ears, it is not surprising that the left/right decision is the more direct and accurate, depending on the interaural amplitude difference (IAD) and interaural time difference (ITD) localization cues.

Although the ear is sensitive to IAD at all audible frequencies, as can be verified by headphone listening, the head provides little shadowing effect at low frequencies, since the wavelengths become much longer than the size of the head, so IAD is mostly useful at middle to high frequencies.

The JND for IAD is about 1 dB, corresponding to about 1 degree of an arc for a front/centered high-frequency source. As a sound moves around to the side of the head, the absolute IAD will tend to increase, and the differential IAD per degree will decrease, causing the angular JND to decrease accordingly.

The ITD lateralization cue has a rather astonishing JND of just 10 ms, for a source on the centerline between the ears, corresponding to an angle of about 1 degree. For signals off to one side, the ITD will typically reach a maximum value of just 750 ms – 3/4 of a millisecond. As with the IAD cue, the sensitivity of the ITD cue declines as a source moves around to the side of the head.

Also like IAD, this cue is usable at all frequencies, albeit with a key qualification. At frequencies below about 1500 Hz, the neural response from the basilar

Amazon
EX1013 Page 758

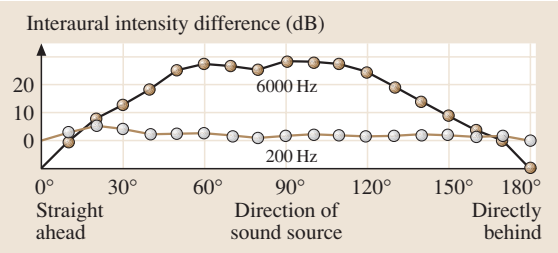


Fig. 18.7 Interaural amplitude difference as a function of direction [18.19]

membrane remains phase locked to the bandpass-filtered audio waveform. At such frequencies, even a modest amount of interaural phase shift of a sine wave can result in a shift in the position of the audio image. Above 1500 Hz, neural phase synchrony is effectively lost, at least as far as ITD perception is concerned, so the interaural phase of a high-frequency sine wave has little or no perceptual effect on its apparent position. However, the brain can still make fairly accurate judgments of the ITD of the *envelopes* of the high-frequency bandpass-filtered signals, assuming there is sufficient variation in the filtered envelopes. This will tend to make spectrally dense high-frequency signals easier to localize than sparse signals, since the latter may only exhibit constant-envelope sine waves after the critical band filtering of the basilar membrane.

While localization in the horizontal direction is firmly mediated by two strong, well-behaved cues, localization in the vertical dimension is decidedly less well behaved. The goal of vertical localization is to differentiate the positions of sources on locus of constant IAD/ITD. For sources equidistant from the two ears, the locus is a plane passing though the center of the head, referred to as the median plane. For sources off to one side, the locus is a cone, sometimes referred to as the *cone of confusion*. The best-case resolution, for front-centered sources, is typically on the order of 5°, which is quite a bit less precise than the horizontal resolution of about 1° under the same conditions. As with horizontal localization, vertical resolution appears to diminish for side sources, although since the cone of confusion becomes progressively narrower, the net spatial error resulting from a vertical localization error is likely to diminish. For sources at the side, the cone of confusion collapses to a line.

One demonstrated vertical localization cue [18.20] is generated by the interaction of the arriving sound and the folds of the pinnae. The pattern of reflections off the folds depends on the direction of arrival of a sound, although not in any simple, monotonic fashion. One consequence of these pinna reflections is the introduction of sharp peaks and dips in the frequency response, which the brain can apparently learn to associate with specific directions. Because of the size of the pinna, and resulting reflection delays, these direction-dependent response delays only appear at high frequencies, above about 7 kHz, and so are less effective for low-frequency or narrow-band signals. Vertical localization of signals below 7 kHz may be possible on the basis of pinna cues if the brain is performing some sort of autocorrelation of each ear's signal, and can detect the delays in the time domain, but so far there is little evidence to support this hypothesis.

A second vertical localization cue is dynamic in nature, namely the modulation of the horizontal localization cues – ITD and IAD – with movement of the head, which will vary with the vertical position of the source. Given a front source on the median plane, for example, a rightward movement of the head will result in a left ear arrival that is louder and sooner than the right ear arrival. For a rear source, the correspondence will be opposite, and for a top source, there will be essentially no change in the ITD/IAD values. By correlating head movement with these sonic changes, the brain can deduce the vertical position of the source.

Finally, the presence and pattern of room echoes, once learned, and in concert with the above cues, can reinforce the vertical position of a source, especially whether in front of or behind the listener.

Even so, the vertical localization cues tend to be somewhat individual, like fingerprints, and, for example, playing a recording made with microphones placed in the ear canals of one subject may not result in good vertical localization for another listener.

18.3 Audio Specifications

With the basic limits of auditory resolution now reviewed, consideration can be given to the requirements for high-quality audio systems in terms of those limits. While individual audio components can be expected to exhibit correspondingly individual performance requirements, as will overall systems, there are some

Amazon
EX1013 Page 759

specifications that can be expected to apply to all elements of the signal chain. Some of these are considered traditional audio specs that have been used as general metrics for many years. Happily, in at least some cases, modern audio systems are of such high quality as to render some specifications moot. In other cases, complete fulfillment of some of the requirements is something that may never be completely realized.

18.3.1 Bandwidth

It is reasonably evident that, given the nominal 20–20 kHz range of human hearing, it will be desirable for audio systems with aspirations to sonic accuracy to support at least that frequency range. Since an audio signal may pass through any number of processors on its way from acquisition to ultimate reproduction, it is generally advisable for each processor to have a wide-enough pass band to avoid progressive bandwidth limitation.

For most audio devices during the formative years of audio development, this requirement was daunting, and rarely achieved outside the laboratory, if then. There were too many mechanical vibrating structures in the chain, and it is very difficult to make them respond consistently over a 1000:1 frequency range. Grooved disc recordings were limited by large-diameter, low-compliance styli, and slow inner-groove velocities of constant-RPM discs. Optical media, like film, and magnetic media, were both limited by linear speed, and the size of the effective apertures used for recording and playback. The linear speed of such media was chosen as a compromise with playing time, and the smallness of the apertures was limited by high noise levels and diminishing signal level. Broadcast media were limited by bandwidth constraints and noise.

Not counting a straight piece of wire, amplifiers are the audio component that have probably had the easiest time achieving full audio bandwidth, particularly small-signal amplifiers, thanks in part to the relatively low mass of the electron. Power amplifiers have had a slightly harder time, partly because devices that handle high powers often come at the expense of compromises in other parameters, and partly because loudspeaker systems often exhibit ill-controlled, non-resistive loads. The response and linearity of amplifiers of all sorts was materially improved by the invention of negative feedback by Harold Black of Bell Laboratories, on August 2, 1927 [18.21].

An early recording system with fairly wide bandwidth was demonstrated on April 9–10, 1940, by Harvey Fletcher of Bell Laboratories, and conductor Leopold Stokowski. This employed three optical audio tracks on high-speed film, with a fourth track for a gain control signal, apparently implementing a wide-band compressor/expander for noise reduction, and possessing a frequency response of about 30–15 000 Hz.

It was probably not until the mid-1950s that such a bandwidth became commonplace in consumer media, with improvements such as the long-play record, commercial FM broadcasting, and refinements in magnetic tape, heads, and circuits.

As with a number of other basic audio specifications, the question of full-bandwidth frequency response was, to a great extent, rendered moot with the commercial introduction of digital audio in 1982. As discussed in a subsequent section, the bandwidth of a digitally sampled signal is largely dependent on the sampling rate chosen, which must be at least twice as high as the highest audio frequency of interest. Thus, conveying a bandwidth of 20 kHz requires a sampling frequency of at least 40 kHz. Current standard sampling frequencies in common use for full-bandwidth audio systems are 44 100 Hz and 48 000 Hz, depending upon the medium. Sometimes, a slight reduction in high-frequency response is tolerated, and a sampling frequency of 32 000 Hz is used to reduce the data rate slightly. Audio perfectionists, on the other hand, often prefer sampling rates of 96 000 Hz or more, to support an effective bandwidth in excess of 20 kHz and thereby provide additional safety margin to avoid alias distortion, preserve harmonics above 20 kHz, avoid possible intermodulation distortion, and mitigate possible filter distortion.

Once converted to digital representation, an audio signal tends to maintain its bandwidth unless explicitly limited by some digital signal processing (DSP) operation.

Despite the advances of digital audio, full-range audio systems are still the exception. The chief culprits for this situation are the transducers, especially loudspeakers. At low frequencies, naturally occurring combinations of mass and compliance lead to a fundamental resonance for most typical loudspeaker drivers that falls well inside the audio band, imparting a high-pass characteristic that usually rolls off at lower frequencies at a rate of at least 12 dB per octave. At high frequencies, a similar fundamental resonance leads to a similar roll-off, with a lowpass characteristic. Some of these roll-offs can be partially mitigated with electronic equalization, but the steepness of the characteristic makes it difficult to extend the response very far

Amazon
EX1013 Page 760

with equalization, without incurring unduly high-power requirements.

In general, most musical and speech energy is concentrated in a subset of the audio band, say from 100 Hz to 10 kHz, allowing the majority of audio signals to be satisfactorily reproduced with modestly compromised system bandwidth, to the benefit of size, cost, and bulk.

18.3.2 Amplitude Response Variation

From the basic amplitude acuity of the ear, it is logical to require that the pass-band deviation in frequency response be less than an amplitude JND, say within ± 0.25 dB of flat. Of course, this is the desired net response of an entire signal chain, and if deviations are known to exist in one part of the chain, it may be possible to apply compensating equalization elsewhere in the chain.

As with the quest for full audio bandwidth, early audio systems for many years had trouble meeting this requirement, in part from mechanical limitations. Amplifiers again have traditionally had the easiest time meeting this specification, and modern-day DSP is limited only by the precision of the analog/digital converters, which is generally quite good in this regard, plus any deliberate signal-processing spectral modifications.

The worst offender is many cases is the listening room, where echoes and resonances are likely to make nearly any sine-wave frequency response plot a jumble of peaks and dips covering many dB. The three-dimensionality of the room and the fact that echoes arrive from diverse directions work to reduce the audibility of many of these measured deviations.

After rooms, loudspeakers have the hardest time meeting this flat-response requirement, being mechanical devices and having to handle a lot of power and move a lot of air. In order to cover the bulk of the audio band, loudspeaker designs typically employ multiple drivers, each optimized for a portion of the audio band, but this in turn introduces potential response deviations from required crossover networks.

In considering reasonable values for most audio specifications based on psychophysical performance, it is prudent to consider both monotic and dichotic performance; that is, each ear considered in isolation, and then the binaural performance of the two ears together. In the case of amplitude response variation, similar JNDs apply, so there is no compelling reason to differentiate the situations in this case.

18.3.3 Phase Response

In the course of passing through an audio component, an audio signal may encounter delays, which may be frequency dependent, and may therefore be categorized as either frequency-dependent time delays, group delays, or phase shifts. It is therefore prudent to establish preferred perceptual limits on the allowable deviations of these quantities.

Of course, a pure wide-band time delay will be manifested as a linear change in phase shift proportional to frequency, so to separate the effects of time delay and phase shift, so the latter ("frequency-differential single-channel phase shift") may commonly be understood to represent the deviation from a linear-with-frequency phase characteristic.

It is a little difficult to be completely definitive about a phase-shift specification, in part because there are multiple neural timing mechanisms involved. For a sound channel considered in isolation, phase perception is rather weak. There does not seem to be an explicit means of comparing the fine-grain phase or timing of signals at significantly different frequencies (more than a critical bandwidth), at least for steady-state signals. This is sometimes used as the basis of a claim that the ear is virtually phase deaf, although that is probably an overstatement. Still, if one constructs a steady-state signal with widely space spectral components that are in slow precession, say a combination of 100 Hz and 501 Hz (360° per second relative phase precession), the resulting percept is almost entirely constant.

So under what conditions are phase shifts audible? There seem to be at least three such mechanisms: fusion time, transient fusion time, and critical band envelopes.

As noted earlier, fusion time refers to the audio event integration time of the ear, generally in the range of 20–50 ms. Separate audio events that occur farther apart than the fusion time will generally be heard as separate. If a spectrally complex signal is put through a system where some frequencies are phase- or time-shifted by more than the fusion time, the signal may temporally defuse, rendering the phase shift quite audible. (A 20 ms phase delay at 30 Hz corresponds to 216 degrees of phase shift, not an unrealistically large value.) This can occur, for example, if a sharp pulse is passed through a third-order, 1/3-octave filter bank tuned for flat magnitude response. What starts out as a click will emerge as a "boing", quite audibly distinct from the source.

Transient fusion time refers to the shortest interval the ear can differentiate temporally, as exemplified by the premasking interval of about 1–2 ms. If a sharp tran-

Amazon
EX1013 Page 761

sient, shorter than the transient fusion time, is passed through an audio channel where the high-frequency phase shift exceeds the transient fusion time, the output signal may sound audibly time-smeared compared to the source. Although the effect can be subtle, it is often manifested as the difference, e.g., between a sharp strike with a hard mallet, and a wire-brush mallet hit.

Critical band envelopes come into play with regard to the audibility of phase shifts with steady-state spectrally dense signals, more specifically signals in which two or more significant signal components fall within a common critical band. Since the basilar membrane filter bank will be unable to segregate these signals into separate bands, they will beat with one another, and the pattern of beats may be quite audible, for example, a complex composed of equal parts 500 Hz and 501 Hz, which can be considered the 500th and 501st harmonic of a 1 Hz pulse train. These two spectral components are certainly within a critical band of one another, and the instantaneous phase relation between them will certainly be audible; indeed the signal will cancel altogether once per second. Compare that to the earlier example of 100 Hz and 501 Hz, the 100th and 501st harmonic of a 1 Hz pulse train, where the phase relation remained virtually inaudible because of the wider frequency spacing.

So, it can be safely hypothesized that keeping frequency-differential single-channel phase shift under 1 ms is probably pretty inaudible, assuming small changes over the width of a critical band, and in practice, with most non-transient music, speech or other natural sounds, it can be expected that phase shifts corresponding to frequency-differential time shifts much under the fusion time of 20–50 ms (at least at low frequencies) will be fairly benign, if not almost totally inaudible under most conditions.

Once again, amplifiers have the easiest time meeting this requirement, with multi-driver loudspeaker systems being one of the more problematic components. Although sharp resonances, the bane of mechanical acoustic components, may exhibit audible amounts of phase shift, they will often be rendered insignificant compared to the amplitude deviations introduced by such resonances. And, again, with DSP systems, audible phase shift is not likely to be an issue as long as the signal remains in the digital domain, unless such phase shift is purposely introduced with signal processing.

As can be readily gleaned from binaural psychophysics, dichotic phase shift – the phase shift between a pair of channels – can be more audible at lower delay levels than monotic phase perception. A timing difference of even a single sample at 48 000 Hz sampling, about 20 μs, is twice the JND of this parameter. Of course, the perceptual metric for binaural phase difference is a change in apparent image position or width, rather than the spectral or amplitude change associated with monotic phase shift. It is possible for audio systems to have audible amounts of monotic phase shift and still meet the dichotic (interchannel) phase requirement, as long as all channels have the same amount of phase shift.

As usual, amplifiers and digital signals have the easiest time with this specification. In the past, some multichannel analog recording media, such as magnetic tape recorders or optical film recorders, tended to exhibit small amounts of interchannel timing error, owing to misalignment of the recording and playback apertures. Real-world rooms and loudspeakers have the hardest time, since differential timing of 10 μs corresponds to a path-length difference of only about one eighth of an inch, and few listeners are willing or able to hold their heads that still. This effect is somewhat mitigated because the plethora of echoes in a typical room leads the brain to discount timing information in favor of first-arrival amplitudes, which are somewhat better behaved. Still, it can be readily appreciated that the virtual imaging of stereo reproduction is likely to be fairly sweet-spot dependent, since listener positions near a loudspeaker will tend to receive sound both sooner and louder from the near loudspeaker, resulting in spatial distortion of the sound stage.

18.3.4 Harmonic Distortion

The presence of nonlinearities in the transfer function of an audio device or system is likely to have the effect of introducing new, spurious spectral components. For a single isolated sine wave, the nonlinearity will result in a new waveform with the same periodicity as the original signal, which can then be expressed via Fourier analysis as a sum of the original sine wave plus harmonics thereof [and perhaps a direct-current (DC) component]. For more spectrally complex signals, each discrete spectral component will be subject to such spurious harmonic generation. This process is referred to as harmonic distortion, and the power sum of the generated harmonics, divided by the power of the original signal, is referred to as total harmonic distortion (THD), usually expressed as a percentage.

The limit of acceptable quantities of THD is the point at which the generated harmonics become audible. In part, this may be signal dependent, since distortion products of a low-level signal may themselves fall be-

Amazon
EX1013 Page 762

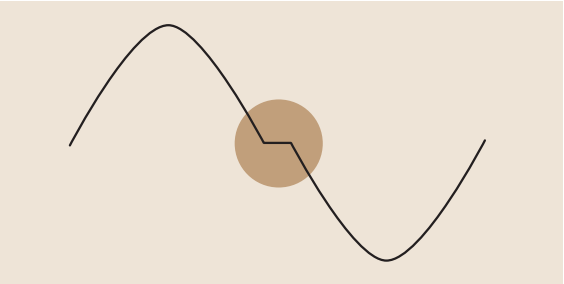


Fig. 18.8 Sine wave with crossover distortion (http://www.duncanamps.com/technical/ampclasses.html)

low the threshold of hearing, rendering them inaudible. For higher-level signals, the distortion products will be inaudible if they are masked, rendering the effect of harmonic distortion decidedly signal dependent. (The amplitude dependency may by further emphasized by nonlinearities that are themselves amplitude dependent, such as clipping.) For example, a wide-band noise signal may mask all harmonic distortion components of a moderate level. Harmonic distortion may therefore be most audible with spectrally sparse signals, with minimal masking components, such as a single sine wave.

For a sine-wave signal, the maximum acceptable amount of harmonic distortion can be deduced from the masking curve of that sine wave. From the masking curve shown in Fig. 18.4, it may be deduced that, since the masking level for a 1 kHz tone is around −35 dB at 2 kHz, the tolerable amount of second-harmonic distortion is about 2%. By 3 kHz, the masking curve is down around 50 dB, or about 0.3% third-harmonic distortion. Higher-order harmonics would benefit from even less masking, making them even more likely to be audible, until their amplitude falls below the threshold of hearing. Thus, the audibility of harmonic distortion will depend on the order of the distortion, with low-order distortion generally being less audible than high-order distortion.

Low-order harmonic distortion is generally associated with smooth, gentle nonlinearities, such as the typical onset of clipping of a vacuum-tube amplifier or a loudspeaker driver. High-order harmonic distortion, in contrast, is likely to result from sudden, sharp nonlinearities, such as the crossover distortion of an improperly designed solid-state power amplifier.

Among more traditional audio devices, playback of grooved media was probably fraught with the greatest amount of high-order THD, caused by sharp mistracking of the stylus. In the age of digital audio, low-level undithered digital signals will likely exhibit stair-step behavior which will also tend to be rich in high harmonic content, rendering them highly objectionable from the perspective of THD. Although this can lead to a desire for very high quantization accuracy (say 32-bits per sample), proper use of dither to spread the harmonic energy out as noise is generally the preferred solution.

18.3.5 Intermodulation Distortion

Intermodulation (IM) distortion results when two or more spectral components interact with a nonlinearity. The resulting distortion products will tend to be generated at sums and differences of the original spectral components. For even low-order IM distortion, these distortion products may be far in frequency from the source signal components, and may in fact be much lower in frequency. Since masking tends to be most effective when the signal to be masked is higher in frequency than the masker, having IM distortion products lower in frequency than the generating components renders them more likely to be audible, in turn making IM distortion potentially more bothersome than THD.

Serious IM distortion tends to be associated with sharp nonlinearities, and was a problem with analog grooved recordings when the stylus mistracked, as well as with poorly designed power amplifiers exhibiting crossover distortion. It can also show up with low-level signals and digital converters. Loudspeakers usually have less trouble with IM distortion, in part because any nonlinearities are usually more gentle (associated with maximum output limiting), and in part because IM distortion products located outside the range of a driver will tend not to be radiated.

Although the theoretical audible limit of IM distortion is a miniscule fraction of 1%, a value of 0.1% IM distortion or less is usually pretty safe.

18.3.6 Speed Accuracy

Considerations of speed accuracy encompass both absolute accuracy and dynamic variations, sometimes subdivided into slow variations called *wow* and faster variations called *flutter*. Except in extreme cases, timing error is usually perceived as an error in pitch, rather than time. Unless it is drastically in error, absolute speed accuracy is likely to be of concern mostly to folks with absolute pitch.

Speed variations such as wow and flutter are likely to be audible to anyone with normal pitch perception. Given a pitch JND of 0.7%, which would correspond to peak-to-peak just-noticeable speed variation, the cor-

Amazon
EX1013 Page 763

responding root-mean-square (RMS) speed variation would be about 0.25%. Since speed variation may be cumulative, a preferred figure of 0.1% is likely to be specified [18.22]. There may also be cases in which cancellations from room echoes cause amplitude variations resulting from flutter that may require still lower limits of flutter to be rendered inaudible.

The analog mechanical and electromechanical recording devices of the past all have had to contend with speed accuracy. Turntables, tape drives of all stripe, and film projectors all fought off the effects of speed variation, usually with large flywheels and elaborate bearings. Getting a small lightweight tape drive like the Sony Walkman® to have low flutter was a good engineering trick. Properly designed and maintained equipment was generally able to meet the required speed accuracy specification.

In the age of digital audio, speed errors of all sorts are mostly a dead issue. Sampling clocks controlled by quartz oscillators provide speed accuracy and steadiness far in excess of what the ear can hope to detect.

18.3.7 Noise

Noise is a first cousin to distortion in its characteristics and considerations. It is a random signal, hopefully low in level, that gets tacked onto an audio signal in the course of almost any processing, analog or digital, and it is generally cumulative in nature. It often sounds like a hiss. Noise will be inaudible if it is either below the threshold of hearing, or masked by the audio signal.

Sources of Noise

Noise sources include:

Acoustical:	Random fluctuations of air molecules against the eardrum or a microphone diaphragm
Electrical:	Tiny fluctuations of electrical charge in analog electronic circuits from discrete electrons
Mechanical:	Surface irregularities of grooved media
Magnetic:	Irregularities of magnetic strength from discrete magnetic particles in recording tape
Arithmetic:	Random errors from DSP quantization of numerical signals to discrete values.

There are other common spurious signals, such as hums and buzzes, that can corrupt an otherwise pristine audio signal, and although strictly speaking they may not be considered noise, they are generally at least as undesirable.

In any case, the level of noise can be considered as a function of frequency, giving rise to the notion of the spectrum or spectral density of the noise. If the value of an audio noise signal at any point in time is uncorrelated with the value of the noise at any other point in time, the signal is an example of what is called *white noise*. Spectrally, white noise is said to be flat in a linear frequency sense. That is, in essence, the output energy of a bandpass filter of bandwidth B Hz. supplied with white noise, will be the same regardless of the center frequency of the filter (providing, of course, that the pass-band of the filter is substantially within the audio band).

Because the filters on the basilar membrane are approximately logarithmic (not linear) in center frequency and bandwidth, and because they tend to integrate the signal energy within each pass-band, white noise will likely be perceived as having a rising high-frequency spectrum, even though it is flat on a linear frequency basis. In other words, as the center frequency of each effective basilar membrane bandpass filter increases, so too does its bandwidth, roughly in proportion, ergo the total filtered energy of a white-noise source will increase with frequency, leading to the bright, top-heavy perception. For each successive octave, or doubling of center frequency, the bandwidth will double, ergo so too will the bandpass white-noise signal energy, leading to a characteristic spectral rise of 3 dB per octave.

Note: dB = 10 log (ratio of power) = 20 log (ratio of amplitude).

If white noise is processed with a filter having a constant roll-off characteristic of −3 dB per octave, the output noise will have constant energy per octave, or fraction thereof, and will sound flatter to the ear. Such a signal is referred to as pink noise.

Although many forms of noise are independent of the presence of an actual signal, there are mechanisms whereby noise, or noise-like artifacts, are signal dependent. The most common such behavior is for the level of the noise to be roughly proportional to signal level. This type of artifact can be especially pernicious, as the dynamic behavior tends to call attention to the noise.

Although there are numerous ways of dynamically suppressing noise if it is unavoidable, it is by far preferable to avoid it in the first place as much as possible via proper low-noise design, proper shielding and grounding, etc.

Amazon
EX1013 Page 764

18.3.8 Dynamic Range

Dynamic range is the amplitude range from the largest to the smallest signal an audio device or channel can handle. In cases where there is no minimum signal, other than zero level, the minimum useable level may be taken as the noise level of the channel.

If the noise level of the channel is substantially independent of signal level, the dynamic range of the channel may alternately be referred to as the signal-to-noise ratio (SNR). If the noise is signal dependent, the audibility of the noise will depend on the instantaneous signal-to-noise ratio, which is likely to be quite different from the dynamic range.

Since both the noise level and the overload point of a channel are likely to be functions of frequency, so too will the dynamic range. A common characteristic of the noise of traditional analog recording systems has been rising noise levels at the ends of the audio band, coupled with diminishing maximum signal levels, a double whammy that tends to limit the resulting dynamic range, in some cases rather severely.

Historically, unadorned analog channels have tended to have overall dynamic-range specifications on the order of 45–60 dB. This is true of such media as FM stereo, grooved records, optical film soundtracks, and magnetic tape. Perhaps the most limited hi-fi medium in this regard has been cassette tape, which manages about 55 dB at midband and on the order of 30 dB dynamic range at high frequencies.

Happily, DSP systems usually have flat overload characteristics and flat, or at least white, noise floors. The dynamic range of a DSP system is largely a question of the number of bits used to represent the signal. A rough rule of thumb is that each additional bit in the representation of sample values adds 6 dB to the dynamic range. The common 16-bit format used on digital compact discs yields a theoretical dynamic range of 96 dB. In practice, the combined effect of real-world imperfections and converter noise limits the practical attainable dynamic range with 16-bit audio to a little short of 96 dB, perhaps about 93 dB. A 24-bit converter can approach 144 dB dynamic range, which not only allows coverage of the full human auditory dynamic range, but allows margin for error in setting recording levels and subsequent mixing and processing operations.

18.4 Audio Components

With the common audio specifications now delineated, consideration can be given to some of the common audio components, their typical specifications, and specific characteristics and limitations.

18.4.1 Microphones

For any acoustical recording, the type, pickup pattern, position, and characteristics of the microphone are likely to have the greatest effect on the resulting quality of the recording. A microphone is, of course, an analog transducer, generally used to convert acoustic signals to proportional electrical signals. It is the functional analog of the eardrum in the human auditory system.

The term microphone was coined by Sir Charles Wheatstone in 1827, although at that point it referred to a purely acoustical device, such as a stethoscope. Its start as a transducer was pretty much coincident with Bell's invention of the telephone. Bell's initial *transmitter* used a vibrating needle connected to a diaphragm and immersed in an acid bath to implement a variable resistance responsive to acoustic waves [18.23]. This was followed in short order by more practical, rugged designs by Bell, Edison, Berliner, Hughes, and others.

Common to virtually all microphones ever made is a diaphragm. Most microphones do not directly sense the instantaneous pressure deviations of an airborne acoustic wave. Instead, the sound waves impinging on a diaphragm cause movement of the diaphragm, and it is this movement that is sensed and converted to an electrical audio signal.

The methods used for the mechanical-to-electrical transduction are many and varied. Bell's telephone patent seems to describe an odd type of electromagnetic microphone, in which the diaphragm was mechanically coupled to an armature in a magnetic circuit. Acoustic vibrations caused vibration of the armature, in turn inducing changes in the reluctance of the magnetic circuit, which induced a dynamic current in a coil of wire.

Later dynamic microphones simply attached a fine coil of wire to the diaphragm, and suspended it in a magnetic field. Movement of the diaphragm induces a voltage in the coil of wire, presumably proportional to the velocity of the diaphragm, and said voltage can

Amazon
EX1013 Page 765

then be amplified as necessary. Dynamic microphones do not require an external source of power. Their popularity was limited prior to World War II by a lack of powerful permanent magnets.

Many early telephone transmitters used carbon in one form or another, typically carbon granules coupled to the diaphragm such that vibration of the diaphragm would dynamically squeeze the granules, causing a variation in resistance that could be electrically sensed by an applied voltage. This arrangement could be scaled up in size and voltage enough to produce a sizeable signal, of particular importance prior to the advent of electrical amplification.

Another type of self-generating microphone is the crystal microphone, based on the piezoelectric effect discovered by the Curies in 1880. Here, the diaphragm constricts a thin crystal made of Rochelle salt, ceramic, or similar material, directly producing a voltage. The low compliance of the piezoelectric element typically limits the performance of such a microphone, especially the extent of its frequency response.

Indeed, the performance of most of the microphone types mentioned so far is somewhat limited by the need to attach some object or structure to the diaphragm to detect its motion, increasing the mass of the moving elements, which in turn tends to limit the high-frequency response.

Two types of microphones avoid any mechanical attachment to the diaphragm: the ribbon microphone and the condenser microphone.

A popular variant of the dynamic microphone, the ribbon microphone, uses a thin metal ribbon diaphragm suspended in a magnetic field. Movement of the diaphragm induces a current in the diaphragm, which can then be amplified. In effect, the diaphragm acts as a single-turn moving coil. The ribbon microphone was developed by Schottky and Gerlach in 1923, and considerably refined by Harry Olsen of RCA in 1930 [18.26].

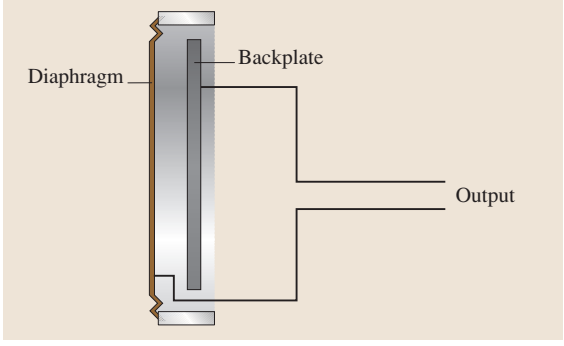


Fig. 18.9 A condenser microphone consists of a very thin diaphragm suspended parallel to a backplate (After [18.24])

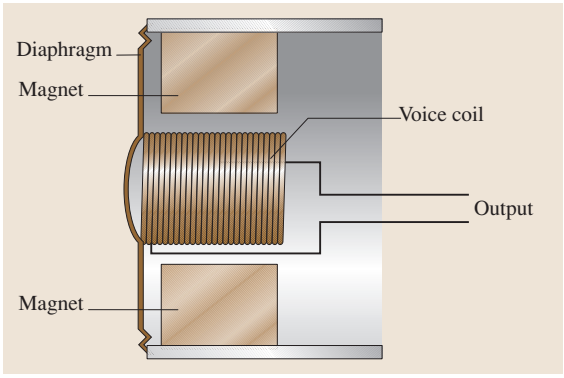


Fig. 18.10 A dynamic microphone uses a coil of wire attached to a diaphragm, suspended in a magnetic field (After [18.24])

The condenser microphone is based on the notion of electrical capacitance. A metal, or other conducting, diaphragm is placed close to a second, fixed conducting plate, forming a capacitor, and a polarizing voltage is applied between the plates. Movement of the diaphragm causes alterations in the capacitance, in turn causing corresponding alteration of the charge on the plates, which is converted to a dynamic voltage by a resistor, and amplified as necessary. The invention of the condenser

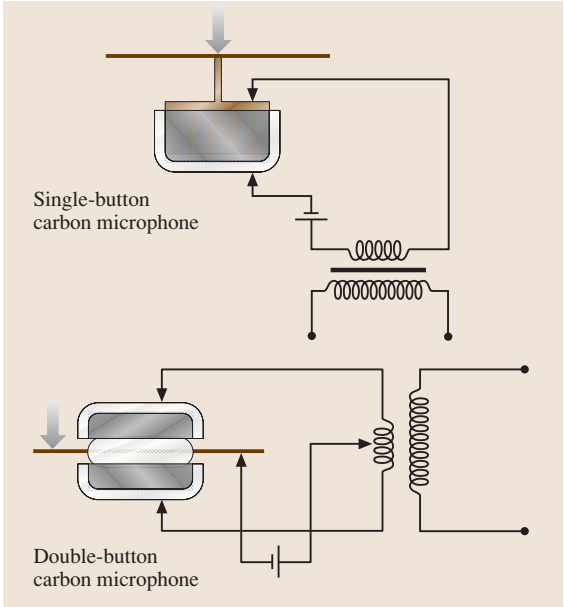


Fig. 18.11 A carbon microphone (After [18.25])

Amazon
EX1013 Page 766

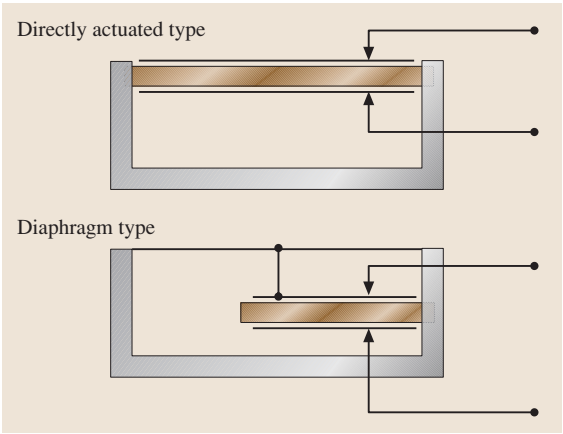


Fig. 18.12 Piezoelectric microphones (After [18.29])

microphone is generally credited to E. C. Wente of Bell Laboratories, in 1917 [18.27]. A popular modern variant is the electret condenser microphone, perfected in 1962 by Sessler and West of Bell Laboratories [18.28]. An electret is an object possessing a permanent electrostatic charge, making it rather like the electrical analog of the permanent magnet. Use of an electret to polarize the capacitor in a condenser microphone eliminates the need to supply a polarizing voltage for the capacitor. Since electret condenser microphones are used almost universally in cell phones, as well as other applications, they are probably the most popular type in current use. The diaphragm of electret condenser microphones is usually made of a thin sheet of flexible plastic which has had a metal coating deposited on it, a notably lightweight arrangement that permits good sensitivity and frequency response. Such microphones also usually exhibit good sample-to-sample consistency, making them practical for multi-microphone arrays.

A more recent variant of the condenser microphone is the silicon microphone. This is a microscopic condenser microphone constructed using micro electromechanical system (MEMS) technology. The required electrostatic polarizing field is provided by a charge pump, and the diaphragm can be free floating, being held in place by electrostatic attraction.

In recent years, the motion of a diaphragm has been measured using optical means, including laser interferometry, which may in time become a practical alternative to the capacitor microphone.

Microphone Dynamic Range

The noise level of a microphone is limited by Brownian motion of air molecules impinging on the diaphragm. Of course, it can be further limited by electrical noise of the amplification which follows it, but even discounting those effects, the so-called self-noise of a microphone is likely to be 10–20 dB higher than the threshold of hearing [18.30]. Lower self-noise is possible with larger diaphragms, say on the order of one inch, but this comes at a price of making the microphone more directional at high frequencies.

The loud signal limit of a microphone is often set by the maximum signal capacity of the associated preamplifier. Using large rail voltages can be of some aid in this regard. Eventually the diaphragm will reach the limits of its travel, and experience soft clipping. There has been some work on using force feedback to linearize the characteristics of the diaphragm [18.31, 32]. With a little effort and care, it is possible to make microphones that can handle high-level signals about as well as the human ear can.

Finally it should be noted that there has been some work on creating a microphone that does not require a diaphragm. Blondell and Chambers did some work on a flame microphone in 1902 and 1910 [18.23]. More recent efforts have used lasers to detect index of refraction changes due to acoustic pressure waves [18.33].

Microphone Pickup Pattern

Interacting with a real, three-dimensional sound field, a microphone is an audio device inextricably tied to spatial considerations. Immersed in the sound field, the microphone responds to the sound at essentially just one point, so its output represents a spatial sample of the total sonic event. This is not to say that it necessarily responds equally to sounds arriving from all directions; on the contrary, most microphones exhibit some directional bias, whether or not intended, and it is probably not possible to specify a single universally preferred pickup pattern.

The simplest arrangement is to put a sealed chamber behind the diaphragm, allowing only a slow leak for stabilization of quiescent atmospheric pressure. With short-term pressure behind the diaphragm fixed, the microphone responds to the sound pressure at the front of the diaphragm. Since pressure is a spatial scalar, it might be expected that a pressure microphone might respond equally to sound from all directions, and indeed this might be the case if the diaphragm had zero dimension. As it is, a finite-dimension diaphragm will spatially integrate the sound pressure across its surface. This will tend to increase output for normal-incident waves and waves with long wavelength, but will have less boosting effect for random incidence sounds, such as Brownian

Amazon
EX1013 Page 767

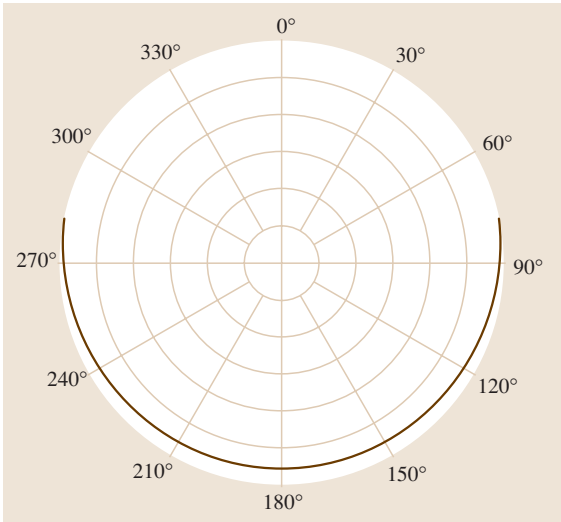


Fig. 18.13 *Omnidirectional* microphone pickup pattern. Note the slight attenuation of sounds arriving from the rear (After [18.35])

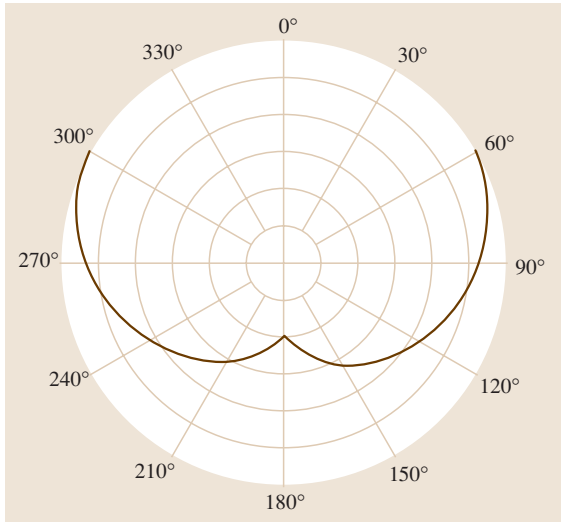


Fig. 18.15 Cardioid microphone pickup pattern, obtained by combining omnidirectional and bidirectional pickup patterns (After [18.35])

noise, and may reduce response to sounds with oblique incidence and wavelengths shorter than twice the diameter of the diaphragm. The resulting pickup pattern may therefore become directional at high frequencies, while remaining omnidirectional at lower frequencies. To remain omnidirectional up to 20 kHz, a diaphragm will have to have a maximum diameter of half the wavelength at the upper frequency limit, or about 8 mm. Whether by accident or design, this happens to be the approximate diameter of the human eardrum [18.34]. The fact that the eardrum is so small, and is coupled to the structures of the inner ear, and yet the ear still manages to exhibit a maximum sensitivity of 0 dB SPL or better has to be considered impressive performance, especially considering that quiet sounds just fade into perceived silence, rather than some inherent background perceptual noise.

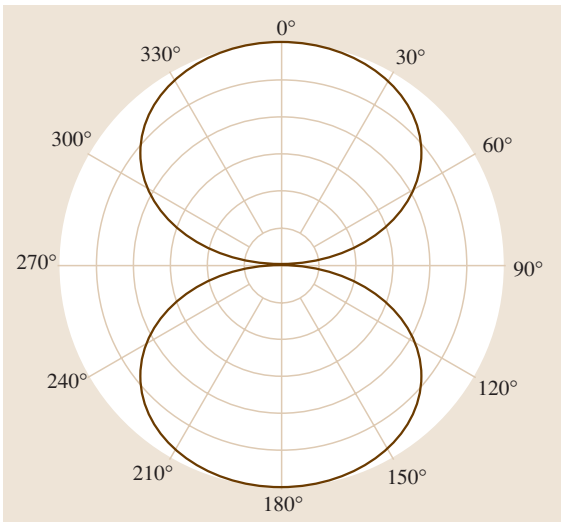


Fig. 18.14 Bidirectional microphone pickup pattern, typical of an open-back ribbon microphone (After [18.35])

More elaborate microphone pickup patterns (Figs. 18.13–18.15) can be achieved by increasing the number and complexity of acoustic paths to the diaphragm. For example, simply opening the back surface of the diaphragm to the atmosphere, the arrangement used in the classic open ribbon microphone, results in a microphone sensitive to the difference between front and back pressure, so sounds arriving from the side produce no response, and the pickup pattern usually resembles that of Fig. 18.8, albeit with the front arrival signal of opposite phase to the rear arrival signal.

If a pickup pattern as in Fig. 18.8 is summed appropriately with an omnidirectional pickup pattern, the rear arrival sound can be made to cancel, producing a cardioid front-directional pickup pattern.

With more elaborate channeling of the audio reading the diaphragm, more specialized or directional pickup patterns may be obtained.

Greater flexibility in the choice of pickup pattern can be obtained by using arrays of microphones, typically arranged in fixed physical orientations, where the

Amazon
EX1013 Page 768

effective pickup pattern of the array is dependent on how the individual microphone signals are electrically summed. By recording the individual signals, the array can be effectively focused and oriented after the fact by postprocessing the recorded signals. The large-volume low-cost manufacture of certain types of microphones has made this approach especially attractive.

18.4.2 Records and Phonograph Cartridges

Although by now semi-obsolete, the phonograph cartridge has historically played a key role in sound reproduction, and is fairly close to a microphone in many aspects of operation. As with a microphone, the intent is to act as a transducer for acoustical vibration, in this case the vibration originating from a stylus following the undulations of a groove, rather than a diaphragm responding to vibrations in air.

Although early phonographs employed direct mechanical-to-acoustical transduction, the advent of electronic amplification made the electromechanical phono cartridge the default device for record playing.

Most of the methods of transduction used for microphones have been tried for phono cartridges as well, with piezoelectric and magnetic being the two most popular. Whereas most magnetic microphones have been of the moving-coil type, the magnetic cartridge design has employed both moving magnet and moving coil designs. Perhaps the use of a cantilever design of the stylus to reduce the effective mass of the coil or magnet seen by the stylus tip, coupled with the lack of a diaphragm, led to the preference of magnetic transduction for phonograph cartridges, while the use of capacitive transduction in condenser microphones has tended to be preferred in that transducer.

Despite the popularity of the grooved phonograph record for well over half a century, it has always been a difficult medium with which to achieve full fidelity.

For one thing, the stylus assembly is fraught with high-frequency resonance problems, much like the microphone, which have tended to limit the high-frequency response. In many cases, the high-frequency resonance has been underdamped, resulting in a serious peak in the mid-high-frequency response. The material used to mount the stylus in a way to secure it but still allow it to vibrate, often little bits of rubber, can sometimes degrade with age, impairing the characteristics of the cartridge. Then there is the need for the cartridge to drag a tonearm across the record with it.

In theory, the tonearm should provide a rock-stable platform for the cartridge at audio frequencies, so that only the stylus vibrates. Below the audio band, the tonearm should move as freely as possible, to follow the slowly spiraling groove and miscellaneous movements induced by the rotating disc, including horizontal eccentricity and vertical warp. In practice, this behavior cannot be achieved perfectly (since a perfect *brick-wall* filter is impossible), but at least the cartridge and tonearm should be matched, and the low-frequency resonance peak resulting from the compliance of the stylus and the effective mass of the tonearm should be properly damped, to avoid a peak either in the low-frequency audio response, or in the subsonic region where a resonance could make the system overly sensitive to record warp. Even in the best arrangement of properly damped resonance centered an octave or so below the audio band, the system is only second order, so some low-frequency tonearm vibration and some subsonic stressing of the stylus from pulling the tonearm is inevitable [18.36], which can lead to distortion and uneven frequency response. To avoid the build up of acoustic traveling waves in the tonearm, it is desirable to have its construction and material chosen to damp out such vibrations; not all tonearms have done this well.

Another problem that has hampered record playback, although not strictly the province of the phono cartridge, is that the playback stylus usually does not exactly follow the same path and orientation as the cutting stylus. This is because discs are cut on a cutting lathe, in which the cutter slowly traverses a straight-line path from the outer grooves to the inner grooves. The playback cartridge, in turn, is usually mounted on a pivoted arm, so its path is an arc. (The longer the tonearm, the smaller the disparity between record and playback orientations.) In order to minimize the path mismatch, a right dogleg bend has often been incorporated into the shape of the tonearm. The cartridge has to be carefully mounted on the tonearm so that the stylus is a specified distance from the pivot if this feature is to be properly exploited. Unfortunately, the bent tonearm traversing an arc results in a steady-state force from friction between the moving record groove and stylus tip, one which does not pull straight down the length of the tonearm, but instead generates a sideways force, called the skating force, which pulls the stylus tip against one or the other of the groove walls, which is another potential source of distortion. The skating force varies with groove speed and stylus tip size and shape, among other things, making it challenging to fully compensate.

At various times straight-line tracking tonearms have been produced, usually with the back of the tonearm mounted via bearings to a platform intended to move

Amazon
EX1013 Page 769

in a straight line to follow the front of the arm (and the stylus). Since the stylus cannot be expected to drag the rear platform along with it, some sort of servo is used to insure that the platform follows the slow movements of the stylus/cartridge combination. Although this should mitigate skating-force problems, assuming the cartridge is mounted precisely, it still leaves the other difficulties of disc playback.

Another issue with phono cartridges is the shape of the stylus tip. The cutting stylus is chisel-shaped, with sharp corners, so it is capable of inscribing very fine undulations. But the traditional stylus tip shape is round, so the points of contact with the groove walls vary with the curvature of the groove. A round stylus will therefore follow a slightly different path than the cutting stylus, another source of distortion. In the 1960s, elliptically shaped styli became popular to reduce the tracing error, by more closely approximating the shape of the cutting stylus. The smaller stylus tip radius of curvature at the point of contact, compared to a spherical stylus tip, should allow the stylus to more accurately track the motion of the cutting stylus, but results in higher localized pressure, which can increase groove wear unless low tracking pressure is employed. Even so, the precision with which a given playback stylus follows the motions of the cutting stylus is unlikely to be sufficient to avoid some alteration in sound quality.

Around the time elliptical styli were becoming popular, RCA attempted to improve tracing performance by pre-distorting the cutter signal, anticipating playback with a round stylus. The system was not universally welcomed [18.37], in part because it arguably impaired performance with elliptical styli, and was eventually dropped.

Ultimately it has to be accepted that there will be some, hopefully slight, difference between the motion of the cutting stylus and the motion of the playback stylus, and that there are some signals that the playback stylus cannot hope to track properly, such as a constant-amplitude-cut square wave. The best one can hope for is that the deviation will be small and gentle, and that the resulting distortion will be mostly low-order THD. Even so, distortion figures for records were not uncommonly in excess of 10%, especially for the slower inner grooves. The technique of stereo recording on discs, pioneered by Blumlein in 1931 and introduced commercially in 1957, further complicated the task of the playback stylus, as it then had to track both horizontal and vertical undulations simultaneously.

Getting adequate dynamic range from records was a tricky process, fraught with compromises. For one thing, the maximum signal level used tended to be traded off against total playing time. Large-amplitude grooves took up more real estate than quiet grooves, which led to the use of variable-pitch, look-ahead record cutters that could dynamically adjust groove spacing in response to groove amplitude. Such cutters functioned by increasing the spacing one revolution before the arrival of a loud sound, and maintaining the spacing for the revolution that followed. Further complicating the amplitude/playing-time conundrum is the desire on the part of some recording engineers to minimize use of the inner groove area, to minimize inner groove distortion and bandwidth limitations.

Maximum amplitude was also limited by considerations of what the playback styli and cartridges of the consumers could accommodate, which could lead to lowest-common-denominator compromises in performance. To better match tracking ability and noise floor to the average statistics of music spectra, the Recording Industry Association of America (RIAA) standard equalization curve was adopted in 1954 that progressively reduced low frequencies and boosted high frequencies on recordings, applying inverse compensation during playback. This, along with the characteristics of the stylus, record surface and groove speed, made the overload characteristic of records strong functions of both frequency and playing time, so the precise specification of overload characteristic of records is problematical at best. To be sure, there are specified suggestions for maximum cutting amplitude and velocity, but given the commercial realities, at least with pop music, that louder all too often equates with better, the suggested limits have not always been observed.

The noise level of a record is dependent on the smoothness with which a groove can be cut, which in turn depends on the cutting technique and the material used. Although wax and shellac were used in the early years of record production, more modern practice has been to use lacquer deposited on an aluminum disc base. In conjunction with the hot-stylus cutting technique introduced in 1951 [18.9], a lacquer master disc is capable of a dynamic range approaching 70 dB, although in practice a somewhat lower figure is sometimes encountered for the end product on a production basis. Most of the steady-state noise encountered with records is concentrated at mid to low frequencies. In pursuit of the goal of fitting a performance to the dynamic-range characteristics and limitations of grooved discs, it has traditionally been common practice to employ a combination of compression and limiting to the signal in the course of cutting a master disc. There have been some

Amazon
EX1013 Page 770

attempts to apply complementary compression and expansion to record audio to improve the system dynamic range, notably discs produced by dbx Inc. in the 1970s.

Successful record playback can also run afoul of a problem known as acoustic feedback, if sound from the loudspeakers causes the disc itself to vibrate as a diaphragm. Solutions include the use of an inert turntable mat, remote positioning of the turntable, and limiting the playback level.

Of course, dust, dirt, scratches and wear will tend not only to degrade the recorded signal but also to raise the noise level, including the possible introduction of clicks and pops.

Given the litany of compromises and limitations in disc recording and playback, the level of performance that has been achieved with this medium is laudable. Although largely eclipsed by digital audio, some refinement of grooved record playback continues, including efforts to achieve the holy grail of disc playback, contactless playback via light or laser beam, notably by Fadeyev and Haber at Lawrence Berkeley National Laboratory [18.38]. The availability of advanced DSP has allowed after-the-fact improvements in noise reduction and distortion reduction that would have been impossible in an earlier age. It is tantalizing to consider the performance that might have been possible with this medium if such processing had been available and incorporated into the basic architecture of the disc recording system.

18.4.3 Loudspeakers

Loudspeakers and headphones provide complementary transduction to microphones, namely converting electrical audio energy to acoustical energy. Many of the principles and notions regarding microphones can be applied to loudspeakers either directly or by, in a sense, turning the ideas upside down.

In a manner reminiscent of microphones, few loudspeakers convert directly between electrical and acoustical energy. Instead, the electrical energy is converted to mechanical energy, specifically the movement of a diaphragm, often cone-shaped or dome-shaped, which then radiates acoustical energy, sometimes aided by an acoustical coupling horn or lens.

Indeed, many of the same transduction methods that have been used in microphones have at least been tried for loudspeakers including magnetic (also known as *dynamic*), piezoelectric, and condenser.

The genesis of loudspeaker-like transducer actually predates Bell's invention of the telephone by a couple of years when, in 1874, Ernst W. Siemens described a dynamic moving-coil transducer, although he does not seem to have initially done much with it [18.39].

Bell's telephone, patented in 1876, also used a dynamic receiver. With a magnetic armature mechanically coupled to a diaphragm secured at its edges, it was probably closer in spirit to a headphone than a loudspeaker.

There followed a series of refinements that progressively brought the output transducer closer to the modern concept of a loudspeaker. In 1898, Oliver Lodge patented a spacer to maintain the air gaps in the magnetic circuit. Three years later, John Stroh came up with the notion of attaching the cone to the frame via a flexible, corrugated surround that improved the linear travel of the cone. Proper, flexible centering of the rear of the cone in the magnetic gap was provided by the *spider*, developed by Anton Pollak in 1909.

Many of these early loudspeakers used an acoustic horn of some sort to project the loudspeaker output into the room. The horn improves efficiency, partly by improving the acoustic coupling from the loudspeaker to the room, and partly by collimating the sound into a beam directed at the listener. Credit for the first true direct radiator loudspeaker design to include all the elements of what can be regarded as the modern-day loudspeaker is generally accorded to Rice and Kellogg in 1925 [18.40]. Their design took explicit account of the baffle on which the driver was mounted in the overall acoustic design.

There were of course, countless follow-on refinements to the basic electrodynamic driver, a process that continues to present day. Other methods of transduction were also explored. In 1929, E. W. Kellogg filed a patent for an electrostatic loudspeaker, consisting of panels not unlike large condenser microphones. The design has not been as widely accepted as the dynamic driver, in part due to some limitations in maximum output. There have also been piezoelectric drivers, but the limited travel of the piezoelectric element has limited their use to mid-range drivers, tweeters, and earphones.

So, as the Rice/Kellogg baffle indicated, with the basic dynamic loudspeaker configuration established, much of the attention focused on refining the loudspeaker/cabinet combination as a system.

The output ultimately obtained from a raw loudspeaker driver is profoundly affected by the surrounding baffle or cabinetry. In isolation, a loudspeaker driver produces two acoustical outputs: one from the front of the diaphragm or cone, and the second from the back. Unfortunately, the two outputs are out of phase. Were they generated at the same place and time, they would can-

Amazon
EX1013 Page 771

cel; the loudspeaker would produce nothing but heat. So part of the goal in designing a complete loudspeaker system is to keep those two outputs from annihilating each other.

At frequencies above the point where the wavelength is comparable to the size of the speaker cone, the radiation will tend to be a directional beam, and the front and back waves will tend to naturally go their own separate ways. At lower frequencies, the front and back radiation will tend to be more omnidirectional, and will therefore tend to cancel. This can be avoided by requiring that the size of the loudspeaker be at least as large as the wavelength of the lowest frequency of interest. For a low-frequency limit of 20 Hz, this would correspond to a diameter of at least 15 m, which at the very least would make for rather unwieldy boom boxes.

Happily, it is not necessary to rely on the loudspeaker itself to avoid low-frequency cancellation. Instead, a smaller driver can be mounted on a baffle, and the baffle can be used to prevent interaction of the front and back radiation. Indeed, a common concept in loudspeaker system theory is that of the *infinite baffle*, which precludes cancellation at any frequency, at the cost of being physically impossible. Even assuming a willingness to forego response below 20 Hz, a baffle of at least 15 m would still be required, which is not much more practical than an unadorned 15 m woofer. This leads to the notion of putting the loudspeaker in a box, or loudspeaker cabinet. (A speaker mounted in a car door or trunk lid is operating in an approximate infinite baffle, although the car interior can be considered a semi-sealed cabinet.)

The simplest arrangement is to make the box completely sealed, so that the rear radiation is trapped within the box, and cannot emerge to cancel the front radiation. Of course, this requires making the box quite rigid, else the rear speaker radiation will simply be conducted through the walls of the box. One problem with sealed-box systems is that the springiness of the air in the cabinet will decrease the natural compliance of the loudspeaker, so that a driver with a respectably low free-air resonance may have a much higher resonance mounted in a sealed box, with correspondingly reduced low-frequency response.

This issue was addressed in 1954 by Edgar Villchur, in a manner which in hindsight seems remarkably straightforward, namely increasing the compliance of the speaker to obtain the desired performance when it was mounted in a sealed box. This approach, referred to as an acoustic suspension design, made the cabinet an inherent part of the design, and is still in use today.

The alternative to the sealed-box loudspeaker enclosure is, logically enough, the vented box, in which at least a portion of the back radiation of the driver is allowed to emerge from the cabinet, often delayed or phase-shifted to better align it with the front radiation. One of the early examples of this approach was the bass reflex design by Albert Thuras of Bell Laboratories, in 1930. Other vented designs have used elaborate internal tubing, internal ductwork (such as the Klipsch folded horns), and or provided mass loading of the exit port via a diaphragm (the *passive radiator*). The goal is to improve efficiency, partially by relieving some of the back pressure behind the driver, but mostly by adding some of the back radiation to the front radiation in a nondestructive fashion, to achieve some increase in output, typically on the order of 3 dB.

More recently, hybrid enclosures have been developed, sometimes called bandpass enclosures, in which a vented or sealed box is used behind the driver and a vented box is used in front of the driver, to provide improved low-frequency efficiency over a somewhat narrow range of frequencies [18.41]. This design is popular in some smaller low-frequency speakers.

Regardless of its other characteristics, a bare loudspeaker cabinet is likely to impart a variety of spurious resonances to the response from sound waves bouncing helter-skelter around the inside of the cabinet. For this reason, loudspeaker cabinets are often lined or stuffed with sound-absorbing material. This alters the overall response as well as damping the resonances, and must be accounted for in the overall design.

Much of the modern modeling of loudspeaker systems is based on the work of Thiele and Small in the 1970s, who established the basic parameters governing the performance of loudspeaker systems, and developed accurate mathematical models incorporating those parameters, allowing rapid and precise mathematical prototyping [18.42].

Still, the implications of achievable, cost-effective drivers and practical cabinet size is that there is precious little that can be done to get usable bass response down to 20 Hz from a small loudspeaker system. It is possible to apply low-level equalization to the signal upstream of the loudspeaker, but this can increase the cost and the power requirements, and will not be effective unless the driver can handle increased signal level without serious distortion. The closed-box design is considered one of the better choices for the use of equalization, as its response roll-off below primary resonance is only 12 dB per octave, which is gentler than most vented systems.

Amazon
EX1013 Page 772

Most of the issues relating to speaker cabinetry pertain to low frequencies. Midrange and high-frequency drivers may also have their backs sealed, but the wavelengths are shorter, and the required resonant frequencies are a lot easier to attain.

The notion of using multiple drivers in a loudspeaker system seems to have originated in 1931 with a two-way system designed by Frederick of Bell Laboratories [18.43]. A three-way system followed within two years. It is simply much easier to design a system to have flat frequency response with multiple drivers, optimized to handle a subset of the audio band, than to try to accomplish the task with a single driver.

Part of the motivation for using multiple drivers is that the loudspeaker, like the microphone, is an inherently spatial entity. Sound radiates from a loudspeaker in all directions, depending on angle and frequency. In an anechoic environment, with a listener situated on the main axis of the loudspeaker, the response on that axis is well defined, and the only one that really matters. In a real room, however, the multidirectional acoustical output of the loudspeaker will reflect off walls and other surfaces, generally producing a complicated multiplicity of arrivals at the listener's ears. There is no easy way to characterize this process, or of the resulting perception, although in the 1970s, Roy Allison made significant strides in characterizing the loudspeaker–room interaction at low frequencies [18.44]. However, it certainly establishes the importance of the loudspeaker spatial response at all angles, and suggests the desirability of a loudspeaker having a consistent radiated frequency response (radiation pattern) at any angle.

This is another area in which multiple drivers can be beneficial. Again, the output of a driver will begin to get directional at a frequency corresponding to a wavelength about twice the diameter of the driver. Assuming consistently wide dispersion is desired, a 30 cm woofer will be useful up to about 500 Hz, a 7.5 cm midrange continuing to 2 kHz and a 2.5 cm tweeter maintaining wide dispersion to at least 6 kHz. These figures may be conservative, in part because controlled decoupling of the speaker cone can result in a progressively smaller radiation area at successively higher frequencies, decreasing the effective diameter and broadening the resulting dispersion. Other approaches to maintaining a desired radiation pattern across frequency include the use of acoustic lenses and phased arrays of multiple drivers within a given frequency range. It is rare for a speaker to have truly constant directionality across its entire pass-band, in part from that fact that most are at least somewhat directional at mid and high frequencies, and, because of the long wavelengths involved, almost unavoidably omnidirectional at low frequencies. (Such long wavelengths may be difficult to localize, enabling the use of satellite systems with only one woofer to support two or more channels.)

The use of two or more frequency-selective drivers generally imposes the need for a crossover network to apportion the signal on the basis of frequency. Not only does this avoid wasting power, but it can be absolutely necessary to avoid overdriving or burning-out a driver. Some care is generally required in the transition frequency region to maintain flat response, and although there are well-regarded canonical topologies, such as the Linkwitz–Riley crossover developed in 1976 [18.45], real-world deviations from ideal driver behavior may engender the use of recursive approximation techniques to optimize system response fully.

The combination of crossover networks, drivers, and room loading will often result in a loudspeaker having an electrical input impedance characteristic that is strongly frequency dependent, and which poses a challenge for a power amplifier.

As is no doubt evident, a loudspeaker system and its associated radiation pattern is a fiendishly complicated device that does not readily lend itself to characterization by traditional audio metrics or specifications. In consideration of such specifications, it is difficult to make definitive statements that fully characterize the sound of a speaker.

The bandwidth of most speakers may approach the ear's 20–20 kHz bandwidth, but few speakers make it all the way, particularly at low frequencies, and bandwidths on the order of, say, 70–14 kHz are far more common, at least in consumer equipment.

The deviation of the frequency response is difficult to talk about meaningfully, but it seems desirable for speakers to be within 1 dB of flat response across their useful bandwidth, at least on axis, and better speakers can approach or achieve this, although at low frequencies the room exerts enough effect as to likely require tuning or experimentation with speaker position to maintain a flat response. The response at other angles should probably be smooth and well controlled, but the exact optimal specification has not been established. Perhaps the ideal speaker would adjust its radiation pattern to match the instruments being reproduced. (But then what is the preferred radiation pattern of a purely electronic instrument, such as a synthesizer?) Although it is necessary to measure the response of a loudspeaker in a room in order to accurately gauge the low-frequency response, a direct narrow-band or sine-wave measure-

Amazon
EX1013 Page 773

ment of the system response in a room at middle and high frequencies is likely to yield a useless, mad scribble of a response, owing to the complex interactions of the echoes at the point of measurement. Fortunately, the perceived response should be a lot smoother than what is measured, in part because the ear differentiates sound arrivals on the basis of direction. Measurement techniques that attempt to address this include the use of directional microphones, anechoic and semi-anechoic measurement, and time-gated response measurement.

Timing precision is another specification that is difficult to pin down with loudspeakers, again partly due to the effects of room echoes. In one respect, most speakers do pretty well in first-arrival timing, since the ear's temporal pre-masking resolution is on the order of 1–2 ms, corresponding to about 30–60 cm of path length, and speaker systems with all drivers mounted on a common front panel will generally be well within this tolerance for inter-driver path-length differences. Still, audible time disparities may result from primary resonances, particularly at low frequencies, and high-order crossovers.

In the area of distortion, loudspeakers also do pretty well, all things considered. At normal listening levels, a properly designed loudspeaker exhibits fairly low nonlinearity, with a fairly gentle limiting characteristic, leading to low IM distortion and fairly low THD, mostly third order, which is likely to be masked by most signals. Most properly operating loudspeakers do not produce drastic amounts of distortion until they reach the limits of their travel. To be sure, some loudspeakers can sound audibly harsh, but this is not necessarily a byproduct of distortion, it may be more related to an uneven response or radiation pattern. One bugaboo of many drivers is cone breakup, wherein the cone no longer acts as a piston. Although this can look unsightly in slow motion, it will not necessarily produce distortion unless there is nonlinearity present, and most speaker cones, being passive structures of paper, plastic, or metal in and of themselves, are not especially nonlinear. While such breakup is likely to affect the response and/or the radiation pattern of the speaker, it may not produce much distortion per se.

Although the basic dynamic loudspeaker has now been in use for over 70 years, and has been greatly refined in that time, it is still the subject of active research, in both design and evaluation. One of the holy grails yet to be achieved is, like the diaphragm-free microphone, the cone-less loudspeaker that more directly converts electrical energy to acoustical energy. Aside from eliminating some troubling mechanical elements and their associated resonances, this might improve the efficiency, which is currently typically around 1% for consumer loudspeakers.

Another long-sought goal is the projection loudspeaker, which can make a sound appear to emanate from a specified remote location. The have been some encouraging results in the use of ultrasonics [18.46] and large-scale speaker arrays [18.47] in this regard.

18.4.4 Amplifiers

Of all the common audio analog devices, electronic amplifiers of all types probably have the easiest time meeting the preferred specifications of a good-quality audio system. This in no small part derives from the fact that they generally contain no mechanical processing stages, and simply have to push a bunch of very lightweight electrons around. Properly designed, an amplifier should have little difficulty achieving a dynamic range approaching 120 dB, nor is it likely to deviate significantly from the preferred specifications for bandwidth, response deviation, phase/timing response, or distortion.

This is despite the fact that most amplification devices, mainly vacuum tubes, transistors, and field-effect transistors (FETs), are not themselves inherently very linear. In the years following de Forest's invention of the triode, considerable and rapid strides were made in not only the refinement of the devices themselves, but in circuit designs which optimized the linearity and response of the complete amplifiers. A major milestone was the invention in 1928 of the negative-feedback amplifier by *Black* of Bell Laboratories [18.9]. This arrangement compares the output signal with the input signal, and to the extent that they may tend to differ, generates an instantaneous correction signal. Successful application of negative feedback carries with it specific requirements on the performance of the raw amplifier, such as maximum phase shift, and some of these requirements were still being uncovered years after the initial development of negative feedback.

Several standard types of amplifiers are commonly found in audio use, each with specific requirements and design challenges, including preamplifiers, line amplifiers, power amplifiers, and radio frequency (RF) and intermediate frequency (IF) amplifiers for use in radio. Preamplifiers, for example, are required to handle the generally very small signal from transducers such as microphones, phonograph cartridges, tape heads, and optical photocells. Linearity is usually not a major problem, because of the small signals involved, but low noise and impedance matching to the transducer are often sig-

Amazon
EX1013 Page 774

nificant design considerations. At the other end of the chain, power amplifiers usually have a fairly easy time with dynamic range, but high required power-handling capability and the need to drive sometimes ill-behaved loudspeaker loads make power amplifier design an art unto itself, with the 1948 Williamson amplifier being just one early example of a notable design [18.48].

The advent of the transistor, by Brattain, Bardeen, and Shockley of Bell Laboratories in 1947, sparked a revolution in active devices, and stands as one of the foremost inventions of the 20th century [18.49, 50]. Operational differences between tubes and transistors were sufficiently great that, to an extent, the discipline of amplifier design had to be reinvented from the ground up, and it was some time before solid-state amplifiers were accepted in some high-end audio systems.

In time, the development of analog integrated circuits gave rise to the multi-transistor operational amplifier, an amazingly versatile device that facilitated cost-effective analog processors of great sophistication and complexity. These days, entire analog audio subsystems can be implemented on monolithic chips with many transistors and associated components.

Special mention should be made of one particularly challenging amplifier variant, the voltage-controlled variable-gain amplifier (VCA). Such a device, basically an analog multiplier, is essential for exercising automatic control of signal level, a common building block in many audio signal processors. Where fixed-gain amplifiers depend on fixed, highly linear passive components such as resistors to maintain operating point and linearity, VCAs require linear active elements, like FETs. In the late 1960s, Barry Blesser designed a novel variable pulse width multiplier, and a few years later, David Blackmer of dbx, Inc. designed a wide-range variable-transconductance VCA, elements of which are still used in commercial VCAs [18.51]. The complexity and cost of VCAs limited their use somewhat, which is perhaps ironic in these days of digital audio, where DSP chips typically perform millions of multiplications per second.

18.4.5 Magnetic and Optical Media

Part of the motivation for the use of magnetic and optical recording on linear media (tape, film) undoubtedly came from the desire to avoid the wear inherent in the playback grooved records. Both media involve the use of narrow apertures to record and play audio striations perpendicular to the motion of a linear medium.

Magnetic recording got its start in 1898, with Valdemar Poulsen recording on steel wire, an especially notable accomplishment considering that the invention of the triode amplifier tube was still several years away. The use of magnetic tape originated with O'Neill's patent of paper tape coated with iron oxide, in 1926, followed by the introduction of plastic tape by BASF in 1935. A serious nonlinearity problem with magnetic tape was largely resolved in 1939 with the development of alternating-current (AC) recording bias. Considerable refinement was achieved during and following World War II, and consumer reel-to-reel tape recording became a reality in the 1950s. There followed a series of formats which expanded the performance envelope and consumer friendliness, most notably the compact cassette and the venerable eight-track cartridge.

Optical recording on film was first investigated around 1901 [18.9], five years before the vacuum tube, and started to receive significant research and development attention starting around 1915, with the work of Arnold and others at Bell Laboratories, followed by a number of efforts by RCA and other groups. Lee De Forest may have come up with the first viable optical sound system, *Phonofilm*, in 1922 [18.52]. By 1928, synchronized optical sound on film was used commercially on Disney's *Steamboat Willie* cartoon.

Both of these formats rely on some sort of device with a narrow aperture to record and playback, and therefore both face a tradeoff between high-frequency response, corresponding to the smallest resolvable feature, and media speed/playback time.

Although there may be no vibrating mechanical elements in either the magnetic system or optical playback, both media exhibit difficulty reaching the full audible bandwidth, or in maintaining ruler-flat response, although, properly designed and adjusted, they can come fairly close. Beyond that, both media bear some similarities to modern record-based reproduction, including the use of fixed equalization, modest, gentle distortion at typical operating levels, a typical dynamic range on the order of 50–70 dB, and both absolute and dynamic speed error sensitivity.

The viability of these formats was significantly enhanced by the invention in 1965 of analog noise reduction by Ray Dolby. Noise-reduction systems usually consist of a compressor/encoder and an expander/decoder. The compressor boosts low-level signals to keep them above the noise floor of the channel, while the expander restores the signals to their proper level, reducing the noise in the process. The Dolby A-type noise-reduction system and follow-on systems exploited a range of solid-state circuit tools, including active filters, precision rectifiers, and lin-

Amazon
EX1013 Page 775

ear FET-based VCAs, to implement a comprehensive solution to the problems faced by noise-reduction systems. The frequency-selective nature of masking by the human ear was addressed by splitting the signal processing into multiple bands, and the problem of potential overload associated with the sudden onset of loud signals was resolved by the use of a novel dual-path limiter. Other well-known NR systems used wide-range bandpass compressors (Telcom C-4), wideband compressors with preemphasis (dbx I and II), sliding shelf filters (Dolby B), cascaded sliding shelf filters (Dolby C), cascaded wideband and variable slope compressors (dbx/MTS TV compressor), and cascaded combined sliding shelf and bandpass compressors (Dolby SR and S). Similar principles would subsequently be adapted for use in digital low-bitrate coders.

18.4.6 Radio

While the medium of analog radio has its own share of unique attributes, it shares with other analog media many of the same sorts of characteristics and limitations, including somewhat limited frequency response and dynamic range, which is partly associated with bandwidth conservation.

Commercial amplitude modulated (AM) radio broadcast kicked off in 1921 when KDKA went on air in Pittsburgh [18.9]. The use of dual-sideband modulation made for simpler receivers at the cost of channel spacing being twice the highest transmitted audio frequency. Although in theory AM radio was capable of supporting wide audio bandwidth, a channel spacing of just 10 kHz was chosen to conserve spectrum space, limiting audio bandwidth to something less than 5 kHz in practice.

Wide-band frequency modulated (FM) radio was developed by Edwin Armstrong in 1933, motivated in part by a desire to reduce the static that was common to AM radio, with commercial broadcast originating in 1941 [18.9, 53]. The dual sidebands of standard FM extend quite a bit farther than the maximum audio frequency supported, with channel spacing of 200 kHz for a typical audio bandwidth of 15 kHz. Dynamic range in excess of 60 dB is possible with mono FM, although this was compromised somewhat by the choice (in the US at least) made by the Federal Communications Commission (FCC) of the stereo multiplex system in 1961 [18.9].

18.5 Digital Audio

The advent of digital audio, perhaps best signified by the introduction of the audio compact disc in 1982 by Sony and Philips, completely revolutionized the world of audio engineering, in effect completely rewriting the rule book on what was readily achievable in mainstream audio systems, not to mention how signals were processed.

The basic notion of digital audio is discreteness. Instead of specifying a signal as a continuous function of time and amplitude, it is expressed as a rapid series of sample values at discrete time intervals, often with integer or finite-precision values. The signal becomes, in effect, a completely specified list of numbers, and as long as no explicit errors are made, can be copied an arbitrary number of generations without any loss or alteration whatsoever. (Error-correction schemes commonly help guard against the occasional error slipping through.)

The mathematical foundations of discrete sampling of continuous functions go back surprisingly far, with at least one claim on behalf of Archimedes, around 250 BC [18.54, 55]. Additional support for sampling theory was provided by *Fourier* in 1822 [18.56, 57] and *Cauchy* in 1841 [18.58, 59]. Modern sampling theory is usually credited to the combined work of *Whittaker* (1915) [18.60], *Nyquist* (1928), *Kotelnikov* (1933), and *Shannon* (1949) [18.61, 62], who provided a rigorous proof [18.63].

The key point of the sampling theorem is that a continuous band-limited signal can be discretely sampled without loss of information, and the original continuous signal subsequently recovered from the sampled version, as long as the sampling frequency is higher than twice the signal bandwidth.

In order to exploit the sampling theorem and digital audio practically, quite a few other pieces of the puzzle had to be supplied. George Boole contributed the binary number system in 1854, which would provide the numerical and logical basis of modern digital processors. Alec Reeves used Boole's binary numbers to represent audio samples, creating the pulse code modulation (PCM) format, in 1937. By the late 1960s, prototype digital audio recorders were operating at companies like Sony, NHK, and Philips [18.58, 59], leading to the 1982 debut of the audio CD.

Amazon
EX1013 Page 776

Some appreciation of the performance of digital audio can be had by considering the traditional audio specifications that analog systems have for so long tried to fully embody.

Bandwidth. Traditionally an uphill battle fighting mechanical resonances and slit-loss effects, successfully meeting a bandwidth specification in the digital domain is simply a question of selecting a high enough sampling rate. Similar principles apply to maintaining the precision of a frequency response specification, usually a flat response: without parasitic or extraneous influences, the response will be ruler practically-flat by default, although it can otherwise be explicitly and predictably altered by application of appropriately designed DSP filters and equalizers.

Distortion. With signal processing implemented as a series of digital additions and multiplications, which are presumably accurate to the least significant bit, nonlinearity and distortion in the digital domain is absent unless, like deviation from flat response, it is explicitly introduced into the processing, to which end it can usually be precisely applied and controlled.

Dynamic range and signal-to-noise ratio. The dynamic range is usually expressed as the logarithmic difference between the maximal signal level and the noise level. By default, the maximal signal capacity of a PCM DSP system as a function of frequency is flat. The noise level is once again a design choice, ergo so is the net dynamic range. In somewhat simplified terms, the noise level of a digital signal process drops by a factor of two (6 dB) for each additional digital bit of precision employed in the representation of the sample values. So, maintaining a desired dynamic range is fundamentally a design choice to employ a sufficient number of bits of precision.

Speed. Once a signal is in the digital domain, there is of course no error or variation in speed unless it is explicitly imposed with appropriate processing.

In short, the attainment of performance that in the analog domain might require considerable design skill, and even then might simply be impractical or impossible, becomes in the digital domain a matter of design choices.

Of course, since most audio signals begin and end as analog vibrations in air (and, so far, correspondingly analog electrical signals associated with input and output transducers), there must be means to convert between analog and digital domains. This is the province of analog-to-digital converters (ADCs) and digital-to-analog converters (DACs). With so much of the overall performance of a digital audio system riding on the performance of these devices, their design has been the subject of intense activity since early experimental digital audio systems. Those early systems were fortunate to muster a precision of just 10-bits per sample, but that has slowly evolved over time to the point that current converters can operate at 24-bits per sample or better in a practical, cost-effective manner. Similarly, the various sources of subtle distortions that might otherwise limit ultimate converter performance have progressively been identified and, on the whole, reduced to the point of insignificance. And, where once speed accuracy of an audio system depended on the skill with which one turned a crank, or perhaps the vagaries of a wind-up clock motor and mechanical speed regulator, the sampling rate of a digital/audio converter is usually controlled by a high-precision electronic timing circuit, such as a quartz oscillator, reducing speed anomalies of all sorts to a level far below that of human perception.

Lest the case for digital audio appear at this point a bit overly rosy, it must be allowed that there are limitations, restrictions and pitfalls; and that it is perfectly possible to produce miserable sound via digital means.

For one thing, there is the matter of the Nyquist frequency, defined as half the sampling frequency and otherwise the highest frequency that can be represented at a particular sampling frequency. In some respects, the Nyquist frequency of a digital audio system corresponds to infinite frequency of an analog system. This introduces certain response-warping effects, such that it may be difficult or impossible to replicate a given analog system response in the digital domain. (Example: the phase shift at the Nyquist frequency can only be 0° or 180°.)

An attempt to represent a frequency greater than the Nyquist frequency in the digital domain will result in a different frequency being generated within the band, usually by reflecting the original frequency value back through the Nyquist frequency, a form of intermodulation distortion. Unless the goal is to produce weird spacey effects, this process is usually not highly desirable, and a low-pass filter is commonly placed ahead of an ADC to prevent that occurrence.

Then there is the issue of performance levels being matters of system architectural choices, rather than the result of possibly elaborate or laborious designs. Such

Amazon
EX1013 Page 777

design choices are not always free, in that they may be tradeoffs with either other system resources, or simply cost more monetarily, which may therefore still have the result of limiting overall performance to something less than idyllic near-perfection.

For example, the dynamic range of the human ear was cited earlier to be on the order of 120 dB or more. If the digital representation of an audio signal yields an approximate dynamic range of 6 dB for each bit used, it would follow that the logical word size to use for digital audio samples would be at least 120 dB/6 dB per bit = 20 bits. Yet the standard chosen for compact discs, as well as many other current generation digital devices, is just 16-bits per sample, corresponding to a best-case dynamic range of only 96 dB. This decision was predicated in part on the performance limitations of analog/digital converters at the time the CD specification was cast, and further from a desire to provide a playing time of 74 min on a 12.7 cm disc. Had a larger word size been chosen, either the disc would have had to be larger, or the playing time reduced.

Does this result in audible impairment of the sound? It can, but it will not necessarily be the case, so it will be evident some hopefully small percentage of the time. For example, there should be no audible alteration if the maximum playback level does not exceed 96 dB SPL. Nor should there be a problem if the playback level is higher, but the room noise is high enough to mask the digital noise. Since the ear is only maximally sensitive in the upper-middle part of the audio band, the elevated thresholds towards the band edges provide additional noise margin. And the presence of signal components will tend to mask digital noise that might otherwise have been heard. So, consistent with practical experience, 16 bit PCM digital audio is free of serious imperfections for most listeners and most source material, most of the time.

18.5.1 Digital Signal Processing

With the practice of representing audio signals in the digital domain as PCM samples comes the need to perform signal processing in the digital domain. To some extent, this required a reinvention of signal processing in terms of discrete numerical sample values, but since the foundations of signal processing were mathematical to begin with, in some respects there was a more direct connection between theory and practice.

The most common signal processing operations – amplification, filtering, and equalization – are mostly implemented with the basic four mathematical functions: addition, subtraction, multiplication, and, perhaps to a lesser extent, division.

More elaborate signal processing functions, such as rectification or level derivation, may involve geometric or transcendental functions such as square roots, cosine, or logarithms, virtually all of which are easier to implement, more direct, more wide-range, and more stable as digital entities than analog alternatives. Even some seemingly simple functions, such as a pure delay, can be challenging to implement well as analog devices, but become more-or-less trivial with DSP.

Clearly, the requirement of audio DSP to perform large numbers of numerical calculations and the ability to perform such calculations with modern digital hardware, including personal computers and specialized DSP chips, whether by accident or design, has been auspiciously symbiotic. Large, elaborate analog breadboards of spaghetti wiring have been replaced with even more elaborate programs of (sometimes) spaghetti C code.

The ability to create and combine a virtually limitless number of new processing functions by simply typing in the code for them, using a computer language with full mathematical support, is an understandably powerful tool to extend signal processing, a little like a bottomless Tinker Toy. The effect is amplified by the ability to nest functions to an almost arbitrary extent, allowing the creation of labyrinthine tree-structured processes. And whereas every block of an analog circuit must be physically implemented, DSP programs usually get by with just one physical copy of each routine, regardless of how many times the function may appear in the block diagram.

With all this processing power available, one may inquire as to what has been applied. Certainly, one function category that has *not* been strongly needed has been aids to correct deficiencies inherent in the recording system. If a clean recording is obtained to begin with, there is often little need to clean it up further. Digital signal processors have, however, been very effective in cleaning up and correcting the imperfections of older, analog recordings, and have been used for noise suppression, speed correction, and even distortion reduction.

The additional processing power afforded by DSP processors has enabled re-engineering of most traditional audio processors, such as equalizers, with greater complexity and precision. The availability of a simple, cheap delay mechanism with equalization has allowed extremely elaborate and faithful reverberators to be employed in both professional and consumer applications. This is one area for which there was arguably never a truly effective analog electronic solution. The best

Amazon
EX1013 Page 778

pre-digital reverberators were electromechanical, using transducers attached to a metal reverberation plate.

In general, the availability of a cheap, low-noise, high-precision multiplication operation has, along with other DSP elements, made it much easier to implement elaborate dynamic systems, once limited by the relative scarcity and high cost of analog VCAs. This has also made it much easier to include specific nonlinearities, where appropriate, usually in the signal analysis of a processor. And the ability to combine multiple processing algorithms into a single DSP program has facilitated multifunctional systems implemented on a single DSP chip.

One of the key building blocks to be brought to bear in a variety of digital signal processors has been a digital model of some or all of the human auditory system. If a processor can be made to hear the same way as humans do, it can make processing decisions that reflect the truly audible elements of a signal, resulting in far more effective processing. A key component of such an approach has been the digital filter bank, essentially a bank of bandpass filters that, in most cases, cover the entire audio band. This component provides a cost-effective analog of the filter-bank-like processing performed by the basilar membrane of the inner ear. A major element of such processing has been the so-called fast Fourier transform (FFT), first devised in 1805 by *Gauss* [18.64–66], then rediscovered in 1965 by *Cooley* and *Tukey* [18.67, 68]. This one algorithm allows full-range filter banks of great complexity, typically implementing hundreds or thousands of bandpass filters, with dramatically less processing power than would otherwise be needed by direct implementation of such filters. The use of this technique does, however, carry with it some compromise, as the bandpass filters of an FFT filter bank all have the same bandwidth, measured in Hertz, and are linearly spaced in center frequency, whereas the filters of the basilar membrane are logarithmically arranged in both bandwidth and center frequency. This discrepancy is sometimes compensated by processing the FFT outputs in groups, or bands, whose width and center frequency are arranged in an approximately logarithmic manner.

The transform-based filter bank is used both as an analysis tool, to present signals in a perceptually relevant domain for subsequent processing, and as a component of the actual signal path. In the latter instance, signals are processed in the transform domain, usually in overlapping blocks of samples, and ultimately restored to time-domain PCM values by performing running inverse transforms. This has elevated transform-based processing to the level of a new category of digital signal processing. In some situations, a variant transform of some sort may be used to provide, for example, logarithmic frequency selectivity, but in many cases such transforms are still based on a core FFT.

A final note on numerical precision: virtually any digital processor will employ finite-word-length symbols for numerical values, be they integers or floating-point numbers. This gives rise to the generation of digital noise, resulting from the round-off error of each arithmetic operation. Since there may be many thousands of such performed in the course of a DSP algorithm, the numerical precision of the processor must be great enough that the accumulated noise remains insignificant. For a 120 dB dynamic range, corresponding to 20 bits of precision, one might prefer at least 24 bit processing (that is, either 24 bit integers or 24 bit mantissas of float values). The actual precision required will depend some on the actual operations performed, with subtraction (or addition of opposite-signed values) especially troublesome, since it can leave a very small residual, potentially exposing the digital noise. As it happens, this is an issue in performing the FFT. As an additional hedge against the buildup of digital noise, many DSP processors employ double-length registers, so accumulated quantities are only rounded back to standard precision when stored back to memory, usually after multiple multiply–accumulate operations.

18.5.2 Audio Coding

The commercial introduction of digital audio in the form of the compact disc brought with it the reincarnation of a tradeoff that had long been a source of compromise in analog audio, namely that between quality and capacity, or equivalently playing time, or bandwidth. Mechanical audio recordings (i.e. records) had started with cylinders that played for a minute or two, and had progressed to long-play records with about 20 minutes per side. Magnetic tape had used media running at speeds like 76 cm per second, and had progressed to cassette tapes with typical playback times of 45 minutes per side, usually aided with noise reduction. Broadcast channels were less flexible, but techniques had at least progressed to the practice of piggy-backing low-energy subcarriers onto to the main signal, e.g. to advance to stereo operation, sometimes accompanied by noise reduction, at least in the case of US multichannel television sound (MTS) audio. If the use of digital audio for recording and broadcast was to be adopted for media beyond the CD, the need to

Amazon
EX1013 Page 779

reduce the number of bits required was apparent. This has given rise to one of the premier pursuits of digital development since that time: the use of audio coding to convey a signal faithfully using the smallest possible number of bits.

The data rate of an audio CD derives from the use of two channels, a sampling rate of 44 100 Hz, and 16 bit PCM samples, a total of 1.4 megabits per second. This is greater than can be accommodated on a standard 35 mm movie print, or a standard cassette tape, or an FM radio channel. (It is also more than will fit on a standard long-play record, but since that medium is not especially compatible with digital storage in the first place, it is a moot point.) Even the CD can only accommodate the two channels; a 5.1-channel PCM format would reduce the playing time to something under 25 min.

So the need to reduce the data rate without sacrificing quality was readily apparent, which motivated interest and activity in the development of low-bitrate audio coders, and made practical such applications as digital cinema soundtracks, portable music players, and satellite radio. It is perhaps not self-evident that the data rate can be reduced in the first place, but for starters, one could simply use smaller PCM words. This would, of course, raise the effective noise floor, which in turn could be addressed by adapting an analog noise-reduction system as an associated DSP algorithm. However, the resulting reduction in data rate would still be relatively modest, as the compression ratios of noise-reduction systems never exceeded a range of about 2–3:1, to avoid mistracking problems, and this is less data compression than is required by many coding applications.

Overall, two general classes of techniques have been exploited in digital audio coders, lossless and lossy, or perceptual, coding. They share in common the arrangement of an encoder and a decoder which, instead of just conveying PCM samples, employ a common vocabulary or protocol for describing and conveying the audio at an elevated level of abstraction. This protocol is sometimes referred to as a bit-stream syntax. Although the signal connecting encoder and decoder is indeed usually just a stream of bits, the meaning and significance of a given bit will depend on the context in which it appears. This represents an additional advantage of digital audio systems over analog systems: a bit stream can consist of interleaved sequences of diverse information packets. Some of these will generally be representations of some derived form of the audio, and the rest will be descriptive information about the signal, referred to as side-chain information.

Lossless Audio Coding

The goal of lossless coding is to convey digital information in a more compact form while reliably recovering the original data in the decoder with bit-for-bit accuracy. Obviously, no psychoacoustic or perceptual considerations need enter into the picture.

The key to the ability to losslessly compress digital data is redundancy. Any time there is a pattern or common element to a block of data, it may be possible to express it in a more compact form by conveying the essence of the pattern along with the deviations from the pattern, such that the original data may be reconstructed. Of course, in order for this to happen, the bit-stream syntax must support a compact means of specifying the necessary range of patterns to the decoder.

The number of potentially useful patterns is actually fairly large, and considerable ingenuity has gone into the development of lossless coding algorithms which can recognize and concisely encapsulate the details of redundant information patterns. Some of the more common data structures that can be losslessly compressed are listed below.

1. A repeating sample value: rather than send the same value multiple times, the sample can be sent just once, along with a replication factor. This technique is sometimes referred to as *run-length coding*.
2. Similar values: a variant of run-length coding, this type of pattern can be compressed by sending the average value, followed by concise representations of the difference between each sample value and the average. Alternatively, differential coding may be used, in which the difference between successive samples is coded.
3. Repeating sinusoids: for an ongoing sinusoidal signal, the PCM samples can be replaced with parameters indicating the amplitude, frequency, phase, and length of the sample sequence. In general form, this is sometimes referred to as *linear prediction*. If there are small deviations, they may be sent as concise difference values.
4. Small values: a sequence of samples, each with one or more leading zeros, may be sent with the zeros omitted, along with a factor indicating how many leading zeros were suppressed. (Note: negative values expressed in two's-complement form will have leading ones, but the same factor should otherwise apply.) This is known as *block floating point* representation.
5. Commonly used values: sample values which appear more often than the average – like the letters 'e'

Amazon
EX1013 Page 780

and 's' in English text – may be assigned special, short transmission symbols to reduce the average data rate. Representative techniques which employ this include arithmetic coding and Huffman coding.
6. Interchannel difference coding: if two or more channels are used and are sufficiently similar, data-rate savings may be attained by sending sum and difference information, or a variant thereof. The expectation is that the difference values will be small, and can be coded in fewer bits using technique (4), above. In the case of stereo content, this approach is sometimes referred to as mid/side (M/S) coding.

In addition to the techniques described above (and others), a lossless coder may either directly code time-domain PCM samples, or may instead use running transforms and operate in the frequency domain. In the latter case, a specialized bit-exact transform may be used to guarantee that the final output is an exact clone of the input.

A fully realized lossless coder is likely to use a combination of techniques like those described above, requiring a flexible and perhaps elaborate bit stream syntax. The process of selecting which coding options will be brought to bear on a given block of samples is often not readily apparent from inspection or simple calculation, so the encoder may be required to try an exhaustive search of all possibilities for each block in order to select the best option. Needless to say, this can make for a rather slow-running encoder, but usually does not affect the speed of the decoder.

As may be evident, the effectiveness of a lossless coder is very much signal dependent: some signals can be compressed considerably more than others, and some, such as full-level white noise, basically cannot be losslessly compressed at all. Typical compression ratios run in the range of 2:1 to 3:1. Because lossless compression is inherently a variable bitrate (VBR) coding technology, it is not well suited to real-time streaming applications, such as broadcast or fixed-speed digital tape.

One weakness of lossless coding, and indeed of lossy coding as well, is that an error of even a single bit (in, say, one of the control codes specifying which method to use to unpack a block of data) can cause the loss of significant audio data. Some care must be used in the design of the bitstream to ensure that re-synchronization of the decoder in the face of data errors is reliably possible within an acceptably short period. More generally, the complexity and low error tolerance of almost any data-reduction algorithm raises the regrettable possibility that digital records, audio and otherwise, may be lost to future generations unless care is taken to avoid data errors, and to ensure that decoding algorithms are carefully documented and preserved in full detail. This is especially critical in the case of proprietary algorithms owned and maintained by commercial entities, whose long-term future operation may not be assured.

Perceptual/Lossy Audio Coding

While one might wish that lossless coding be used exclusively for digital audio compression, its limited performance and variable bitrate render it inappropriate for a large number of coding applications. For more-aggressive compression and acceptable performance within the constraint of constant bit rate, one must resort to the use of perceptual audio coding.

The goal of perceptual audio coding is to preserve and convey all of the perceptual aspects of audio content, without regard to preserving the signal literally, usually employing the fewest bits possible. Recall that the data rate for a standard audio CD was about 1.4 Mbit/s. As this is written, the lowest data rate in commercial use for conveying high-quality stereo audio is no higher than 48 Kbits/s, used by satellite radio and for Internet streaming, a data-compression ratio on the order of 30:1 or better.

How is this accomplished? For starters, the use of lossless coding is not retained and incorporated in overall lossy algorithms. Every packet of information generated and reduced to its ultimate compactness by any part of a perceptual coder is a candidate for further reduction in size if one or more lossless techniques can be brought to bear effectively.

Beyond that, the keys to high-efficiency perceptual coding are to:

1. suppress any audio components not contributing to the ultimate perception,
2. ensure that all the audible components are conveyed with no alteration in perceived quality, while using the minimum possible number of bits, and
3. employ, as much as possible, high-level abstractions to describe audio signals in purely perceptual terms, while retaining enough precision and detail to allow the decoder to accurately reconstitute a signal indistinguishable from the original encoder input signal.

One fundamental building block that has almost always been part of the foundation of perceptual coders is a digital filter bank. This, of course, mirrors the functionality of the basilar membrane; and the frequency-domain out-

Amazon
EX1013 Page 781

put signals are most commonly used for both the analysis sections of the coder and the signal path. Indeed, the fundamental data conveyed by most perceptual coders is not PCM time-domain data, but quantized frequency-domain information from which the decoder eventually produces PCM output by way of a synthesis filter bank. The most commonly used digital filter bank is a variant of the FFT called the time-domain alias cancellation (TDAC) transform, developed by Princen and Bradley in 1986 [18.69, 70]. This transform has the highly desirable property of being critically sampled, meaning that it produces exactly the same number of output frequency-domain samples as there are input PCM samples. There is also a companion inverse TDAC transform for producing output PCM from a set of decoded frequency-domain input samples.

In operation, PCM samples input to the encoder are divided into regular blocks of some predetermined length, and each block is transformed to the frequency domain, conveyed to the decoder, reconstituted as a block of PCM samples, and the successive blocks strung together, often in an overlapping manner, to recover the final outputs. This process is sometimes described as block-oriented processing using running transforms. The practice of coding an entire block of samples as a single entity facilitates data rates of less than one bit per original audio sample. One downside of block processing is that events which occupy a small fraction of a block, like a sharp transient, may become smeared across a range of frequencies, consuming a large number of bits. A number of techniques have evolved to deal with this situation, most commonly the use of block switching, wherein such short events are detected (by, for example, a transient detector routine), and the block size is temporarily shortened to more closely isolate the transient within the shortened block.

The filter-bank output is of direct use in pursuit of the first two perceptual coding techniques listed above, principally by deriving the magnitude of the frequency-domain signals as a function of frequency to obtain a discrete approximation to the power spectral density (PSD) of the signal block. This in turn is processed by a routine implementing a perceptual model of human hearing to derive a masking curve, which specifies the threshold of hearing as a function of frequency for that signal block. Any spectral component falling below the masking curve will be inaudible, so need not be coded, as per the first listed coding technique. The remaining spectral components must be preserved if audible alteration is to be avoided, but the quantization precision is only that which is required to render the level of the quantization noise below the masking curve. Instead of the 120 dB/20 bit range one might require for PCM audio, the instantaneous masking range is more often on the order of 20–30 dB, about 4–5 bits per sample, assuming 6 dB per bit SNR. Thus, between suppression of inaudible components and dynamic quantization of audible components, the data rate can be expected to be reduced to something less than 4 bits per sample. Of course, an effective bit-stream syntax protocol must be devised to signal these conditions efficiently to the decoder, and considerable ingenuity has been brought to bear on that issue to ensure the requirement is met.

Notable lossy audio coders based on these principles include AC-3, MP3, DTS, WMA, Ogg, and AAC. These have been instrumental enablers of such technologies as portable music players, DVD's, digital soundtracks on 35 mm film, and satellite radio.

It should be noted that perceptual coding is much more compatible with constant-bit-rate operation than is lossless coding, for as the complexity of a signal increases, which might otherwise increase the required data rate, the masking afforded by that signal also increases, reducing the average quantization accuracy required, thereby holding the required data rate to a more nearly constant rate. In effect, the human auditory system can only absorb so much information per unit time, so as long as the coder accurately models that behavior, the required data rate should be largely constant. Of course, very simple signals, like silence, are not likely to require the same data rate as more complex signals, in which case a constant-bitrate coder may simply use far more than the minimum required data rate in order to maintain a constant transmission rate. Some coders can defer use of bits in the presence of simple signals, saving them in a *bit bucket* until needed by a more complex signal.

The third listed technique used in perceptual coding, the application of higher-level abstractions to describe the signal, is much more general and open-ended; and is still an area of active investigation. It can be something as simple as using a single PSD curve to approximate the actual PSD curves of several successive transform blocks, or using decoder-generated noise in place of actually transmitting noise-like signals. More abstract approaches may fall under the heading of *parametric coding*.

One notable example of parametric coding is *bandwidth extension*, in which the entire high-frequency end of the spectrum is suppressed by the encoder, and is reconstituted by the decoder from analysis of the harmonic

Amazon
EX1013 Page 782

content of rest of the signal spectrum. This technique alone can reduce the data rate by 30% or more.

Another widely pursued approach to parametric coding is to segregate signal components into categories, such as sinusoidal, noise-like, or transient, then convey numerical descriptions of each to a synthesizing decoder. Although this is arguably far closer to a true perceptually grounded coder than a simple transform-based coder, the complexity of the descriptors needed to accurately portray the full range of possible signals has so far made it difficult to reduce the data rate required by more conventional coders dramatically.

Spatial Audio Coding

Although strictly speaking it may be a subset of perceptual audio coding, spatial audio coding (SAC) is a sufficiently independent approach to justify separate examination. SAC can be used to improve the transmission efficiency of multichannel content, typically either stereo or 5.1-channel surround material.

The approach involves coding the ensemble of channels as a single entity to achieve data savings that would be unavailable if each channel were to be coded individually. An element common of many SAC algorithms is to combine the source channels additively to a smaller number of channels, perhaps as few as a single channel, from which the decoded output channels are eventually reconstructed. The data rate is thereby reduced by the ratio of the number of input channels to the number of down-mixed channels, minus any side-chain information needed by the decoder to reconstruct the output channels.

Probably the simplest approach to SAC is to downmix the input channels to mono in the encoder and then simply distribute the common signal equally to the output channels. If this is done to the full-bandwidth signal, the result will simply be monaural sound reproduced from multiple speakers, which is not likely to be a very satisfactory multichannel experience, but if limited to a subset of the audio range, say just high frequencies, the ear will tend to associate the spread mono content of the reproduced channels with the direction of the lower frequency, discrete content, and the alteration will be less noticeable.

A slightly more elaborate approach is to send channel-by-channel scale factors, representing relative signal amplitude in each input channel, which are subsequently used by the decoder to distribute signal components to each output channel. Sometimes referred to as either intensity stereo or channel coupling, the technique can be surprisingly effective, especially at high frequencies when individual scale factors are sent on something like a critical band frequency basis.

More recent spatial coding systems have augmented the use of amplitude coupling with transient flags and information about interchannel phase and coherence, allowing extension to progressively lower frequencies, with commensurately greater amounts of data rate reduction.

Although in most cases the spatial information about individual channels is sent as side-chain information comprising part of the total transmitted bit stream, there has been some success in encoding the information in the signal characteristics of the actual transmitted audio, which can reduce the net data rate.

Spatial audio coding remains at this time an area of very active investigation.

18.6 Complete Audio Systems

So far, this discussion of audio and electroacoustics has focused on the task of capturing, transmitting, and reproducing one or more individual channels of audio as accurately as possible. However, the goal of sound reproduction is not strictly to convey some designated number of channels, but to faithfully reproduce a full acoustic event, be it live, synthesized, or some combination thereof, to an audience that, at any given venue, may range in number from one to many hundreds or more.

The use of individual channels is mandated by the available technology. Microphones and loudspeakers are basically point-in-space devices, and most audio media and processors are designed to handle individual channels. But sound is, by nature, a three-dimensional entity, not counting time and frequency, and a few isolated channels can at best represent a sparse spatial sampling of a full acoustic sound field.

So the question is raised as to just how effective are complete audio systems, from recording to reproduction, at least in perceptual, psychoacoustic terms.

Any attempt to perform such an evaluation is immediately faced with the conundrum as to just what should be the ultimate goal of sound reproduction. Should we be trying to transport the listener to the performance site ("we are there"), or is the goal to transport the performers to the listener's living room ("they are here").

Amazon
EX1013 Page 783

The answer perhaps depends on the nature of the source material, as well as listener preference. For situational acoustics, from walking down a busy street to riding in a submarine, it would seem logical to go with "we are there". This is most commonly the case for, say, cinema soundtracks, making a good case for surround-sound systems in commercial and home theaters. A similar case can be made for symphonic music, since the concert hall is such an integral part of the experience. However, a small ensemble or a solo artist might be more compelling if holographically projected into the listening room. Given unlimited wishes, one might hope to have both, with the ability choose which option according to one's predilections. In actual practice, we might be well served if either one is possible. But in any case, it is perhaps worth keeping in mind in evaluating the effectiveness of complete audio systems.

18.6.1 Monaural

For the first half century of recorded sound, monaural was all there was. In that time, considerable expertise was developed in microphone techniques for capturing sound with a viable balance of direct and ambient content. In this era of multichannel audio, it is easy to dismiss mono for failing to support any formal spatial audio cues. Still, there are times when a single channel should, at least in theory, be adequate for the task, such as a solo performer or other single, localized source, in a dry (relatively echo-free) ambient environment, or "here". Even in that case, however, the reproducing loudspeaker is unlikely to have the same spatial radiation pattern as the original source, a problem common to many sound formats to one degree or another, so that even with good frequency response and low distortion, the sound may be recognizable as emanating from a loudspeaker, rather than being a clone of the original performance. To its credit, mono does not exhibit the double-arrival problem of phantom images common to stereo and other multichannel formats (below), arguably resulting in a purer, less time-smeared sound. It is also gratifyingly free of sweet-spot dependence: one can sit pretty much anywhere in a listening room without worrying about impairment of imaging. And historically at least, mono was a big improvement over nothing at all.

18.6.2 Stereo

Two-channel stereophonic sound has been a commercial reality for nearly half a century, and shows little sign of being rendered obsolete anytime soon. The intuitive attraction of stereo is self-evident: we have two ears, and stereo provides two channels. On the surface, this is an apparently compelling argument. The result should perfect sound reproduction, yes? Well, no, although stereo did pretty much render mono obsolete for most media within a few years of its commercial deployment.

Stereo provides the smallest possible increment in spatial fidelity over mono: a horizontal line spread before the listener. But this is likely the most important single improvement one could hope to make, as the human ear is most sensitive to front-center horizontal position. In addition, performing ensembles are most often arrayed in such a manner. The arrangement also allows rendering, where appropriate, of frontal ambiance in a manner more natural than can be achieved with a single audio channel.

An important element of stereo reproduction is the creation of virtual images. These are sonic images that appear suspended in space between the two speakers, and they undoubtedly add considerably to the viability of the format. Virtual images depend for their existence on the primary horizontal localization cues: interaural time differences and interaural amplitude differences. The former cue is especially critical to the creation of virtual images, as the total range of the cue is only about $\pm 700\,\mu s$, corresponding to a horizontal displacement of the listener of only about a foot. This rather undermines the viability of stereo virtual imaging somewhat, and renders the format decidedly sweet-spot dependent. The situation is partially compensated by the flexibility of the human auditory system, and its apparent ability to use dynamically whatever cues appear to be providing reliable localization information. A listener a little off the center line may still experience some degree of virtual imaging by virtue of the interaural amplitude cue, which may not degrade as rapidly as the interaural time delay cue. Indeed, in the course of mixing a multichannel source to a stereo track, a recording engineer will typically position a virtual source between the speakers using simple amplitude panning, with zero interaural time delay, and yet the ear generally accepts this.

But sensitivity to listener position is not the only issue relating to virtual images. There is also that double-arrival problem. The direct sound from almost any real sound source will generally arrive at each ear once, and that is it. But a virtual image from a stereo speaker array will result in two direct arrivals at each ear, one from each speaker. Assuming a centered listener, the left ear will hear from the left speaker at about the same time that the right ear hears the first arrival from the right speaker. Then the sound from each speaker will

Amazon
EX1013 Page 784

cross the head, in opposite directions, and the left ear will hear the right speaker arrival, and vice versa, a process sometimes referred to as loudspeaker crosstalk. The time delay between arrivals is typically on the order of 250 μs, resulting in a small, subtle, but audible notch in the resulting response around 2 kHz. Some temporal smearing of sharp transients may also be evident.

A final difficulty with virtual imaging of stereo systems is the relative inability to position virtual images outside the range of the speakers. This is another consequence of loudspeaker crosstalk, limiting the maximum interaural time difference to that corresponding to the position of the speaker, which again is about 250 μs, and similarly limiting the maximum interaural amplitude difference. One way around this would be to put the speakers in different rooms and sit between them, one ear in each room, with a custom-fitted baffle doorway to block any crosstalk, which is a little impractical. Another approach is to employ a crosstalk cancellation (virtual-speaker) system, to cancel the second arrival at each ear, but such systems are themselves very sweet-spot dependent, and even then usually limited to creating images on the horizontal plane. The most direct way to provide sound from additional directions is to position additional speakers around the listener, including above and possibly below, assuming additional source channels are available to feed them, but that is no longer a stereo sound system.

18.6.3 Binaural

Like stereophonic sound, binaural makes use of just two channels, but it uses headphones instead of loudspeakers to avoid the crosstalk problem. On paper, this system seems like it should be even more perfect than stereo. However, the effect is often to place the sound images within the head, with little or no front/back differentiation.

Part of the problem appears to be that the presentation does not get altered when the listener's head moves, so that the image follows the head movement, which is rather unnatural. There has been some work exploring the use of tracking headphones, which alter the signal in response to head motion, but these have not yet been completely successful, and require more than two channels to be available.

Another issue with binaural reproduction seems to be the lack of pinna cues, those alterations in timing and spectra that are imparted by the pinnae to an arriving sound as a function of direction. Efforts to impart pinna cues electronically have had some success, but they tend to be as individual as fingerprints, making it difficult to have a single binaural track which works for everyone. There has also been some success in producing a universal binaural track by imparting a direction-dependent room reverberation characteristic to the signals, although some purists object to the addition, and the effect is still somewhat listener dependent.

One experimental system that mitigates some of the shortcomings of two-channel binaural is four-channel binaural. This employs two dummy heads or equivalently arranged directional microphones, with one head/microphone-pair in front of the other, preferably with an isolating baffle separating them. The resulting four-channel recording is played over four speakers, preferably placed close to the listener in fixed positions around the listener's head, corresponding to the microphone positions, possibly with a left–right isolating panel or crosstalk-canceling circuits to minimize crosstalk. Because the front–back pair each preserve ITD and IAD, the left–right space is properly preserved, while front–back differentiation and externalization are provided by using separate front and back external drivers. Dynamic IAD and ITD cues with head motion are also at least crudely preserved by the use of front and back channels. The resulting presentation can exhibit good spatial fidelity.

Although it may never be suitable for general audience presentation, the binaural format has compelling qualities for the single listener, and remains an area of active investigation.

18.6.4 Ambisonics

This system, devised by Michael Gerzon, approximates the sound field in the vicinity of a listener using an array of precisely positioned loudspeakers, It is sweet-spot dependent, so not amenable to audience presentation, but can provide compellingly good spatial reproduction.

18.6.5 5.1-Channel Surround

The 5.1-channel format is a well-thought-out response to the shortcomings of stereo. The format employs three front speakers, two surround speakers placed to the side and/or back of the listening room, and the 0.1 low-frequency channel that only occupies a small fraction of the total bandwidth.

The addition of the center-front speaker anchors the front left–right spread, significantly mitigating the problems stereo has with virtual images, even with audience presentation. The surround channels provide sound more

Amazon
EX1013 Page 785

to the side than can be achieved with the front speakers (or stereo), exercising the full range of the ITD/IAD sensitivity of the ear, and providing the option of a far more immersive experience than can be attained with stereo.

The 5.1 format with high-quality source material provides the most accurate overall audio reproduction of any widely used format to date, within either the "here" or "there" reproduction paradigms.

18.7 Appraisal and Speculation

We have been witness, from the crude but noble origins of sound reproduction, to an astonishing march of progress that has, by degrees, rolled down the problems and imperfections of sound reproduction, reaching a notably high level of performance, when considered in basic perceptual terms.

A cynic might note that, fifty years ago, children played little yellow plastic 78 RPM 17.78 cm discs on phonographs that had 12.7 cm loudspeakers, and now, after fifty years of dramatic, hard-won progress, many people play little 12.7 cm plastic (compact) discs on boom boxes with 12.7 cm loudspeakers. Yes, the fidelity is better, as is the playing time, but have we really come that far?

The apparent answer from all the above is that a high-quality system of modern vintage can now satisfy a major percentage of the electroacoustic and spatial requirements implied by the capabilities of the human auditory system.

This does not mean that audio engineers call all just pack up and go home just yet. Indeed, the quantities of audio research papers and patents published each year show no sign of abating, giving rise to the speculation that further meaningful improvements may yet be attained.

One can certainly wish for a few more channels, to fill in the remaining gaps and perhaps even provide explicit coverage above and below the listener. The issue in that regard at this point is more of practicality than of theory. There have been some experiments with the addition of height or ceiling (*voice of God*) channels, and the specifications for digital cinema systems allow for the use of over a dozen channels.

One can also hope for enhanced ability to project sound into a room, rather than having it just arrive from the walls. That is another area of active investigation. One approach, referred to as wave field synthesis [18.71] has been to approximate a complete sound field by reproducing the sound-pressure profile around a periphery, and relying on the Huygens principle [18.72] to recreate the complete interior sound field. This requires large number of loudspeakers (192 in one experimental setup) to avoid spatial aliasing effects.

The other well-known approach to generating interior sounds is to use an ultrasonic audio projector, which projects acoustic signals above the audio band which heterodyne to produce audible sound which appears to come from a remote location [18.73].

Another problem that has only partly been solved is that of sweet-spot sensitivity. This has long been an issue with cinema surround-sound systems, but is no less challenging for automotive surround systems, where there are often no centrally located seats and speaker locations may be less than optimal. The issues go beyond balanced sound to trying to provide a coherent, immersive experience at all listening positions.

Indeed, each new venue for audio seems to come with its own idiosyncratic problems, leading to active investigation of how to make compelling presentations for portable music players, cell phones, and interactive media such as game consoles.

It is also expected that there will be continuing refinement of binaural techniques for personal listening.

Ultimately there may come a wholesale revolution in audio techniques, perhaps with laser transducers replacing loudspeakers and microphones, along with holographic recording of entire sound fields. For personal listening, your music player might interface directly to your brain, bypassing the ear altogether, allowing otherwise deaf people to hear, and getting rid of that troublesome 3 kHz peak in the ear canal once and for all.

It should make for some interesting vibes.

References

18.1 N. Aujoulat: *The Cave of Lascaux* (Ministry of Culture and Communication, Paris August 2005), http://www.culture.gouv.fr/culture/arcnat/lascaux/en/

Amazon
EX1013 Page 786

18.2 J. Sterne: *The Audible Past* (Duke Univ. Press, Durham 2003)
18.3 Thomas Edison (Wikipedia Aug. 2005) http://en.wikipedia.org/wiki/Thomas_Edison
18.4 J.H. Lienhard: *The Telegraph, Engines of Our Ingenuity* (Univ. Houston, Houston 2004), http://www.uh.edu/engines/epi15.htm
18.5 P. Ament: The Electric Battery (The Great Idea Finder, Mar. 2005) http://www.ideafinder.com/history/inventions/story066.htm
18.6 Phonograph (Wikipedia, Aug. 2005) http://en.wikipedia.org/wiki/Phonograph#History
18.7 D. Marty: The History of the Phonautograph (Jan. 2004) http://www.phonautograph.com/
18.8 Emile Berliner and the Birth of the Recording Industry, Library of Congress, Motion Picture, Broadcasting and Recorded Sound Division (April, 2002) http://memory.loc.gov/ammem/berlhtml/berlhome.html
18.9 J. Bruck, A. Grundy, I. Joel: *An Audio Timeline* (Audio Engineering Society, New York 1997), http://www.aes.org/aeshc/docs/audio.history.timeline.html
18.10 M.F. Davis: History of spatial coding, J. Audio Eng. Soc. **51**(6), 554–569 (2003)
18.11 NIH: *Hearing Loss and Older Adults, NIH Publication No. 01-4913, National Institute on Deafness and Other Communications Disorders* (National Inst. Health, Bethesda 2002), http://www.nidcd.nih.gov/health/hearing/older.asp
18.12 T. Weber: *Equal Loudness* (WeberVST, Kokomo 2005), http://www.webervst.com/fm.htm
18.13 E. Terhardt: *Pitch Perception* (Technical Univ. Munich, Munich 2000), http://www.mmk.e-technik.tu-muenchen.de/persons/ter/top/pitch.html
18.14 B. Truax: *Sound pressure level [SPL], Handbook for Acoustic Ecology, Simon Fraser University, Vancouver* (Cambridge Street Publ., Burnaby 1999), http://www2.sfu.ca/sonic-studio/handbook/Sound_Pressure_Level.html
18.15 B. Truax: *Phon, Handbook for Acoustic Ecology, Simon Fraser University, Vancouver* (Cambridge Street Publ., Burnaby 1999), http://www2.sfu.ca/sonic-studio/handbook/Phon.html
18.16 B.G. Crockett, A. Seefeldt, M. Smithers: *A new loudness measurement algorithm, AES 117th Convention Preprint 6236* (Audio Engineering Society, New York 2004)
18.17 L.H. Carney: *Studies of Information Coding in the Auditory Nerve* (Syracuse Univ., Syracuse 2005)
18.18 E.D. Haritaoglu: *Temporal masking* (University of Maryland, College Park 1997), http://www.umiacs.umd.edu/ desin/Speech1/node11.html
18.19 K.M. Steele: *Binaural processing* (Appalachian State Univ., Boone April 2003), http://www.acs.appstate.edu/kms/classes/psy3203/SoundLocalize/intensity.jpg
18.20 J. Blauert: *Spatial Hearing* (MIT Press, Cambridge 1983)
18.21 R. Kline: Harold Black and the negative-feedback amplifier, IEEE Control Syst. Mag. **13**(4), 82–85 (1993), DOI: 10.1109/37.229565
18.22 N. Moffatt: *Wow and Flutter* (Vinyl Record Collectors, United Kingdom 2004), http://www.vinylrecordscollector.co.uk/wow.html
18.23 B. Paquette: *History of the microphone* (Belgium 2004), http://users.pandora.be/oldmicrophones/microphone_history.htm
18.24 J. Strong: *Understanding microphone types, Pro Tools for Dummies* (Wiley, New York 2005), http://www.dummies.com/WileyCDA/DummiesArticle/id-2509.html
18.25 P.V. Murphy: *Carbon Microphone* (Concord University, Athens 2003), http://students.concord.edu/murphypv/images/35107.gif
18.26 D.A. Bohn: *Ribbon Microphones* (Rane Corp., Mukilteo 2005), http://www.rane.com/par-r.html
18.27 H. Robjohns: *A brief history of microphones, Rycote Microphone Data Book* (Rycote Microphone Windshields, Stroud 2001), http://www.microphone-data.com/pdfs/History.pdf (Rycote Microphone Windshields, Stroud 2003)
18.28 G. M. Sessler, J. E. West: Electroacoustic Transducer Electret Microphone, US Patent Number 3118022 (1964) http://www.invent.org/hall_of_fame/132.html (National Inventors Hall of Fame, 2002)
18.29 P.V. Murphy: *Piezoelectric Microphone* (Concord University, Athens 2003), http://students.concord.edu/murphypv/images/35108.gif
18.30 D. Stewart: *Understanding microphone self noise* (ProSoundWeb, Niles 2004), http://www.prosoundweb.com/studio/sw/micnoise.php
18.31 A.G.H. van der Donk, P.R. Scheeper, P. Bergveld: Amplitude-modulated electro-mechanical feedback system for silicon condenser microphones, J. Micromech. Microeng. **2**, 211–214 (1992), http://www.iop.org/EJ/abstract/0960-1317/2/3/024
18.32 A. Oja: *Capacitive Sensors and their readout electronics, Helsinki Univ. Technology, Physics Laboratory* (Helsinki Univ. Technology, Espoo 2005), http://www.fyslab.hut.fi/kurssit/Tfy3.480/Spring2004/Jan28_Oja/Microsensing_2004_Oja2.pdf
18.33 J.N. Caron: *Gas-coupled laser acoustic detection* (Quarktet, Lanham 2005), http://www.quarktet.com/GCLAD.html
18.34 H.M. Ladak: *Anatomy of the middle ear, Finite-Element Modelling Of Middle- Ear Prostheses in Cat, Master of Engineering Thesis* (McGill University, Montreal 1993), Chap. 2 http://audilab.bmed.mcgill.ca/ funnell/AudiLab/ladak_master/chapter2.htm

Amazon
EX1013 Page 787

18.35 Microphone Types, The Way of the Classic CD (Musikproduktion Dabringhaus und Grimm, Detmold 2005), http://www.mdg.de/mikroe.htm

18.36 G. Carol: *Tonearm/Cartridge Capability* (Galen Carol Audio, San Antonio 2005), http://www.gcaudio.com/resources/howtos/tonearmcartridge.html

18.37 J.G. Holt: *Down With Dynagroove, Stereophile, December 1964* (Primedia Magazines, New York 1964), http://www.stereophile.com/asweseeit/95/

18.38 D. Krotz: *From Top Quarks to the Blues, Berkeley Lab physicists develop way to digitally restore and preserve audio recordings, Berkeley Lab. Res. News 4/16/2004* (Lawrence Berkeley National Lab., Berkeley 2004), http://www.lbl.gov/Science-Articles/Archive/Phys-quarks-to-blues.html

18.39 S.E. Schoenherr: *Loudspeaker history, Recording Technology History* (Univ. San Diego, San Diego July 2005), http://history.sandiego.edu/gen/recording/loudspeaker.html

18.40 C.W. Rice, E.W. Kellogg: Notes on the development of a new type of hornless loudspeaker, Trans. Am. Inst. El. Eng. **44**, 461–475 (1925)

18.41 AudioVideo 101: *Speakers Enclosures* (AudioVideo 101, Frisco 2005), http://www.audiovideo101.com/learn/knowit/speakers/speakers16.asp

18.42 Eminence: *Understanding loudspeaker data* (Eminence Loudspeaker Corp., Eminence 2005), http://editweb.iglou.com/eminence/eminence/pages/params02/params.htm

18.43 M. Sleiman: *Speaker Engineering Analysis* (Northern Illinois University, DeKalb 2005), www.students.niu.edu/ z109728/UEET101third%20draft.doc

18.44 Allison Acoustics: *Company History* (Allison Acoustics LLC, Danville 2005), http://www.allisonacoustics.com/history.html

18.45 D. Bohn: A fourth-order state-variable filter for Linkwitz-Riley active crossover designs, J. Audio Eng. Soc. Abstr. **31**, 960 (1983), preprint 2011

18.46 W. Norris: *Overview of hypersonic technology* (American Technology Corp., San Diego 2005), http://www.atcsd.com/pdf/HSSdatasheet.pdf

18.47 T. Sporer: *Conveying sonic spaces* (Fraunhofer Elektronische Medientechnologie, Ilmenau 2005), http://www.emt.iis.fhg.de/projects/carrouso/

18.48 B. DePalma: Analog Audio Power Amplifier Design (January, 1997) http://depalma.pair.com/Analog/analog.html

18.49 E. Augenbraun: *The Transistor, Public Broadcasting System* (ScienCentral, The American Institute of Physics, New York 1999), http://www.pbs.org/transistor (July 2005)

18.50 S. Portz: *Who Invented The Transistor* (PhysLink, Long Beach 2005), http://www.physlink.com/Education/AskExperts/ae414.cfm

18.51 THAT Corp.: *A brief history of VCA development* (THAT Corp., Milford 2005), http://www.thatcorp.com/vcahist.html

18.52 M. Hart: *History of Sound Motion Pictures* (American WideScreen Museum, 2000), http://www.widescreenmuseum.com/sound/sound03.htm

18.53 L.P. Lessing: *Edwin H. Armstrong, Dictionary of American Biography, Supplement V* (Scribner, New York 1977) pp. 21–23, http://users.erols.com/oldradio/ (2005)

18.54 A. Antoniou: *Digital Filters: Analysis, Design, and Applications*, 2nd edn. (McGraw-Hill, New York 1993)

18.55 A. Antoniou: *On the Origins of Digital Signal Processing* (Univ. Victoria, Victoria 2004), http://www.ece.uvic.ca/ andreas/OriginsDSP.pdf

18.56 R.W.T. Preater, R. Swain: Fourier transform fringe analysis of electronic speckle pattern interferometry fringes from hinge-speed rotating components, Opt. Eng. **33**, 1271–1279 (1994)

18.57 E.O. Bringham: *The Fast Fourier Transform* (Prentice-Hall, Upper Saddle River 1974)

18.58 K.C. Pohlmann: *The Compact Disc Handbook*, Comp. Music Digital Audio Ser., Vol. 5, 2nd edn. (A-R Editions, Middleton 1992)

18.59 J. Despain: *History of CD Technology* (OneOff Media, Salt Lake City 1999), http://www.oneoffcd.com/info/historycd.cfm?CFID=435613&CFTOKEN=10955739

18.60 E.T. Whittaker: On the functions which are represented by the expansions of interpolation-theory, Proc. R. Soc. Edinburgh. **35**, 181–194 (1915)

18.61 C.E. Shannon: A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379–423 (1948)

18.62 C.E. Shannon: A mathematical theory of communication, Bell Syst. Tech. J. **27**, 623–656 (1948)

18.63 E. Meijering: Chronology of interpolation, Proc. IEEE **90**(3), 319–342 (2002), http://imagescience.bigr.nl/meijering/research/chronology/

18.64 M.T. Heideman, D.H. Johnson, C.S. Burrus: Gauss and the history of the fast Fourier transform, IEEE Acoust. Speech Signal Process. Mag. **1**, 14–21 (1984),

18.65 M.T. Heideman, D.H. Johnson, C.S. Burrus: Gauss and the history of the fast Fourier transform, Arch. Hist. Exact Sci. **34**, 265–277 (1985)

18.66 C.S. Burrus: *Notes on the FFT* (Rice University, Houston 1997), http://www.fftw.org/burrus-notes.html

18.67 P. Duhamel, M. Vetterli: Fast Fourier transforms: A tutorial review and a state of the art, Signal Process. **19**, 259–299 (1990)

18.68 Fast Fourier Transform (Wikipedia, August 2005) http://en.wikipedia.org/wiki/Fast_Fourier_transform

18.69 M.F. Davis: *The AC-3 multichannel coder, Audio Engineering Society. 95th Convention 1993* (Audio Engineering Society, New York 1993), Preprint Number 3774

18.70 M.F. Davis: *The AC-3 multichannel coder* (Digital Audio Development, Newark 1993), http://www.dadev.com/ac3faq.asp

Amazon
EX1013 Page 788

18.71 T. Sporer: *Creating, assessing and rendering in real time of high quality audio- visual environments in MPEG-4 context, Information Society Technologies* (CORDIS, Luxemburg 2002), http://www.cordis.lu/ist/ka3/iaf/projects/carrouso.htm

18.72 G. Theile, H. Wittek: Wave field synthesis: A promising spatial audio rendering concept, Acoust. Sci. Tech. **25**, 393–399 (2004), DOI: 10.1250/ast.25.393

18.73 E.I. Schwartz: *The sound war, Technol. Rev. May 2004* (MIT, Cambridge 2004), http://www.woodynorris.com/Articles/TechnologyReview.htm

Amazon
EX1013 Page 789

Part F Biological

Part F Biological and Medical Acoustics

19 Animal Bioacoustics
Neville H. Fletcher, Canberra, Australia

20 Cetacean Acoustics
Whitlow W. L. Au, Kailua, USA
Marc O. Lammers, Kailua, USA

21 Medical Acoustics
Kirk W. Beach, Seattle, USA
Barbrina Dunmire, Seattle, USA

Amazon
EX1013 Page 790

19. Animal Bioacoustics

Animals rely upon their acoustic and vibrational senses and abilities to detect the presence of both predators and prey and to communicate with members of the same species. This chapter surveys the physical bases of these abilities and their evolutionary optimization in insects, birds, and other land animals, and in a variety of aquatic animals other than cetaceans, which are treated in Chap. 20. While there are many individual variations, and some animals devote an immense fraction of their time and energy to acoustic communication, there are also many common features in their sound production and in the detection of sounds and vibrations. Excellent treatments of these matters from a biological viewpoint are given in several notable books [19.1, 2] and collections of papers [19.3–8], together with other more specialized books to be mentioned in the following sections, but treatments from an acoustical viewpoint [19.9] are rare. The main difference between these two approaches is that biological books tend to concentrate on anatomical and physiological details and on behavioral outcomes, while acoustical books use simplified anatomical models and quantitative analysis to model whole-system behavior. This latter is the approach to be adopted here.

19.1 **Optimized Communication** 785
19.2 **Hearing and Sound Production** 787
19.3 **Vibrational Communication** 788
19.4 **Insects** 788
19.5 **Land Vertebrates** 790
19.6 **Birds** 795
19.7 **Bats** 796
19.8 **Aquatic Animals** 797
19.9 **Generalities** 799
19.10 **Quantitative System Analysis** 799
References 802

19.1 Optimized Communication

Since animals use their acoustic and vibrational senses both to monitor their environment and to communicate with other animals of the same species, we should expect that natural selection has optimized these sensing and sound production abilities. One particular obvious optimization is to maximize the range over which they can communicate with others of the same species. Simple observation shows that small animals generally use high frequencies for communication while large animals use low frequencies – what determines the best choice? The belief that there is likely to be some sort of universal scaling law involved goes back to the classic work of *D'Arcy Thompson* [19.10], while a modern overview of physical scaling laws (though not including auditory communication) is given by *West* and *Brown* [19.11].

The simplest assumption is that the frequency is determined simply by the physical properties of the sound-producing mechanism. For a category of animals differing only in size, the vibration frequency of the sound-producing organ depends upon the linear dimensions of the vibrating structure, which are all proportional to the linear size L of the animal, and upon the density ρ and elastic modulus E of the material from which it is made. We can thus write that the song frequency $f = A\rho^x E^y L^z$, where A, x, y, and z are constants. Since the dimensions of each side of the equation must agree, we must have $x = -1$, $y = 1$ and $z = -1$, which leads to the conclusion that song frequency should be inversely proportional to the linear size of the animal or, equivalently, that $f \propto M^{-1/3}$, where M is the mass of the animal. This simple result agrees quite well with

Amazon
EX1013 Page 791

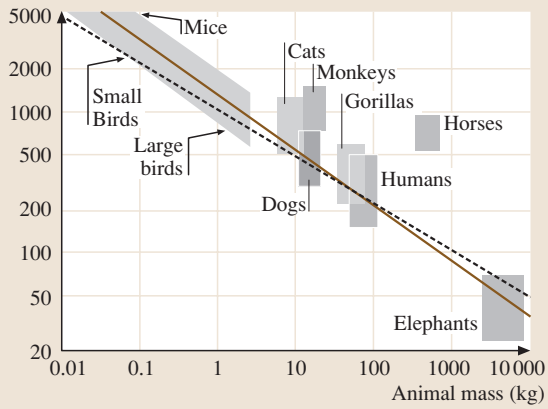


Fig. 19.1 The frequency ranges of the emphasized frequencies of vocalization in a large range of land-dwelling animals, plotted as a function of the mass of the animal. The *dashed line* shows the regression $f \propto M^{-0.33}$ while the *solid line* is the regression $f \propto M^{-0.4}$, as discussed in the text [19.12]

observation, as shown by the dashed line in Fig. 19.1, for a wide variety of animals, but a more detailed analysis is desirable.

Over reasonable distances in the air, sound spreads nearly hemispherically, so that its intensity decays like $1/R^2$, where R is the distance from the source. But sound is also attenuated by atmospheric absorption, with the attenuation coefficient varying as f^2. An optimally evolved animal should have maximized its communication distance under the influence of these two effects. An analysis [19.12] gives the result that $f \propto M^{-0.4}$ which, as shown in Fig. 19.1, fits the observations even better than the simpler result. There are, however, many outliers even among the animals considered, due to different anatomies and habitats. Insects, which must produce their sound in an entirely different way as is discussed later, are not shown on this graph, but there is a similar but not identical relative size relation for them too.

The total sound power produced by an animal is also a function of its size, typically scaling about as $M^{0.53}$ for air-breathing animals of a given category. When the variation of song frequency is included, this leads [19.12] to a conspecific communication distance proportional about to $M^{0.6}$. Again, while this trend agrees with general observations, there are many very notable outliers and great differences between different types of animals. Thus, while mammals comparable in size with humans typically produce sound power in the range 0.1–10 mW, and large birds may produce comparable power, some small insects such as cicadas of mass not much more than 1 g can also produce almost continuous calls with a power of 1 mW, as will be discussed in Sect. 19.4. At intermediate sizes, however, many animals, particularly reptiles, are almost mute.

Elephants represent an interesting extreme in the animal world because of their very large mass – as much as 10 tonnes. Their calls, which have a fundamental in the range 14–35 Hz, can have an acoustic power as large as 5 W, leading to a detection distance as large as 5 km, or even up to 10 km after sunset on a clear night [19.13] when very low frequency propagation is aided by atmospheric inversion layers (Chap. 4). Vegetation too, of course, can have a significant effect upon transmission distance. These elephant calls are often misleadingly referred to as "infrasonic" in the biological literature, despite the fact that only the fundamental and perhaps the second harmonic satisfy this criterion, and the calls have many harmonics well within the hearing range of humans. Indeed, even other elephants depend upon these upper harmonics to convey information, and usually cannot recognize the calls of members of the same family group above a distance of about 1.5 km because of the attenuation of harmonics above about 100 Hz through atmospheric absorption [19.14].

When it comes to sound detection, rather similar scaling principles operate. As will be discussed briefly in Sect. 19.2, the basic neuro-physiological mechanisms for the conversion of vibrations to neural impulses are remarkably similar in different animal classes, so that it is to be expected that auditory sensitivity should vary roughly as the area of the hearing receptor, and thus about as $M^{2/3}$, and this feature is built into the analysis referred to above. While a few animals have narrow-band hearing adapted to detecting particular predators – for example caterpillars detecting wing beats from wasps – or for conspecific communication, as in some insects, most higher animals require a wide frequency range so as to detect larger predators, which generally produce sounds of lower frequency, and smaller prey, which may produce higher frequencies. The auditory range for most higher animals therefore extends with reasonable sensitivity over a frequency range of about a factor 300, with the central frequency being higher for smaller animals. In the case of mammals comparable in size to humans, the range for mature adults is typically about 50 Hz to 15 kHz. This wide range, however, means that irrelevant background noise can become a problem, so that conspecific call frequencies may be adapted, even over a short time, to optimize communication. Small echo-

Amazon
EX1013 Page 792

locating animals, such as bats, generally have a band of enhanced hearing sensitivity and resolution near their call frequency.

Aquatic animals, of course, operate in very different acoustic conditions to land animals, because the density of water nearly matches that of their body tissues. In addition, while the water medium may be considered to be essentially three-dimensional over small distances, it becomes effectively two-dimensional for the very long-distance communication of which some large aquatic animals are capable. Some of these matters will be discussed in Sect. 19.8.

19.2 Hearing and Sound Production

Despite the large differences in anatomy and the great separation in evolutionary time between different animals, there are many surprising similarities in their mechanisms of sound production and hearing. The basic mechanism by which sound or vibration is converted to neural sensation is usually one involving displacement of a set of tiny hairs (cilia) mounted on a cell in such a way that their displacement opens an ion channel and causes an electric potential change in the cell, ultimately leading to a nerve impulse. Such sensory hairs occur in the lateral line organs of some species of fish, supporting the otoliths in other aquatic species, and in the cochlea of land-dwelling mammals. Even the sensory cells of insects are only a little different. In the subsequent discussion there will not be space to consider these matters more fully, but they are treated in more detail in Chap. 12.

Somewhat surprisingly, the auditory sensitivities of animals differing widely in both size and anatomy vary less than one might expect. Human hearing, for example, has a threshold of about 20 μPa in its most sensitive frequency range between 500 Hz and 5 kHz, where most auditory information about conspecific communication and the cries of predators and prey is concentrated, and the sensitivity is within about 10 dB of this value over a frequency range from about 200 Hz to 7 kHz. Other mammals have very similar sensitivities with frequency ranges scaled to vary roughly as the inverse of their linear dimensions, as discussed in Sect. 19.1. Insects in general have narrower bandwidth hearing matched to song frequencies for conspecific communication. Of course, with all of these generalizations there are many outliers with very different hearing abilities.

Sound production mechanisms fall into two categories depending upon whether or not the animal is actively air-breathing. For air-breathing animals there is an internal air reservoir, the volume and pressure of which are under muscular control, so that air can be either inhaled or exhaled. When exhalation is done through some sort of valve with mechanically resonant flaps or membranes, it can set the valve into vibration, thus producing an oscillatory exhalation of air and so a sound source. In the case of aquatic mammals, which are also air-breathing, it would be wasteful to exhale the air, so that it is instead moved from one reservoir to another through the oscillating valve, the vibration of the thin walls of one of the reservoirs then radiating the sound, as is discussed in more detail in Chap. 19. Animals such as insects that do not breathe actively must generally make use of muscle-driven mechanical vibrations to produce their calls, though a few such as the cockroach *Blaberus* can actually make "hissing" noises through its respiratory system when alarmed.

The amount of muscular effort an animal is prepared to expend on vocalization varies very widely. Humans, apart from a few exceptions such as trained singers, can produce a maximum sound output of a few hundred milliwatts for a few seconds, and only about 10 mW for an extended time. In terms of body mass, this amounts to something around 0.1 mW kg^{-1}. At the other end of the scale, some species of cicada with body mass of about 1 g can produce about 1 mW of sound output at a frequency of about 3 kHz, or about 1000 mW kg^{-1} on a nearly continuous basis. An interesting comparative study of birdsongs by *Brackenbury* [19.15] showed sound powers ranging from 0.15 to 200 mW and relative powers ranging from 10 to 870 mW kg^{-1}. The clear winner on the relative power criterion was the small *Turdus philomelos* with a sound output of 60 mW and a body mass of only 69 g, equivalent to 870 mW kg^{-1}, while the loudest bird measured was the common rooster *Gallus domesticus* with a sound output during crowing of 200 mW. Its body mass of 3500 g meant, however, that its relative power was only 57 mW kg^{-1}.

Nearly all sound-production mechanisms, ranging from animals through musical instruments to electrically driven loudspeakers, are quite inefficient, with a ratio of radiated acoustic power to input power of not much more than 1%, and often a good deal less than this. In the case of air-breathing animals, the losses

Amazon
EX1013 Page 793

are largely due to viscosity-induced turbulence above the vocal valve, while in insects that rely upon vibrating body structures internal losses in body tissue are dominant. Wide-bandwidth systems such as the human vocal apparatus are generally much less efficient than narrow-band systems, the parameters of which can be optimized to give good results at a single song frequency. To all this must then be added the internal metabolic loss in muscles, which has not been considered in the 1% figure.

19.3 Vibrational Communication

Most animals are sensitive to vibrations in materials or structures with which they have physical contact. Even humans can sense vibrations over a reasonably large frequency range by simply pressing finger tips against the vibrating object, while vibrations of lower frequency can also be felt through the feet, without requiring any specialized vibration receptors. Many insects make use of this ability to track prey and also for conspecific communication, and some for which this mode of perception is very important have specialized receptors, called subgenual organs, just below a leg joint. A brief survey for the case of insects has been given by *Ewing* [19.16].

For water gliders and other insects that hunt insects that have become trapped by surface tension on the surface of relatively still ponds, the obvious vibratory signal is the surface waves spreading out circularly from the struggles of the trapped insect. The propagation of these surface waves on water that is deep compared with the wavelength has been the subject of analysis, and the calculated wave velocity c_s, which is influenced by both the surface tension T of water and the gravitational acceleration g, is found to depend upon the wavelength λ according to the relation

$$c_s = \left(\frac{2\pi T}{\rho_w \lambda} + \frac{g\lambda}{2\pi}\right)^{1/2}, \qquad (19.1)$$

where ρ_w is the density of water. From the form of this relation it is clear that the wave speed has a minimum value when $\lambda = 2\pi(T/g\rho_w)^{1/2}$, and substituting numerical values then gives a wavelength of about 18 mm, a propagation speed of about 24 cm s^{-1}, and a frequency of about 1.3 Hz. Waves of about this wavelength will continue to move slowly across the surface after other waves have dissipated. The attenuation of all these surface waves is, however, very great.

From (19.1) it is clear that, at low frequencies and long wavelengths, the wave speed c_s is controlled by gravitational forces and $c_s \approx g/\omega$. These are called gravity waves. At high frequencies and short wavelengths, surface tension effects are dominant and $c_s \approx (T\omega/\rho_w)^{1/3}$. These are called capillary waves. The frequencies of biological interest are typically in the range 10 to 100 Hz, and so are in the in the lower part of the capillary-wave regime, a little above the crossover between these two influences, and the propagation speeds are typically in the range 30 to 50 cm s^{-1}.

For insects and other animals living aloft in trees and plants, the waves responsible for transmission of vibration are mostly bending waves in the leaves or branches. The wave speed is then proportional to the square root of the thickness of the structure and varies also as the square root of frequency. Transmission is very slow compared with the speed of sound in air, although generally much faster than the speed of surface waves on water. Once again, the attenuation with propagation distance is usually large.

Disturbances on heavy solid branches or on solid ground also generate surface waves that can be detected at a distance, but these waves propagate with nearly the speed of shear waves in the solid, and thus at speeds of order 2000 m s^{-1}. On the ground they are closely related to seismic waves, which generally have much larger amplitude, and animals can also detect these. Underground, of course, as when an animal is seeking prey in a shallow burrow, the vibration is propagated by bulk shear and compressional waves.

19.4 Insects

The variety of sound production and detection mechanisms across insect species is immense [19.3, 16–18], but their acoustic abilities and senses possess many features in common. Many, particularly immature forms such as caterpillars but also many other insects, use external sensory hairs to detect air movement and these can sense both acoustic stimuli, particularly if tuned to resonance, or the larger near-field air flows produced by the wings of predators [19.19]. These hairs respond to the viscous drag created by acoustic air flow past them and, since

Amazon
EX1013 Page 794

this is directional, the insect may be able to gain some information about the location of the sound source.

Analysis of the behavior of sensory hairs is complex [19.19], but some general statements can be made. Since what is being sensed is an oscillatory flow of air across the surface of the insect, with the hair standing normal to this surface, it is important that the hair be long enough to protrude above the viscous boundary layer, the thickness of which varies inversely with the square root of frequency and is about 0.1 mm at 150 Hz. This means that the length of the hair should also be about inversely proportional to the frequency of the stimulus it is optimized to detect, and that it should typically be at least a few tenths of a millimeter in length. The thickness of the hair is not of immense importance, provided that it is less than about 10 μm for a typical hair, since much of the effective mass is contributed by co-moving fluid. At the mechanical resonance frequency of the hair, the sensitivity is a maximum, and the hair tip displacement is typically about twice the fluid displacement, provided the damping at the hair root is not large, but the response falls off above and below this frequency.

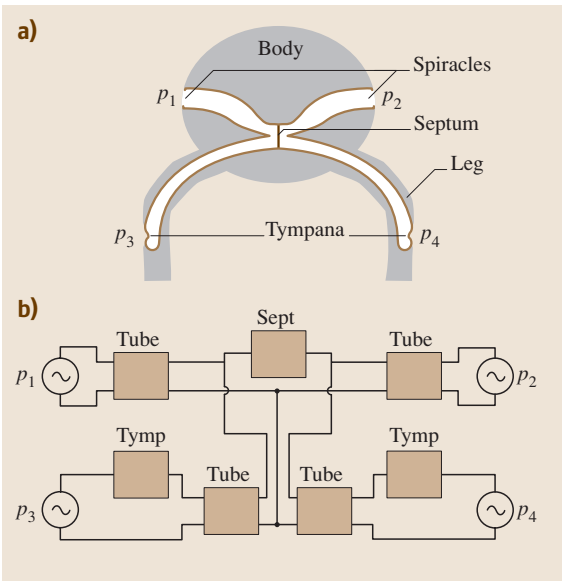


Fig. 19.2 (**a**) Schematic drawing of the auditory anatomy of a cricket. (**b**) Simplified electric network analog for this system. The impedances of the tympana and the septum are each represented by a set of components L, R, C connected in series; the tube impedances are represented by 2×2 matrices. The acoustic pressures p_i acting on the system are generally all nearly equal in magnitude, but differ in phase depending upon the direction of incidence of the sound

Mature insects of some species, however, have auditory systems that bear a superficial resemblance to those of modern humans in their overall structure. Thus for example, and referring to Fig. 19.2a, the cricket has thin sensory membranes (tympana or eardrums) on its forelegs, and these are connected by tubes (Eustachian tubes) that run to spiracles (nostrils) on its upper body, the left and right tubes interacting through a thin membrane (nasal septum) inside the body. Analysis of this system [19.9] shows that the tubes and membranes are configured so that each ear has almost cardioid response, directed ipsilaterally, at the song frequency of the insect, allowing the insect to detect the direction of sound arrival. Some simpler auditory systems showing similar directional response are discussed in Sect. 19.5.

Detailed analysis of the response of anatomically complex systems such as this is most easily carried out using electric network analogs, in which pressure is represented by electric potential and acoustic flow by electric current. Tubes are then represented as 2×2 matrices, diaphragms by L, R, C series resonant circuits, and so on, as shown in Fig. 19.2b. Brief details of this approach are given in Sect. 19.10 and fuller treatments can be found in the published literature [19.9, 20].

When it comes to sound production, insects are very different from other animals because they do not have the equivalent of lungs and a muscle-driven respiratory systems. They must therefore generally rely upon muscle-driven vibration of diaphragms somewhere on the body in order to produce sound. One significant exception is the Death's Head Hawk Moth *Acherontia atropos*, which takes in and then expels pulses of air from a cavity closed by a small flap-valve that is set into oscillation by the expelled air to produce short sound pulses [19.16].

Some insects, such as cicadas, have a large abdominal air cavity with two areas of flexible ribbed cartilage that can be made to vibrate in a pulsatory manner under muscle action, progressive buckling of the membrane as the ribs flip between different shapes effectively multiplying the frequency of excitation by as much as a factor 50 compared with the frequency of muscle contraction. The coupling between these tymbal membranes and the air cavity determines the oscillation frequency, and radiation is efficient because of the monopole nature of the source. The sound frequency is typically in the range 3 to 5 kHz, depending upon insect size, and the radiated acoustic power can be as large as 1 mW [19.21] on an almost continuous basis, though the

Amazon
EX1013 Page 795

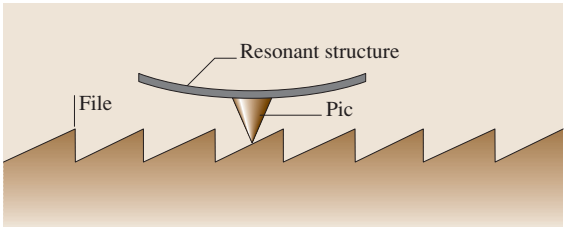


Fig. 19.3 Configuration of a typical file and pic as used by crickets and other animals to excite resonant structures on their wing cases

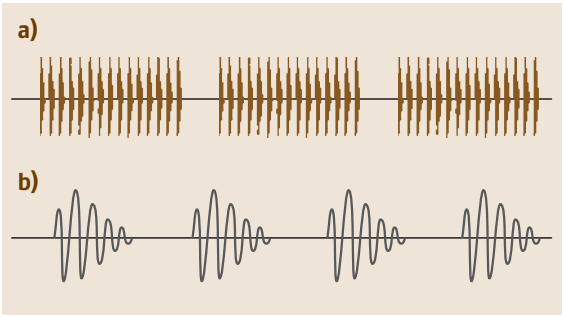


Fig. 19.4a,b Structure of the sound emitted by a typical insect: **(a)** three strokes of the file, the individual oscillations being the pic impacts on the file teeth; **(b)** expanded view of four of the pulses in **(a)**, showing resonant oscillations of the structure produced by successive pic impacts on the file

actual sound signal may consist of a series of regularly spaced bursts corresponding to tymbal collapses and rebounds. Again, there are occasional insects that have developed extreme versions of this sound producing mechanism, such as the bladder cicada *Cystosoma Saundersii (Westwood)*, in which the whole greatly extended abdomen is a resonant air sac and the song frequency is only 800 Hz.

Most other insects do not have significant bodily air cavities and must produce sound by causing parts of their body shell or wings to vibrate by drawing a ribbed file on their legs, or in some cases on their wings, across the hard rim of the membrane, as shown in Fig. 19.3. The anatomical features making up the file and vibrating structure vary very greatly across species, as has been discussed and illustrated in detail by *Dumortier* [19.22]. In some cases the file is located on the vibrating structure and the sharp pic on the part that is moving. The passage of the pic across each tooth or rib on the file generates a sharp impulse that excites a heavily damped transient oscillation of the vibrator, and each repetitive motion of the leg or other structure passes the pic over many teeth. A typical insect call can therefore be subdivided into (a) the leg motion frequency, (b) the pic impact frequency on the file, and (c) the oscillation frequency of the vibrating structure. Since the file may have 10 or more teeth, the frequency of the pic impacts will be more than 10 times the frequency of leg motion, and the structural oscillation may be 10 times this frequency. The structure of a typical call is thus of the form shown in Fig. 19.4. The dominant frequency is usually that of the structural vibration, and this varies widely from about 2 to 80 kHz depending upon species and size [19.23]. Because of the structure of the call, the frequency spectrum appears as a broad band centered on the structural vibration frequency.

Such a vibrating membrane is a dipole source and so is a much less efficient radiator than is a monopole source, except at very high frequencies where the wing dimensions become comparable with the sound wavelength excited. Small insects therefore generally have songs of higher frequency than those that are larger, in much the same way as discussed in Sect. 19.1 for other animals. Some crickets, however, have evolved the strategy of digging a horn-shaped burrow in the earth and positioning themselves at an appropriate place in the horn so as to couple their wing vibrations efficiently to the fundamental mode of the horn resonator, thus taking advantage of the dipole nature of their wing source and greatly enhancing the radiated sound power [19.24].

19.5 Land Vertebrates

The class of vertebrates that live on land is in many ways the most largely studied, since they bear the closest relation to humans, but these animals vary widely in size and behavior. Detailed discussion of birds is deferred to Sect. 19.6, since their songs warrant special attention, while human auditory and vocal abilities are not considered specifically since they are the subject of several other chapters. Reptiles, too, are largely omitted from detailed discussion, since most of them are not notably vocal, and bats are given special attention in Sect. 19.7.

The feature possessed by all the land-dwelling vertebrates is that they vocalize using air expelled from their

Amazon
EX1013 Page 796

inflated lungs through some sort of oscillating vocal valve. The available internal air pressure depends upon the emphasis put by the species on vocalization, but is typically in the range 100–2000 Pa (or 1–20 cm water gauge) almost independent of animal size, which is what is to be expected if the breathing muscles and lung cavity diameter scale about proportionally with the linear size of the animal. There are, however, very wide variations, as discussed in Sect. 19.2. The auditory abilities of most animals are also rather similar, the 20 dB bandwidth extending over a factor of about 100 in frequency and with a center-frequency roughly inversely proportional to linear size, though again there are outliers, as shown in Fig. 19.1. Although the auditory and vocalization frequencies of different animals both follow the same trend with size, the auditory systems generally have a much larger frequency range than the vocalizations for several reasons. The first is that hearing serves many purposes in addition to conspecific communication, particularly the detection of prey and predators. The second is that, particularly with the more sophisticated animals, there are many nuances of communication, such as the vowels and consonants in human speech, for which information is coded in the upper parts of the vocalization spectrum.

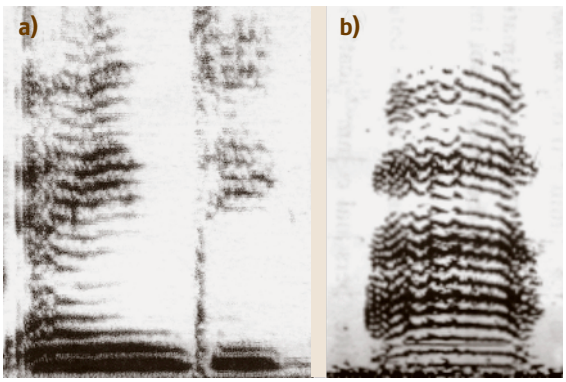


Fig. 19.5a,b A typical time-resolved spectrum for **(a)** human female speech, and **(b)** the cry of a raven, the level of each frequency component being indicated by its darkness. The frequency range is 0–6 kHz and the duration about 0.7 s in **(a)** and 0.5 s in **(b)**. It can be seen that at any instant the spectrum consists of harmonics of the fundamental, with particular bands of frequencies being emphasized. These are the formant bands that distinguish one vowel from another in humans and enable similar distinctions to be made by other animals. Consonants in human speech consist of broadband noise, an example occurring at the beginning of the second syllable in **(a)**, and other animals may have similar interpolations between the tonal sounds

For humans, for example, the primary vocalization frequency is typically in the range 100 to 300 Hz, but the frequency bands that are emphasized in vowel formants lie between 1 and 3 kHz. The same pattern is exhibited in the vocalizations of other air-breathing animals, as shown for the case of a raven in Fig. 19.5. The same formant structure is seen in the cries of elephants, with a fundamental in the range 25–30 Hz and in high-pitched bird songs, which may have a fundamental above 4 kHz.

In most such animals, sound is produced by exhaling the air stored in the lungs through a valve consisting of two taut tissue flaps or membranes that can be made to almost close the base of the vocal tract, as shown in Fig. 19.6a. The lungs in mammals are a complicated

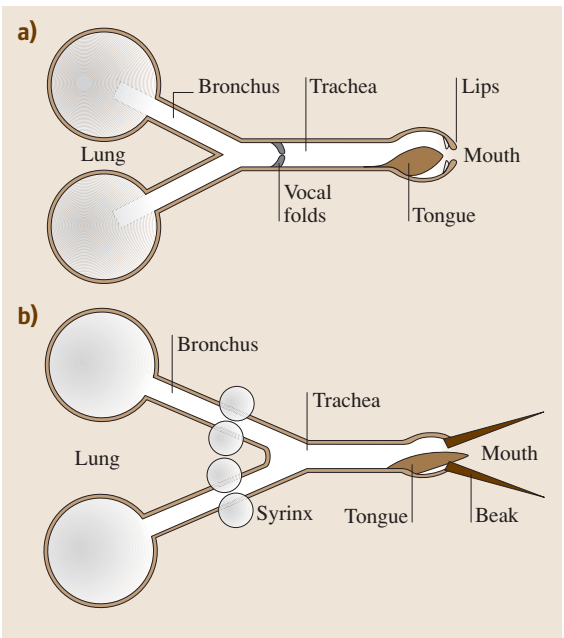


Fig. 19.6 **(a)** Sketch of the vocal system of a typical land mammal. The lungs force air through the dual-flap vocal-fold valve, producing a pulsating flow of air that is rich in harmonics of the fundamental frequency. Resonances of the upper vocal tract, which can be modified by motion of the tongue and lips, produce formants in the sound spectrum which encode vocal information. **(b)** Sketch of the vocal system of a typical song-bird. There are two inflated-membrane valves in the syrinx just below the junction of the bronchi with the trachea, which may be operated either separately or together. Again a harmonic-rich airflow is produced that can either have formants imposed by resonances of the vocal tract and beak, or can be filtered to near pure-tone form by an inflated sac in the upper vocal tract

Amazon
EX1013 Page 797

quasi-fractal dendritic network of tubules with as many as 16 stages of subdivision, the final stage being terminated by alveolar sacs that provide most of storage volume. The interaction of exhaled air pressure and air flow with the flaps of this valve when they have been brought together (or adducted) causes them to vibrate at very nearly their natural frequency, as determined by mass and tension, and this in turn leads to an oscillating air flow through the valve. The classic treatment of the human vocal valve is that of *Ishizaka* and *Flanagan* [19.25], but there have been many more recent treatments exploring modifications and refinements of this model. Very similar vocal valves are found in other mammals, while the major difference in the case of birds is that the flaps of the vocal folds are replaced by opposed taut membranes inflated by air pressure in cavities behind them, and there may be two such valves, as shown in Fig. 19.6b and discussed later in Sect. 19.6.

In most cases, the vibration frequency of the valve is very much lower than the frequency of the first maximum in the upper vocal tract impedance. The valve therefore operates in a nearly autonomous manner at a frequency determined by its geometry, mass and tension and, to a less extent, by the air pressure in the animal's lungs. Because any such valve is necessarily nonlinear in its flow properties, particularly if it actually closes once in each oscillation cycle, this mechanism generates an air flow, and thus a sound wave in the vocal tract, containing all harmonics of the fundamental oscillation. The radiated amplitudes of these harmonics can be modified by resonances in the air column of the vocal tract to produce variations in sound quality and peaks in the radiated spectrum known generally as vocal formants. Changing the tension of the valve flaps changes the fundamental frequency of the vocal sound, but the frequencies of the formants can be changed independently by moving the tongue, jaws, and lips or beak.

To be more quantitative, suppose that the pressure below the valve is p_0 and that above the valve p_1. If the width of the opening is W and its oscillating dimension is $x(t)$, then the volume flow U_1 through it is determined primarily by Bernoulli's equation and is approximately

$$U_1(t) = \left(\frac{2(p_0 - p_1)}{\rho}\right)^{1/2} Wx(t)\,, \tag{19.2}$$

where ρ is the density of air. In order for the valve to be maintained in oscillation near its natural frequency f, the pressure difference $p_0 - p_1$ must vary with the flow and with a phase that is about 90° in advance of the valve opening $x(t)$. This can be achieved if the acoustic impedance of the lungs and bronchi is essentially compliant, and that of the upper vocal tract small, at the oscillation frequency. The fact that this pressure difference appears as a square root in (19.2) then introduces upper harmonic terms at frequencies $2f$, $3f, \ldots$ into the flow $U(t)$. Allowance must then be made for the fact that the vocal valve normally closes for part of each oscillation cycle, so that $x(t)$ is no longer simply sinusoidal, and this introduces further upper-harmonic terms into the flow. This is, of course, a very condensed and simplified treatment of the vocal flow dynamics, and further details can be found in the literature for the case of mammalian, and specifically human, animals [19.25] and also for birds [19.26].

This is, however, only the beginning of the analysis, for it is the acoustic flow out of the mouth that determines the radiated sound, rather than the flow from the vocal valve into the upper vocal tract. Suppose that the acoustic behavior of the upper vocal tract is represented by a 2×2 matrix, as shown in Fig. 19.11 of Sect. 19.10, and that the mouth or beak is regarded as effectively open so that the acoustic pressure p_2 at this end of the tract is almost zero. Then the acoustic volume flow U_2 out of the mouth at a particular angular frequency ω is

$$U_2 = \frac{Z_{21}}{Z_{22}} U_1\,, \tag{19.3}$$

where U_1 is the flow through the larynx or syrinx at this frequency. For the case of a simple cylindrical tube of length L at angular frequency ω, this gives

$$U_2 = \frac{U_1}{\cos kL}\,, \tag{19.4}$$

where $k = (\omega/c) - \mathrm{i}\alpha$, c is the speed of sound in air, and α is the wall damping in the vocal tract. If the spectrum of the valve flow U_1 has a simple envelope, as is normally the case, then (19.4) shows that the radiated sound power, which is proportional to $\omega^2 U_2^2$, has maxima when $\omega L/c = (2n-1)\pi/2$, and thus in a sequence $1, 3, 5, \ldots$ These are the formant bands, the precise frequency relationship of which can be varied by changing the geometry of the upper vocal tract, which in turn changes the forms of Z_{21} and Z_{22}.

Some other air-breathing animals (and even human sopranos singing in their highest register) however, adjust their vocal system so that the frequency of the vocal valve closely matches a major resonance of the upper vocal tract, usually that of lowest frequency but not necessarily so. Some species of frogs and birds achieve this by the incorporation of an inflatable sac in the upper vocal tract. In some cases the sac serves simply to provide

Amazon
EX1013 Page 798

a resonance of appropriate frequency, and the sound is still radiated through the open mouth, but in others the walls of the sac are very thin so that they vibrate under the acoustic pressure and provide the primary radiation source. This will be discussed again in Sect. 19.6 in relation to bird song, but the analysis for frog calls is essentially the same.

The overall efficiency of sound production in vertebrates is generally only 0.1–1%, which is about the same as for musical instruments. The acoustic output power output is typically in the milliwatt range even for quite large animals, but there is a very large variation between different species in the amount of effort that is expended in vocalization.

The hearing mechanism of land-dwelling vertebrates shows considerable similarity across very many species. Sound is detected through the vibrations it induces in a light taut membrane or tympanum, and these vibrations are then communicated to a transduction organ where they are converted to nerve impulses, generally through the agency of hair cells. The mechanism of neural transduction and frequency discrimination is complex, and its explanation goes back to the days of *Helmholtz* [19.27] in the 19th century, with the generally accepted current interpretation being that of *von Békésy* [19.28] in the mid-20th century. Although these studies related specifically to human hearing, the models are generally applicable to other land-dwelling vertebrates as well, as surveyed in the volume edited by *Lewis* et al. [19.29]. All common animals have two auditory transducers, or ears, located on opposite sides of the head, so that they are able to gain information about the direction from which the sound is coming.

In the simplest such auditory system, found for example in frogs, the two ears open directly into the mouth cavity, as shown in Fig. 19.7a, one neural transducer being closely coupled to each of the two tympana. Such a system is very easily analyzed at low frequencies where the cavity presents just a simple compliance with no resonances of its own. Details of the approach are similar to those for the more complex system to be discussed next. For a typical case in which the ears are separated by 20 mm, the cavity volume is 1 cm^3, and the loaded tympanum resonance is at 500 Hz, the calculated response has the form shown in Fig. 19.7b. Directional discrimination is best at the tympanum resonance frequency and can be as much as 20 dB. In a more realistic model for the case of a frog, the nostrils must also be included, since they lead directly into the mouth and allow the entry of sound. The calculated results in this case show that the direction of maximum response for each ear is

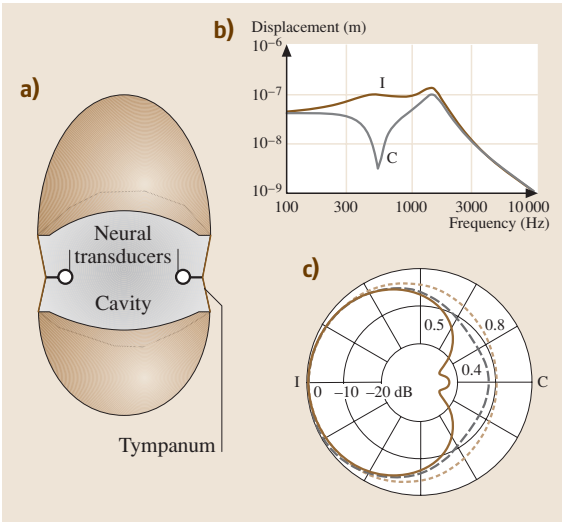


Fig. 19.7 **(a)** A simple frog-ear model; **(b)** the calculated frequency response for ipsilateral (I) and contralateral (C) sound signals for a particular optimized set of parameter values; **(c)** the calculated directivity at different frequencies, shown in kHz as a parameter (from [19.9])

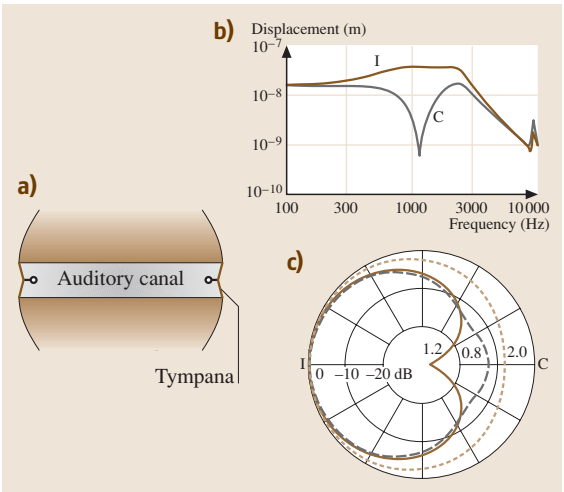


Fig. 19.8 **(a)** Idealized model of the auditory system of a lizard or bird. Each tympanum is connected to a neural transducer and the two are joined by a simple tube. **(b)** Response of a particular case of the system in **(a)** to ipsilateral (I) and contralateral (C) sound of amplitude 0.1 Pa (equivalent to 74 dB re 20 μPa). The tube diameter is assumed to be 5 mm and its length 20 mm, while the tympana have a thickness of 10 μm and are tensioned to a resonance frequency of 1000 Hz. **(c)** Directional response of this hearing system (from [19.9])

Amazon
EX1013 Page 799

shifted towards the rear of the animal, typically by as much as 30° [19.9].

In another simple auditory system such as that of most reptiles, and surprisingly birds, the two tympana are located at the ends of a tube passing through the head and each is supplied with its own neural transducer, as shown in Fig. 19.8a. The tympana may be recessed at the end of short tubes to provide protection, these tubes only slightly modifying the frequency response. The behavior of such a system can be analyzed using the electric network analogs discussed in Sect. 19.10, with each tympanum involving a series L, R, C combination and the tube being represented by a 2×2 impedance matrix Z_{ij}. If the sound comes from straight in front, then it is the same in magnitude and phase at each ear, so that their responses are necessarily the same. When the sound comes from one side, however, there is a phase difference between the two ear pressures, along with a rather less significant amplitude difference. The motion of each tympanum is determined by the difference between the internal and external pressures acting upon it, and the internal pressure depends upon the signal transferred from the opposite tympanum, modified by the phase delay or resonance of the coupling tube. The analysis is straightforward, but the resulting behavior, which depends upon the mass and tension of the tympana and the length and diameter of the connecting tube, can only be determined by explicit calculation [19.9]. When these parameters are appropriately chosen, each ear has a strongly directional response near the resonance frequency of the tympana, as shown in Fig. 19.8b, c. The balance between the depth and sharpness of the contralateral minimum is affected by the acoustic absorption in the walls of the connecting canal. In some animals, the bone lining of this canal is actually porous, which reduces the effective sound-wave velocity inside it and so increases the phase shift between internal and external signals, a feature that is of assistance when the physical separation of the ears is small.

In most mammals and other large animals, the auditory system is modified in several significant ways. The first is that the auditory canal joining the two ears in birds and reptiles has generally degenerated in mammals to the extent that each ear functions nearly independently. The connecting canal in humans has become the pair of Eustachian tubes running from the middle ear cavity, which contains the bones linking the tympanum to the cochlea, down to the nasal cavities, and its main purpose is now simply to equalize internal and external static pressures and to drain excess fluids. The middle ear cavity itself is necessary in order that the enclosed air volume be large enough that it does not raise the resonance frequency of the tympanum by too much. The topology of the whole system is surprisingly similar to that of the cricket auditory system shown in Fig. 19.2a, but the functions of some of the elements are now rather different.

The other major change is that, instead of the tympana being located almost on the surface of the animal, they are buried below the surface at the end of short auditory canals (meati), which lead to external ears (pinnae) in the shape of obliquely truncated horns. As well as protecting the tympana from mechanical damage, the canals add a minor resonance of their own, generally in the upper part of the auditory range of the animal concerned. The pinnae both increase the level of the pressure signal, typically by about 10 dB and in some cases even more, and impart a directionality to the response [19.9, 30]. The convoluted form of some pinnae also imparts directionally-excited resonances that help distinguish direction from tonal cues, a feature that is necessary in order to distinguish up/down directions for sounds coming from immediately ahead or immediately behind. Some animals with particularly large pinnae, such as kangaroos, are able to rotate these independently to help locate a sound source. In all such cases, however, the neural system plays a large part in determining sound direction by comparing the phases and amplitudes, and sometimes the precise timing of transients, in the signals received by the two ears.

In a hearing system of any simple type it is clear that, if geometrical similarity is maintained and the system is simply scaled to the linear size of the animal concerned, then the frequency of maximum discrimination will vary as the inverse of the linear size of the animal, giving an approximate match to general trend of vocalization behavior. There is, however, one anomaly to be noted [19.31]. Vocalizations by some frogs have fundamental frequencies as high as 8 kHz and the vocal signal contains prominent formant bands in the case studied at 20 kHz and 80 kHz, so that they might almost be classed as ultrasonic. Despite this, behavioral evidence indicates that these frogs cannot hear signals much above 10 kHz, so that perhaps the formant bands are simply an epiphenomenon of the sound production system. The same is true of the calls of some birds.

Amazon
EX1013 Page 800

19.6 Birds

Because of their generally melodious songs, birds have excited a great deal of attention among researchers, as described in several classic books [19.32, 33]. Hearing does not require further discussion because the auditory system is of the simple tube-coupled type discussed in Sect. 19.5 and illustrated in Fig. 19.8. Research interest has focused instead on the wide range of song types that are produced by different species. These range from almost pure-tone single-frequency calls produced by doves, and whistle-like sounds sweeping over more than a 2:1 frequency range produced by many other species, through wide-band calls with prominent formants, to the loud and apparently chaotic calls of the Australian sulfur-crested cockatoo. Some birds can even produce two-toned calls by using both syringeal valves simultaneously but with different tuning.

As noted in Sect. 19.5 and illustrated in Fig. 19.6b, song birds have a syrinx consisting of dual inflated-membrane valves, one in each bronchus just below their junction with the trachea. These valves can be operated simultaneously and sometimes at different frequencies, but more usually separately, and produce a pulsating air-flow rich in harmonics. In birds such as ravens, the impedance maxima of the vocal tract, as measured at the syringeal valve, produce emphasized spectral bands or formants much as in human speech, as is shown in Fig. 19.5. The vocal tract of a bird is much less flexible in shape than is that of a mammal, but the formant frequencies can still be altered by changes in tongue position and beak opening. Studies with laboratory models [19.34] show that the beak has several acoustic functions. When the beak is nearly closed, it presents an end-correction to the trachea that is about half the beak length at high frequencies, but only about half this amount at low frequencies, so that the formant resonances are brought closer together in frequency. As the beak gape is widened, the end-correction reduces towards the normal value for a simple open tube at all frequencies. The beak also improves radiation efficiency, particularly at higher frequencies, by about 10 dB for the typical beak modeled. Finally, the beak enhances the directionality of the radiated sound, particularly at high frequencies and for wide gapes, this enhanced directionality being as much as 10 dB compared with an isotropic radiator.

The role of the tongue has received little attention as yet, partly because it is rather stiff in most species of birds. It does, however, tend to constrict the passage between the top of the trachea and the root of the beak, and this constriction can be altered simply by raising the tongue, thereby exaggerating the end-correction provided to the trachea by the beak. These models of beak behavior can be combined with models of the syrinx and upper vocal tract [19.26] to produce a fairly good understanding of vocalization in birds such as ravens. The fundamental frequency of the song can be changed by changing the muscle tension and thus the natural resonance frequency of the syringeal valve, while the formants, and thus the tone of the song, can be varied by changing the beak opening and tongue position. The role of particular anatomical features and muscular activities in controlling the fundamental pitch of the song has since been verified by careful measurement and modeling [19.35, 36].

Some birds have developed the ability to mimic others around them, a notable example being Australian Lyrebirds of the family *Menuridae* which, as well as imitating other birds, have even learned to mimic human-generated sounds such as axe blows, train whistles, and even the sound of film being wound in cameras. There has also been considerable interest in the vocalization of certain parrots and cockatoos, which can produce, as well as melodious songs, quite intelligible imitations of human speech. The important thing here is to produce a sound with large harmonic content and then to tune the vocal tract resonances, particularly those that have frequencies in the range 1–2.5 kHz, to match those of the second and third resonances of the human vocal tract, which are responsible for differentiating vowel sounds. Consonants, of course, are wide-band noise-like transients. Careful studies of various parrots [19.37, 38] show that they can achieve this by careful positioning of the tongue, much in the way that humans achieve the same thing but with quantitative differences because of different vocal tract size. The match to human vowels in the second and third human formants can be quite good. The first formant, around 500 Hz in humans, is not reproduced but this is not important for vowel recognition.

These studies produce a good understanding of the vocalization of normal song-birds, the calls of which consist of a dominant fundamental accompanied by a range of upper harmonics. Less work has been done on understanding the nearly pure-tone cries of other birds. For many years there has been the supposition that the sound production mechanism in this case was quite different, perhaps involving some sort of aerodynamic whistle rather than a syringeal vibration. More recent experimental observations on actual singing birds, using

Amazon
EX1013 Page 801

modern technology [19.39], have established however that, at least in the cases studied, the syringeal valve is in fact normally active, though it does not close completely in any part of the cycle, thus avoiding the generation of large amplitudes of upper harmonics. Suppression of upper harmonics can also be aided by filtering in the upper vocal tract [19.40].

In the case of doves, which produce a brief pure-tone "coo" sound at a single frequency, the explanation appears to lie in the use of a thin-walled inflated esophageal sac and a closed beak, with fine tuning provided by an esophageal constriction [19.41]. The volume V of air enclosed in the sac provides an acoustic compliance $C = V/\rho c^2$ that is in parallel with an acoustic inertance $L = m/S^2$ provided by the mass m and surface area S of the sac walls. The resonance frequency f of this sac filter is given by

$$f = \frac{1}{2\pi}\left(\frac{\rho c^2 S^2}{mV}\right)^{1/2}, \qquad (19.5)$$

where ρ is the density of air and c is the speed of sound in air. When excited at this frequency by an inflow of air from the trachea, the very thin expanded walls of the sac vibrate and radiate sound to the surrounding air. For a typical dove using this strategy, the wall-mass and dimensions of the inflated sac are about right to couple to the "coo" frequency of about 900 Hz. From (19.5), the resonance frequency varies only as the square root of the diameter of the sac, so that a moderate exhalation of breath can be accommodated without much disturbance to the resonance. The dove must, however, learn to adjust the glottal constriction to couple the tracheal resonance efficiently to that of the sac.

Some other birds, such as the Cardinal, that produce nearly pure-tone calls over a wide frequency range, appear to do so with the help of an inflated sac in the upper vocal tract that leads to the opened beak, with the tracheal length, sac volume, glottal constriction, tongue position, and beak opening all contributing to determine the variable resonance frequency of the filter. Because the sac is not exposed as in the dove, its walls are heavy and do not vibrate significantly. It thus acts as a Helmholtz resonator, excited by the input flow from the trachea and vented through the beak. The bird presumably learns to adjust the anatomical variables mentioned above to the syringeal vibration frequency, which is in turn controlled by syringeal membrane tension, and can then produce a nearly pure-tone song over quite a large frequency range. Despite this explanation of the mechanism of pure-tone generation in some cases, others have still to be understood, for the variety of birdsong is so great that variant anatomies and vocalization skills may well have evolved.

With most of the basic acoustic principles of birdsong understood, most of the interest in birdsong centers upon the individual differences between species. The variety found in nature is extremely great and, while sometimes correlated with large variations in bird anatomy, it also has many environmental influences. Some birds, surprisingly, have songs with prominent formant bands extending well into the ultrasonic region, despite the fact that behavioral studies and even auditory brainstem measurements show that they have no response above about 8 kHz [19.31, 42]. The formant bands in these cases are therefore presumably just an incidental product of the vocalization mechanism.

19.7 Bats

Since human hearing, even in young people, is limited to the range from about 20 Hz to about 20 kHz, frequencies lying outside this range are referred to as either infrasonic or ultrasonic. Very large animals such as elephants may produce sounds with appreciable infrasonic components, though the higher harmonics of these sounds are quite audible. Small animals such as bats, however, produce echo-location calls with dominant frequencies around 80 kHz that are inaudible to humans, though some associated components of lower frequency may be heard. In this section brief attention will be given to the sonar abilities of bats, though some other animals also make use of ultrasonic techniques [19.43].

Sound waves are scattered by any inhomogeneities in the medium through which they are propagating, the density of an animal body in air provides a large contrast in acoustic impedance and therefore a large scattered signal. The basic physics shows that, for a target in the form of a sphere of diameter d and a signal of wavelength $\lambda \gg d$, the echo strength is proportional to d^6/λ^4 times the incident intensity. This incident intensity is itself determined by the initial signal power and the angular width of the scanning signal, which varies about as λ/a, where a is the radius of the mouth of the probing animal. If R is the distance between the probing animal and the target, then the sound power incident on the target varies as $a^2/\lambda^2 R^2$ and the returned echo strength E thus varies

Amazon
EX1013 Page 802

as

$$E \propto \frac{d^6 a^2}{\lambda^6 R^4} , \tag{19.6}$$

provided the wavelength λ is much greater than the diameter d of the scattering object and the radius a of the mouth. Another factor must, however, be added to this equation and that is the attenuation of the sound due to atmospheric absorption. This attenuation depends on relative humidity, but is about 0.1 dB m^{-1} at 10 kHz, increasing to about 0.3 dB m^{-1} at 100 kHz, and so is not a very significant factor at the short ranges involved.

Once the size of the target becomes comparable with the incident sound wavelength, the scattering tends towards becoming reflection and geometrical-optics techniques can be used. In this limit, the reflected intensity is proportional to the area of the target, but depends in a complicated way upon its shape and orientation. Finally, if a pure-tone sound is used, any Doppler shift in the frequency of the echo can give information about the relative motion of prey and predator. The reflected sound can therefore give information about the nature of the target and its motion, both of which are important to the pursuing bat.

The sonar-location cries of bats and other animals are, however, not pure-tone continuous signals but rather sound pulses with a typical frequency in the range 40–80 kHz and so a wavelength of about 4–8 mm. This wavelength is comparable to the size of typical prey insects so that the echo can give information about size, shape and wing motion, while the echo delay reveals the distance. The bat's large ears are also quite directional at these frequencies, so that further information on location is obtained. Rather than using short bursts of pure tone, some predators use longer bursts in which the frequency is swept rapidly through a range of several kilohertz. This technique is known as "chirping", and it allows the returned echo to be reconstructed as a single pulse by frequency-dependent time delay, although animal neural systems may not actually do this. The advantage of chirping is that the total energy contained in the probing pulse, and thus the echo, can be made much larger without requiring a large peak power.

Because of the high frequencies involved in such sonar detection, bats have a hearing system that is specialized to cover this high-frequency range and to provide extended frequency resolution near the signal frequency, perhaps to aid in the neural equivalent of "dechirping" the signal and detecting Doppler frequency shifts. The auditory sensitivity curve of bats is similar in shape to that of humans (threshold about 20 μPa), but shifted up in frequency by a factor of about 20 to give high sensitivity in the range 10–100 kHz [19.1]. The threshold curve generally shows a particular sensitivity maximum near the frequency of the echo-location call, as might be expected. Bats also generally have forward-facing ears that are large in comparison with their body size. The functions of these in terms of increased directionality and larger receiving area are clear [19.30].

Sound production in bats is generally similar to that in other animals, but with the system dimensions tailored to the higher frequencies involved, and often with the sound emitted through the nose rather than the mouth. The complicated geometry of the nasal tract, which has several spaced enlargements along its length, appears to act as an acoustic filter which suppresses the fundamental and third harmonic of the emitted sound while reinforcing the second harmonic, at least in some species [19.44]. The physics of this arrangement is very similar to that of matching stubs applied to high-frequency transmission lines.

19.8 Aquatic Animals

The generation and propagation of sound under water, as discussed in Chap. 19, have many features that are different from those applying to sound in air. From a physical point of view, a major difference is the fact that water is often quite shallow compared with the range of sound, so that propagation tends to be two-dimensional at long distances. From the present biological viewpoint the major difference is that the acoustic impedance of biological tissue is nearly equal to that of the surrounding water, while in air these differ by a factor of about 3500. This leads to some very significant differences in the auditory anatomy and behavior between aquatic and air-dwelling animals.

The variety of aquatic animal species is comparable to that of land-dwelling animals, and they can be divided into three main categories for the purposes of the present chapter. First come the crustaceans, such as shrimp, lobsters and crabs, which are analogous to insects, then the wide variety of fish with backbones, and finally the air-breathing aquatic mammals such as the

Amazon
EX1013 Page 803

cetaceans (whales, dolphins and porpoises, which are the subject of Chap. 19), though animals such as turtles also come into this category. There are, of course, many other categories of marine animal life, but they are not generally significant from an acoustic point of view. A good review has been given by *Hawkins* and *Myrberg* [19.45].

The crustaceans, like most insects, generally produce sound by rubbing a toothed leg-file against one of the plates covering their body, thus producing a pulse of rhythmic clicks. Some animals, such as shrimp, may produce a single loud click or pop by snapping their claws in much the same way that a human can do with fingers. Because this activity takes place under water, which has density ρ about 800 times greater and sound velocity c about 4.4 times greater than that of air, and thus a specific acoustic impedance ρc about 3500 times greater than air, the displacement amplitude of the vibration does not need to be large, and the vibrating structure is generally small and stiff. The peak radiated acoustic power in the "click" can, however, be very large – as much as several watts in the Pistol shrimp *Alpheidae* – probably from a mechanism involving acoustic cavitation. The click lasts for only a very short time, however, so that the total radiated energy is small.

Fish species that lack an air-filled "swim-bladder", the main purpose of which is to provide buoyancy, must produce sound in a similar manner. For fish that do have a swim-bladder, however, the sound production process is much more like that of the cicada, with some sort of plate or membrane over the bladder, or even the bladder wall itself, being set into oscillation by muscular effort and its displacement being aided by the compressibility of the air contained in the bladder. The surrounding tissue and water provide a very substantial mass loading which also contributes to determining the primary resonance frequency of the bladder. This is a much more efficient sound production process than that of the crustaceans and leads to much louder sounds, again generally consisting of a train of repetitive pulses at a frequency typically in the range 100–1000 Hz.

When it comes to sound detection, hair cells again come into play in a variety of ways. Because the density of the animal body is very similar to that of the surrounding water, a plane wave of sound tends to simply move the animal and the water in just the same way. The slight difference in density, however, does allow for some differential motion. The animal body is, however, relatively rigid, so that its response to diverging waves in the near field is rather different from that of the liquid. Any such relative motion between the body and the surrounding water can be detected by light structures such as hairs protruding from the body of the animal, and these can form the basis of hair-cell acoustic detectors. Even more than this, since such cells are generally sensitive to hair deflection in just one direction, arrays of hair cells can be structured so as to give information about the direction of origin of the sound. Although some such detector hairs may be mechanically strong and protrude from the body of the animal, others may be very small and delicate and may be protected from damage by being located in trench-like structures, these structures being called lateral-line organs. Their geometry also has an effect on sound localization sensitivity.

There is another type of hair detector that responds to acceleration rather than relative displacement and that therefore takes advantage of the fact that the body of the animal tends to follow the acoustic displacement of the surrounding water. This is the otolith, which consists of a small dense calcareous stone supported by one or more hair cells,as shown in Fig. 19.9. Since the motion of this stone tends to lag behind the motion of the liquid, and thus of its supports, the supporting hairs are deflected in synchrony with the acoustic stimulus, so that it can be detected by the cells to which they are attached. These sensors will normally have evolved so that combined mass of the otolith and stiffness of the hair give a resonance frequency in the range of interest to the animal, and detection sensitivity will be greatest at this frequency.

If the fish possesses an air-filled swim-bladder, then this can also be incorporated in the hearing sys-

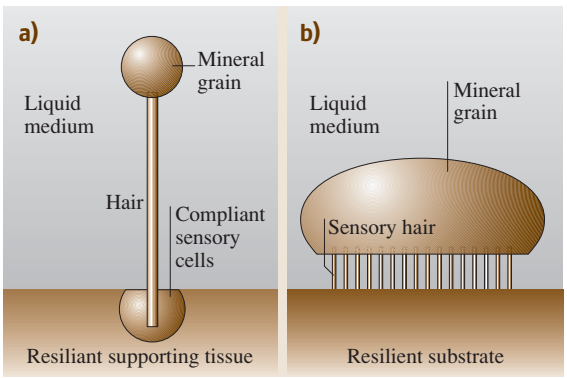


Fig. 19.9a,b Two possible forms for a simple otolith detector: **(a)** a single-cell detector, **(b)** a multi-cell detector. As the mineral grain moves sideways relative to the substrate, the sensory hairs are deflected and open ion channels in the supporting transducer cells (from [19.9])

Amazon
EX1013 Page 804

tem [19.46]. An air bubble in water is made to move in a sound wave by the differential pressure across it. Because its mass is small and the pressure gradient is the same as that on an equivalent volume of water, the bubble tends to move more than the surrounding liquid but, in order to do so, it must displace a significant amount of this liquid. The final result is that the amplitude of free bubble motion is about three times the amplitude of the motion of the surrounding liquid. Since the body of the fish tends to follow the liquid motion because its density is nearly the same, there is relative motion between the bladder and the rest of the fish, provided it is not rigidly enclosed. This relative motion can be detected by neural cells pressed against the bladder wall and provides an efficient sensor of underwater sound. The radial compressive resonances of the swim bladder can also be important in the detection of sound and, since these are also involved in sound production in these fish, there is a good match between call frequency and the frequency of maximum hearing sensitivity, which typically lies in the range 100 to 1000 Hz. Fish with swim bladders have thus evolved to use these primarily buoyancy-regulating structures for both the production and detection of sound.

19.9 Generalities

This brief survey has shown the strong resemblance between the basic sound-producing and auditory mechanisms in a wide variety of animals. Certainly there is a distinction in the matter of sound production between those animals that store air within their bodies and those that do not, and a rather similar distinction in the matter of sound detection, but the similarities are more striking than the differences. In particular, the universal importance of hair-cells in effecting the transduction from acoustic to neural information is most striking.

Also notable in its generality is the scaling of the dominant frequency of produced sound and the associated frequency of greatest hearing sensitivity approximately inversely with the linear dimensions of the animal involved. Certainly there are notable outliers in this scaling, but that in itself is a clue to the significantly different acoustic behavior of the animal species concerned.

Fortunately, quite straightforward acoustic analysis provides a good semi-quantitative description of the operation of most of these animal acoustic systems, and in some cases such analysis can reveal evolutionarily optimized solutions to modern acoustical design problems.

19.10 Quantitative System Analysis

Biological systems are almost always anatomically complex, and this is certainly true of vocal and auditory systems, which typically comprise sets of interconnecting tubes, horns, cavities and membranes driven over a wide frequency range by some sort of internal or external excitation. While a qualitative description of the way in which such a system works can be helpful, this must be largely based upon conjectures unless one is able to produce a quantitative model for its behavior. Fortunately a method has been developed for the construction of detailed analytical models for such systems from which their acoustic behavior can be quantitatively predicted, thus allowing detailed comparison with experiment. This section will outline how this can be done, more detail being given in [19.9] and [19.20].

In a simple mechanical system consisting of masses and springs, the quantities of interest are the force F_i applied at point i and the resulting velocity of motion v_j at point j, and we can define the mechanical impedance Z^{mech} by the equation $Z^{\text{mech}}_{jj} = F_j/v_j$. Note that the subscripts on F and v are here both the same, so that Z^{mech}_{jj} is the ratio of these two quantities measured at a single point j. It is also possible to define a quantity $Z^{\text{mech}}_{ij} = F_i/v_j$ which is called the transfer impedance between the two points i and j. There is an important theorem called the reciprocity theorem, which shows that $Z_{ji} = Z_{ij}$.

In such a simple mechanical system, as well as the masses and springs, there is normally some sort of loss mechanism, usually associated with internal viscous-type losses in the springs, and the situation of interest involves the application of an oscillating force of magnitude F and angular frequency ω. We denote this force in complex notation as $F\,\mathrm{e}^{\mathrm{i}\omega t}$ where $\mathrm{i} = \sqrt{-1}$. (Note

Amazon
EX1013 Page 805

that, in the interest of consistency throughout this Handbook, this notation differs from the standard electrical engineering notation in which i is replaced by j.) For the motion of a simple mass m under the influence of a force F we know that $F = m\,\mathrm{d}v/\mathrm{d}t$ or $F = \mathrm{i}\omega m v$, so that the impedance provided by the mass is $\mathrm{i}\omega m$. Similarly, for a spring of compliance C, $F = \int v\,\mathrm{d}t/C = -\mathrm{i}/(\omega C)$, giving an impedance $-\mathrm{i}/(\omega C)$. For a viscous loss within the spring there is an equivalent resistance R proportional to the speed of motion, so that $F = Rv$. Examination of these equations shows that they are closely similar to those for electric circuits if we assume that voltage is the electrical analog of force and current the analog of velocity. The analog of a mass is thus an inductance of magnitude m, the analog of a compliant spring is a capacitance C, and the analog of a viscous resistance is an electrical resistance R. A frictional resistance is much more complicated to model since its magnitude is constant once sliding occurs but changes direction with the direction of motion, so that its whole behavior is nonlinear.

These mechanical elements can then be combined into a simple circuit, two examples being shown in Fig. 19.10. In example (a) it can be seen that the motion of the mass, the top of the spring, and the viscous drag are all the same, so that in the electrical analog the same current must flow through each, implying a simple series circuit as in (b). The total input impedance is therefore the sum of the three component impedances, so that

$$Z^{\mathrm{mech}} = R + \mathrm{i}\omega m - \frac{\mathrm{i}}{\omega C} = R + \frac{\mathrm{i}m}{\omega}\left(\omega^2 - \frac{1}{mC}\right), \tag{19.7}$$

which has an impedance minimum, and thus maximum motion, at the resonance frequency $\omega^* = 1/(mC)^{1/2}$. In example (c), the displacement of the top of the spring is the sum of the compressive displacement of the spring and the motion of the suspended mass, and so the analog current is split between the spring on the one hand and the mass on the other, the viscous drag having the same motion as the spring, leading to the parallel resonant circuit shown in (d). There is an impedance maximum, and thus minimum motion of the top of the spring, at the resonance frequency, which is again $\omega^* = 1/(mC)^{1/2}$. In a simple variant of these models, the spring could be undamped and the mass immersed in a viscous fluid, in which case the resistance R would be in series with the inductance rather than the capacitance. These analogs can be extended to include things such as levers, which appear as electrical transformers with turns ratio equal to the length ratio of the lever arms.

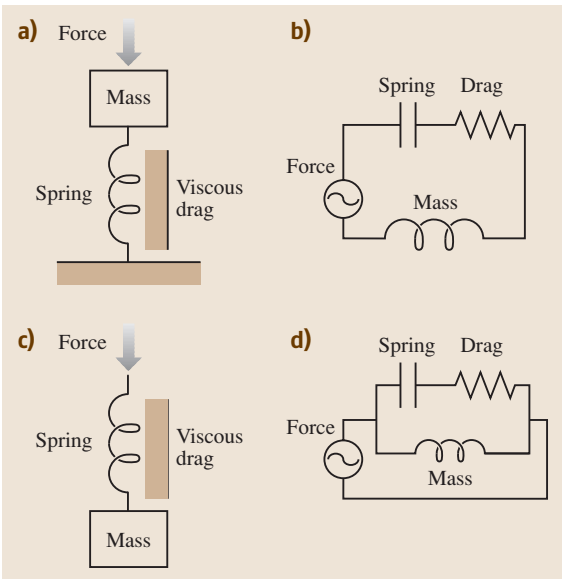


Fig. 19.10 **(a)** A simple resonant assembly with an oscillating force applied to a mass that is resting on a rigidly supported spring, the spring having internal viscous losses. **(b)** Analog electric network for the system shown in **(a)**. **(c)** An alternative assembly with the mass being freely suspended from the spring and the oscillating force being applied to the free end of the spring. **(d)** Analog electric network for the system shown in **(c)**. (Note the potential source of confusion in that an electrical inductance symbol looks like a physical spring but actually represents a mass)

While such combinations of mechanical elements are important in some aspects of animal acoustics, such as the motion of otolith detectors, most of the systems of interest involve the generation and detection of acoustic waves rather than of mechanical vibrations. The important physical variables are then the acoustic pressure p and the acoustic volume flow U, both being functions of the frequency ω. By analogy with mechanical systems, the acoustic impedance is now defined as $Z^{\mathrm{acoust}} = p/U$. For a system with a well-defined measurement aperture of area S, it is clear that $p = F/S$ and $U = vS$, so that $Z^{\mathrm{acoust}} = Z^{\mathrm{mech}}/S^2$, and once again we can define both the impedance at an aperture and the transfer impedance between two apertures. In what follows, the superscript "acoust" will be omitted for simplicity.

Consider first the case of a circular diaphragm under elastic tension and excited by an acoustic pressure $p\mathrm{e}^{\mathrm{i}\omega t}$ on one side. This problem can easily be solved using a simple differential equation [19.9], but here we seek a network analog. The pressure will deflect the dia-

Amazon
EX1013 Page 806

phragm into a nearly spherical-cap shape, displacing an air volume as it does so. If the small mass of the deflected air is neglected, then deflection of the diaphragm will be opposed by its own mass inertia, by the tension force and by internal losses. The mass contributes an acoustic inertance (or inductance in the network analog) with magnitude about $2\rho_s d/S$, where ρ_s is the density of the membrane material, d is the membrane thickness, and S is its area, while the tension force T contributes an equivalent capacitance of magnitude about $2S^2/T$. This predicts a resonance frequency $\omega^* \approx 0.5(T/S\rho_s d)^{1/2}$, while the rigorously derived frequency for the first mode of the diaphragm replaces the 0.5 with 0.38. Because biological membranes are not uniform in thickness and are usually loaded by neural transducers, however, it is not appropriate to worry unduly about the exact values of numerical factors. For a membrane there are also higher modes with nodal diameters and nodal circles, but these are of almost negligible importance in biological systems.

The next thing to consider is the behavior of cavities, apertures, and short tubes. The air enclosed in a rigid container of volume V can be compressed by the addition of a further small volume $\mathrm{d}V$ of air and the pressure will then rise by an amount $p_0\gamma\,\mathrm{d}V/V$ where p_0 is atmospheric pressure and $\gamma = 1.4$ is the ratio of specific heats of air at constant pressure and constant volume. The electric analog for this case is a capacitance of value $\gamma p_0/V = \rho c^2/V$ where c is the speed of sound in air and ρ is the density of air.

In the case of a tube of length l and cross-section S, both very much less than the wavelength of the sound, the air within the tube behaves like a mass ρlS that is set into motion by the pressure difference p across it. The acoustic inertance, or analog inductance, is then $Z = \rho l/S$. For a very short tube, such as a simple aperture, the motion induced in the air just outside the two ends must be considered as well, and this adds an "end-correction" of 0.6–0.8 times the tube radius [19.47] or about $0.4S^{1/2}$ to each end of the tube, thus increasing its effective length by twice this amount.

These elements can be combined simply by considering the fact that acoustic flow must be continuous and conserved. Thus a Helmholtz resonator, which consists of a rigid enclosure with a simple aperture, is modeled as a capacitor and an inductor in series, with some resistance added due to the viscosity of the air moving through the aperture and to a less extent to radiation resistance. The analog circuit is thus just as in Fig. 19.10b and leads to a simple resonance.

If any of the components in the system under analysis is not negligibly small compared with the wavelength of the sound involved, then it is necessary to use a more complex analysis. For the case of a uniform tube of length l and cross section S, the pressure and flow at one end can be denoted by p_1 and U_1 respectively and those at the other end by p_2 and U_2. The analog impedance is then a four-terminal element, as shown in Fig. 19.11. The equations describing the behavior of this element are

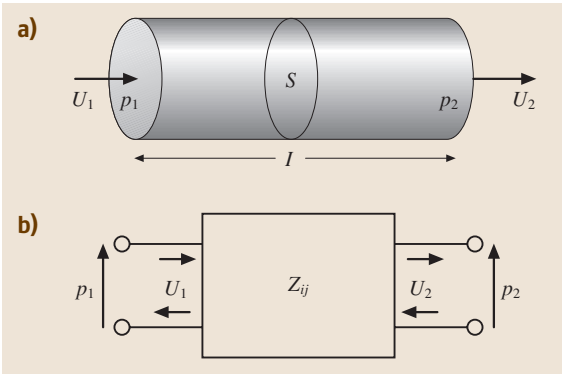


Fig. 19.11 **(a)** A simple uniform tube showing the acoustic pressures p_i and acoustic volume flows U_i. **(b)** The electrical analog impedance element Z_{ij}. Note that in both cases the current flow has been assumed to be antisymmetric rather than symmetric

$$p_1 = Z_{11}U_1 - Z_{12}U_2\,,$$
$$p_2 = Z_{21}U_1 - Z_{22}U_2\,. \qquad (19.8)$$

Note that the acoustic flow is taken to be inwards at the first aperture and outwards at the second, and this asymmetry gives rise to the minus signs in (19.8). Some texts take both currents to be inflowing, which removes the minus signs. For a simple uniform tube, symmetry demands that $Z_{21} = Z_{12}$ and that $Z_{22} = Z_{11}$. The first of these relations, known as reciprocity, is universal and does not depend upon symmetry, while the second is not true for asymmetric tubes or horns.

For a uniform tube of length l and cross-section S, consideration of wave propagation between the two ends shows that

$$Z_{11} = Z_{22} = -\mathrm{i}Z_0 \cot kl\,,$$
$$Z_{21} = Z_{12} = -\mathrm{i}Z_0 \csc kl\,, \qquad (19.9)$$

where $Z_0 = \rho c/S$ is the characteristic impedance of the tube. The propagation number k is given by

$$k = \frac{\omega}{c} - \mathrm{i}\alpha \qquad (19.10)$$

Amazon
EX1013 Page 807

where α represents the attenuation caused by viscous and thermal losses to the tube walls, and has the approximate value

$$\alpha \approx 10^{-5}\frac{\omega^{1/2}}{a} \qquad (19.11)$$

where a is the tube radius. Since auditory systems often involve quite narrow tubes, this term can become very important.

Similar but much more complicated relations apply for horns of various shapes, the simplest being conical, exponential or parabolic in profile [19.9]. As discussed elsewhere [19.47], horns act as acoustic transformers, basically because $Z_{22} \neq Z_{11}$, and can raise the acoustic pressure in the narrow throat by about the ratio of the mouth diameter to the throat diameter. This amplification is limited, however, to frequencies high enough that the horn diameter does not change greatly over an axial distance of about one wavelength.

With the aid of these electric circuit element analogs it is now possible to construct an analog network to represent the whole acoustic system, as shown by the example in Fig. 19.2. The major guiding principle in constructing such a network is the conservation of acoustic flow, or equivalently of electric current. In the case of an auditory system, there may be several openings connecting to the environment, and the acoustic pressures at these openings will generally differ in both phase and amplitude, depending upon the direction from which the sound is coming. Phase differences are generally more important for directional discrimination than are amplitude differences, particularly for small animals, just as in directional microphones, [19.47] and the combination of tubes, cavities and membranes in the system provides additional phase shifts. The sensory input is normally generated by the motion of tympanic membranes on the two sides of the animal, with the combination of external and internal pressure determining the motion of each membrane, as shown for example in Figs. 19.7 and 19.8.

When sound production is considered, it is usual to make the approximation of separating the basic sound-generation mechanism, such as the air flow through vibrating vocal folds, from the acoustics of the subsequent tubes and cavities, thus leading to a "source/filter model". With this simplification, the source can be assumed to produce an oscillating acoustic flow at a particular frequency ω, generally accompanied by a set of harmonics at frequencies $n\omega$ generated by nonlinearity in the vibration of the flow valve. The sound production at each frequency is then determined by the final acoustic flow $U(n\omega)$ through an aperture into the surroundings. The square of this flow, multiplied the acoustic radiation resistance of the aperture, which depends upon its size, [19.47] then determines the radiated acoustic power at that frequency. In a much more complex model, the pressure generated by the acoustic input flow can be recognized to act back upon the acoustic source itself, thus modifying its behavior. Typically this results in a small shift in the source frequency to align it more closely with a neighboring strong resonance of the filter. Such interactions are well known in other systems such as musical wind instruments and probably assist in coupling the oscillations of the syrinx to the resonance of the vocal filter in "pure tone" birdsong.

Finally, it is worthwhile to emphasize that biological acoustic systems do not consist of straight tubes of uniform diameter with rigid walls, unloaded circular membranes, or other idealized elements. While it is possible in principle to include such refinements in a computational model, this will generally make that model so complex to analyze that it reveals little about the actual operation of the system. It is usually better, therefore, to construct the simplest possible model that includes all the relevant acoustic elements and to examine its predictions and the effect that changing the dimensions of some of those elements has on the overall behavior. In this way the behavior of the real biological system can be largely understood.

Remark

The number of books and papers published in this field is very large, so that the reference list below is a rather eclectic selection. Generally descriptive and behavioral references have not been included, and the list concentrates on those that apply physical and mathematical analysis to bioacoustic systems.

References

19.1 W.C. Stebbins: *The Acoustic Sense of Animals* (Harvard Univ. Press, Cambridge 1983)

19.2 J.W. Bradbury, S.L. Vehrenkamp: *Principles of Animal Communication* (Sinauer, Sunderland 1998)

19.3 R.-G. Busnel (Ed.): *Acoustic Behavior of Animals* (Elsevier, New York 1963)

19.4 A.N. Popper, R.R. Fay (Eds.): *Comparative Studies of Hearing in Vertebrates* (Springer, Berlin, New York 1980)

Amazon
EX1013 Page 808

19.5 B. Lewis (Ed.): *Bioacoustics: A Comparative Approach* (Academic, London 1983)
19.6 D.B. Webster, R.R. Fay, A.N. Popper (Eds.): *The Evolutionary Biology of Hearing* (Springer, Berlin, New York 1992)
19.7 R.R. Fay, A.N. Popper (Eds.): *Comparative Hearing: Mammals* (Springer, Berlin, New York 1994)
19.8 R.R. Hoy, A.N. Popper, R.R. Fay (Eds.): *Comparative Hearing: Insects* (Springer, Berlin, New York 1998)
19.9 N.H. Fletcher: *Acoustic Systems in Biology* (Oxford Univ. Press, New York 1992)
19.10 D.A.W. Thompson: *On Growth and Form (abridged)*, 2nd edn., ed. by J.T. Bonner (Cambridge Univ. Press, Cambridge 1961)
19.11 G.B. West, J.H. Brown: Life's universal scaling laws, Phys. Today **57**(9), 36–42 (2004)
19.12 N.H. Fletcher: A simple frequency-scaling rule for animal communication, J. Acoust. Soc. Am. **115**, 2334–2338 (2004)
19.13 M. Garstang, D. Larom, R.Raspet, M. Lindeque: Atmospheric controls on elephant communication, J. Exper. Biol. **198**, 939–951 (1995)
19.14 K. McComb, D. Reby, L. Baker, C. Moss, S. Sayialel: Long-distance communication of acoustic cues to social identity in African elephants, Anim. Behav. **65**, 317–329 (2003)
19.15 J.H. Brackenbury: Power capabilities of the avian sound-producing system, J. Exp. Biol. **78**, 163–166 (1979)
19.16 A.W. Ewing: *Arthropod Bioacoustics: Neurobiology and Behavior* (Cornell Univ. Press, Ithaca 1989)
19.17 A. Michelsen, H. Nocke: Biophysical aspects of sound communication in insects, Adv. Insect Physiol. **10**, 247–296 (1974)
19.18 K. Kalmring, N. Elsner (Eds.): *Acoustic and Vibrational Communication in Insects* (Parey, Berlin, Hamburg 1985)
19.19 N.H. Fletcher: Acoustical response of hair receptors in insects, J. Comp. Physiol. **127**, 185–189 (1978)
19.20 N.H. Fletcher, S. Thwaites: Physical models for the analysis of acoustical systems in biology, Quart. Rev. Biophys. **12**(1), 25–65 (1979)
19.21 H.C. Bennet-Clark: Resonators in insect sound production: how insects produce loud pure-tone songs, J. Exp. Biol. **202**, 3347–3357 (1999)
19.22 B. Dumortier: Morphology of sound emission apparatus in arthropoda. In: *Acoustic Behaviour of Animals*, ed. by R.-G. Busnel (Elsevier, Amsterdam 1963) pp. 277–345
19.23 B. Dumortier: The physical characteristics of sound emission in arthropoda. In: *Acoustic Behaviour of Animals*, ed. by R.-G. Busnel (Elsevier, Amsterdam 1963) pp. 346–373
19.24 A.G. Daws, H.C. Bennet-Clark, N.H. Fletcher: The mechanism of tuning of the mole cricket singing burrow, Bioacoust. **7**, 81–117 (1996)
19.25 K. Ishizaka, J.L. Flanagan: Synthesis of voiced sounds from a two-mass model of the vocal cords, Bell Syst. Tech. J. **51**(6), 1233–1268 (1972)
19.26 N.H. Fletcher: Bird song – a quantitative acoustic model, J. Theor. Biol. **135**, 455–481 (1988)
19.27 H.L.F. Helmholtz: *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Dover, New York 1954), trans. A.J. Ellis
19.28 G. von Békésy: *Experiments in Hearing* (McGraw-Hill, New York 1960)
19.29 E.R. Lewis, E.L. Leverenz, W.S. Bialek: *The Vertebrate Inner Ear* (CRC, Boca Raton 1985)
19.30 N.H. Fletcher, S. Thwaites: Obliquely truncated simple horns: idealized models for vertebrate pinnae, Acustica **65**, 194–204 (1988)
19.31 P.M. Narins, A.S. Feng, W. Lin, H.-U. Schnitzler, A. Densinger, R.A. Suthers, C.-H. Xu: Old world frog and bird vocalizations contain prominent ultrasonic harmonics, J. Acoust. Soc. Am. **115**, 910–913 (2004)
19.32 C.H. Greenewalt: *Bird Song: Acoustics and Physiology* (Smithsonian Inst. Press, Washington 1968)
19.33 D.E. Kroodsma, E.H. Miller (Eds.): *Acoustic Communication in Birds*, Vol. 1-2 (Academic, New York 1982)
19.34 N.H. Fletcher, A. Tarnopolsky: Acoustics of the avian vocal tract, J. Acoust. Soc. Am. **105**, 35–49 (1999)
19.35 F. Goller, O.N. Larsen: A new mechanism of sound generation in songbirds, Proc. Nat. Acad. Sci. U.S.A. **94**, 14787–14791 (1997)
19.36 G.B. Mindlin, T.J. Gardner, F. Goller, R.A. Suthers: Experimental support for a model of birdsong production, Phys. Rev. E **68**, 041908 (2003)
19.37 D.K. Patterson, I.M. Pepperberg: A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels, J. Acoust. Soc. Am. **96**, 634–648 (1994)
19.38 G.J.L. Beckers, B.S. Nelson, R.A. Suthers: Vocal-tract filtering by lingual articulation in a parrot, Current Biol. **14**, 1592–1597 (2004)
19.39 O.N. Larsen, F. Goller: Role of syringeal vibrations in bird vocalizations, Proc. R. Soc. London B **266**, 1609–1615 (1999)
19.40 G.J.L. Beckers, R.A. Suthers, C. ten Cate: Pure-tone birdsong by resonant filtering of harmonic overtones, Proc. Nat. Acad. Sci. U.S.A. **100**, 7372–7376 (2003)
19.41 N.H. Fletcher, T. Riede, G.A. Beckers, R.A. Suthers: Vocal tract filtering and the "coo" of doves, J. Acoust. Soc. Am. **116**, 3750–3756 (2004)
19.42 C. Pytte, M.S. Ficken, A. Moiseff: Ultrasonic singing by the blue-throated hummingbird: A comparison between production and perception, J. Comp. Physiol. A **190**, 665–673 (2004)
19.43 G. Sales, D. Pye: *Ultrasonic Communication in Animals* (Chapman Hall, London 1974)

Amazon
EX1013 Page 809

19.44 D.J. Hartley, R.A. Suthers: The acoustics of the vocal tract in the horseshoe bat, Rhinolophus hildebrandti, J. Acoust. Soc. Am. **84**, 1201–1213 (1988)
19.45 A.D. Hawkins, A.A. Myrberg: Hearing and sound communication under water. In: *Bioacoustics: A Comparative Approach*, ed. by B. Lewis (Academic, London 1983) pp. 347–405
19.46 N.A.M. Schellart, A.N. Popper: Functional aspects of the evolution of the auditory system in Actinopterygian fish. In: *The Evolutionary Biology of Hearing*, ed. by D.B. Webster, R.R. Fay, A.N. Popper (Springer, Berlin, New York 1992) pp. 295–322
19.47 T.D. Rossing, N.H. Fletcher: *Principles of Vibration and Sound* (Springer, New York 2004), Chapts. 8–10

Amazon
EX1013 Page 810

20. Cetacean Acoustics

The mammalian order cetacea consist of dolphins and whales, animals that are found in all the oceans and seas of the world. A few species even inhabit fresh water lakes and rivers. A list of 80 species of cetaceans in a convenient table is presented by *Ridgway* [20.1]. These mammals vary considerably in size, from the largest living mammal, the large blue whale (*balaenoptera musculus*), to the very small harbor porpoise (*phocoena phocoena*) and Commerson's dolphin (*cephalorhynchus commersonnii*), which are typically slightly over a meter in length.

Cetaceans are subdivided into two suborders, odontoceti and mysticeti. Odontocetes are the toothed whales and dolphins, the largest being the sperm whale (*physeter catodon*), followed by the Baird's beaked whale (*berardius bairdii*) and the killer whale (*orcinus orca*). Within the suborder odontoceti there are four superfamilies: *platanistoidea*, *delphinoidea*, *ziphioidea*, and *physeteridea*. Over half of all cetaceans belong to the superfamily *delphinoidea*, consisting of seven species of medium whales and 35 species of small whales also known as dolphins and porpoises [20.1]. Dolphins generally have a sickle-shaped dorsal fin, conical teeth, and a long rostrum. Porpoises have a more triangular dorsal fin, more spade-shaped teeth, and a much shorter rostrum [20.1].

Mysticetes are toothless, and in the place of teeth they have rigid brush-like whalebone plate material called baleen hanging from their upper jaw. The baleen is used to strain shrimp, krill, micronekton, and zooplankton. All the great whales are mysticetes or baleen whales and all are quite large. The sperm and Baird s beaked whales are the only odontocetes that are larger than the smaller mysticetes such as Minke whales and pygmy right whales. Baleen whales are subdivided into four families, *balaenidae* (right and bowhead whales), *eschrichtiidae* (gray whales), *balaenopteridae* (Minke, sei, Bryde's, blue, fin, and humpback whales), and *neobalaenidae* (pygmy right whale).

Acoustics play a large role in the lives of cetaceans since sound travels underwater better than any other form of energy. Vision underwater is limited to tens of meters under the best conditions and less than a fraction of a meter in turbid and murky waters. Visibility is also limited by the lack of light at great depths during the day and at almost any depth on a moonless night. Sounds are used by marine mammals for myriad reasons such as group cohesion, group coordination, communications, mate selection, navigation and locating food. Sound is also used over different spatial scales from tens of km for some species and tens of meters for other species, emphasizing the fact that different species utilize sound in different ways. All odontocetes seem to have the capability to echolocate, while mysticetes do not echolocate except in a very broad sense, such as listening to their sound bouncing off the bottom, sea mounts, underwater canyon walls, and other large objects.

The general rule of thumb is that larger animals tend to emit lower-frequency sounds and the frequency range utilized by a specific species may be dictated more from anatomical constraints than any other factors. If resonance is involved with sound production, then anatomical dimensions become critical, that is, larger volumes resonate at lower frequencies than smaller volumes. The use of a particular frequency band will also have implications as to the distance other animals, including conspecifics, will be able to hear the sounds. Acoustic energy is lost in the propagation process by geometric spreading and absorption. Absorption losses are frequency dependent, increasing with frequency. Therefore, the sounds of baleen whales such as the blue whale that emit sounds with fundamental frequencies as low as 15 Hz can propagate to much longer distances than the whistles of dolphins that contain frequencies between 5 and 25 kHz.

20.1 **Hearing in Cetaceans** 806
20.1.1 Hearing Sensitivity of Odontocetes 807

Amazon
EX1013 Page 811

20.1.2 Directional Hearing in Dolphins .. 808
20.1.3 Hearing by Mysticetes 812
20.2 **Echolocation Signals** 813
20.2.1 Echolocation Signals of Dolphins that also Whistle 813
20.2.2 Echolocation Signals of Smaller Odontocetes that Do not Whistle . 817
20.2.3 Transmission Beam Pattern 819
20.3 **Odontocete Acoustic Communication** 821
20.3.1 Social Acoustic Signals 821
20.3.2 Signal Design Characteristics 823
20.4 **Acoustic Signals of Mysticetes** 827
20.4.1 Songs of Mysticete Whales 827
20.5 **Discussion** 830
References 831

20.1 Hearing in Cetaceans

One of the obvious adaptations for life in the sea is the absence of a pinna in cetaceans. The pinna probably disappeared through a natural selection process because it would obstruct the flow of water of a swimming animal and therefore be a source of noise. In the place of a pinna, there is a pin-hole on the surface of a dolphin's head which leads to a ligament inside the head, essentially rendering the pinna nonfunctional in conducting sounds to the middle ear. So, how does sounds enter into the heads of cetacean? Several electrophysiological experiments have been performed in which a hydrophone is held at different locations on an animal's head and the electrophysiological thresholds are determined as a function of the position of the hydrophone [20.2–4]. All three studies revealed greatest sensitivity on the dolphin's lower jaw.

The experimental configuration of *Møhl* et al. [20.2] is shown in Fig. 20.1a, which shows a *suction-cup hydrophone* and attached a bottlenose dolphin's lower jaw and the surface contact electrodes embedded in suction cups used to measure the auditory brainstem potential signals for different levels of sound intensity. The important differences in the experiment of *Møhl* et al. [20.2] are that the subject was trained and the measurements were done in air so that the sounds were limited to the immediate area where they were applied. The results of the experiment are shown in Fig. 20.1b. The location of maximum sensitivity to sound is slightly forward of the pan-bone area of the lower jaw, a location where *Norris* [20.5] hypothesized that sounds enter the head of a dolphin. Numerical simulation work by *Aroyan* [20.6] suggest that sounds enter the dolphin's head forward of the pan bone, but a good portion of the sound propagates below the skin to the pan bone and enters the head through the pan bone. The fact that sounds probably propagate through the lower jaw of dolphins and other

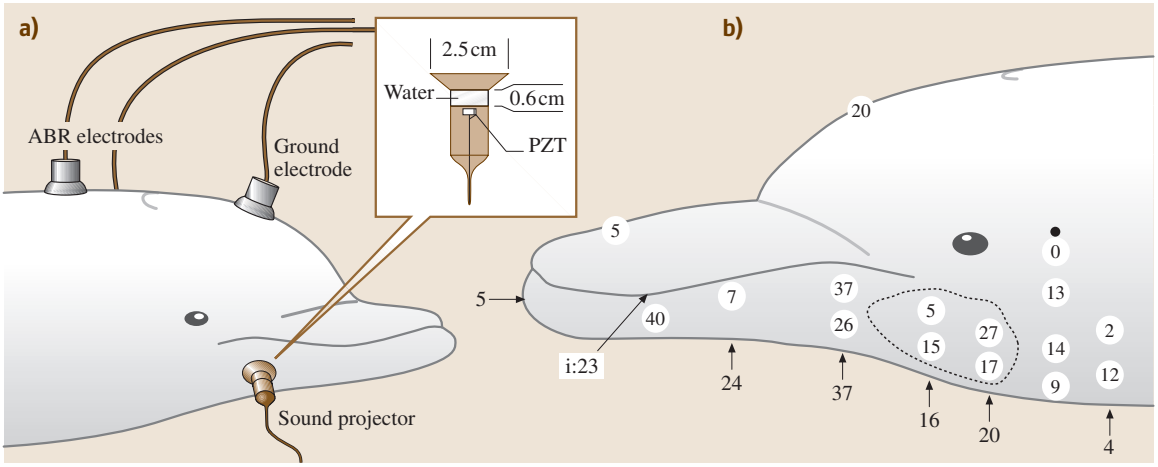


Fig. 20.1 **(a)** The experimental geometry used by *Møhl* et al. [20.2], **(b)** results of the auditory brainstem response (ABR) threshold measurements. The numerical values represent the amount of attenuation of the sound needed to obtain an ABR threshold. Therefore, the higher the number the more sensitive the location

Amazon
EX1013 Page 812

odontocetes does not necessarily mean that the same or a similar propagation process is occurring with baleen whales.

20.1.1 Hearing Sensitivity of Odontocetes

Almost all our knowledge of hearing in cetaceans comes from studies performed with small odontocetes. The most studied species is the Atlantic bottlenose dolphin (*tursiops truncatus*). Despite the amount of research performed with the bottlenose dolphin, our understanding of auditory processes in these animals lags considerably behind that for humans and other terrestrial mammals. There are still many large gaps in our knowledge of various auditory processes occurring within the most studied odontocetes. The first audiogram for a cetacean was measured by *Johnson* [20.9] for a *tursiops truncatus*. Since then, audiograms have been determined for the harbor porpoise (*phocoena phocoena*) by *Andersen* [20.10] and *Kastelein* et al. [20.11], the killer whale (*orcinus orca*) by *Hall* and *Johnson* [20.12] and *Szymanski* et al. [20.13], the beluga whale (*delphinapterus leucas*) by *White* et al. [20.14], the Pacific bottlenose dolphin (*tursiops gilli*) by *Ljungblad* et al. [20.15], the false killer whale (*pseudorca crassidens*) by *Thomas* et al. [20.16], the Chinese river dolphin (*lipotes vexillifer*) by *Wang* et al. [20.17], Risso's dolphins (*grampus griseus*) by *Nachtigall* et al. [20.18], the tucuxi (*sotalia fluviatilis*) by *Sauerland* and *Dehnhardt* [20.19], and the striped dolphin (*stenella coeruleoalba*) by *Kastelein* et al. [20.20]. The audiograms of these odontocetes are shown in Fig. 20.2. It is relatively striking to see how similar the audiograms are between species considering the vastly different habitats and areas of the world where some of these animals are found and the large differences in body size. All the audiograms suggest high-frequency hearing capabilities, with the smallest animal, *phocoena phocoena* having the highest hearing limit close to 180 kHz. However, the *orcinus orca*, which is over 95 times heavier and about six times longer can hear up to about 105 kHz. The actual threshold values shown in Fig. 20.2 should not be compared between species because the different methods of determining the threshold can lead to different results and because of the difficulties of obtaining good sound pressure level measurements in a reverberant environment. For example, *Kastelein* et al. [20.11] used a narrow-band frequency-modulated (FM) signal to avoid multi-path problems and the FM signals may provide additional cues not present in a pure-tone signal. Nevertheless, the audiograms shown in Fig. 20.2 suggest that all the animals had similar thresholds of 10–15 dB.

A summary of some important properties of the different audiograms depicted in Fig. 20.2 is given in Table 20.1. In the table, the frequency of best hearing is arbitrary defined as the frequency region in which the auditory sensitivity is within 10 dB of the maximum sensitivity depicted in each audiogram of Fig. 20.2. With the exception of the *Orcinus* and the *Lipotes*, the maximum sensitivity of the rest of the species represented

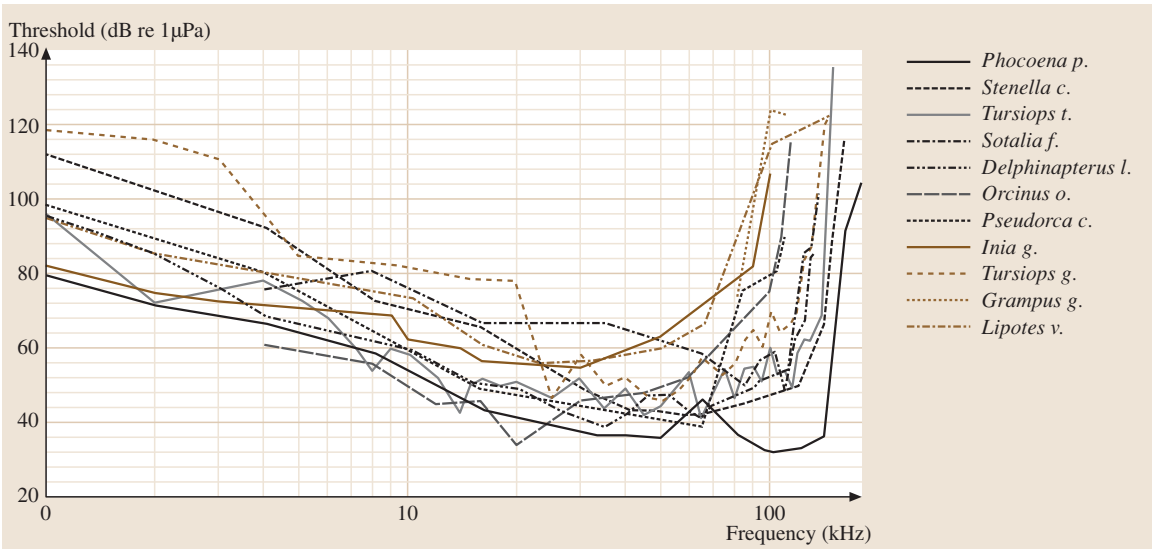


Fig. 20.2 Audiogram for different odontocetes species (after [20.7, 8])

Amazon
EX1013 Page 813

Table 20.1 Some important properties of the audiograms plotted in Fig. 20.2

Species	Maximum sensitivity (dB re 1 : Pa)	Frequency of best hearing (kHZ)	Upper frequency limit (kHz)
phocoena phocoena	32	18 – 130	180
stenella coeruleoalba	42	30 – 125	160
tursiops truncatus	42	15 – 110	150
tursiops gilli	47	30 – 80	135
sotalia fluviatilis	50	35 – 50	135
delphinapterus leucas	40	11 – 105	120
pseudorca crassidens	39	17 – 74	115
orcinus orca	34	15 – 40	110
inia geoffrensis	51	12 – 64	100
lipotes vexillifer	55	15 – 60	100
grampus griseus	–	–	100

in Fig. 20.2 and Table 20.1 are very similar, and within experimental uncertainties, especially for audiograms obtained with the staircase procedure using relatively large step sizes of 5 dB or greater. At the frequency of best hearing, the threshold for *Orcinus* and *Phocoena* are much lower than for the other animals. It is not clear whether this keen sensitivity is a reflection of a real difference or a result of some experimental artifact. The data of Table 20.1 also indicate that *Phocoena phocoena*, *Stenella coeruleoalba* and *Tursiops truncatus* seem to have the widest auditory bandwidth.

20.1.2 Directional Hearing in Dolphins

Sound Localization

The ability to localize or determine the position of a sound source is important in order to navigate, detect prey and avoid predators and avoid hazards producing an acoustic signatures. The capability to localize sounds has been studied extensively in humans and in many vertebrates (see [20.21]). Lord Rayleigh in 1907 proposed that humans localized in the horizontal plane by using interaural time differences for low-frequency sounds and interaural intensity differences for high-frequency sounds. The sound localization acuity of a subject is normally defined in terms of a minimum audible angle (MAA), defined as the angle subtended at the subject by two sound sources, one being at a reference azimuth, at which the subject can just discriminate the sound sources as two discrete sources [20.22]. If the sound sources are separated symmetrically about a midline, the MAA is one half the threshold angular separation between the sound sources. If one of the sound sources is located at the midline, then the MAA is the same as the threshold angular separation between the two sources.

Renaud and *Popper* [20.23] examined the sound localization capabilities of a *tursiops truncatus* by measuring the MAA in both the horizontal and vertical planes. During a test trial the dolphin was required to station on a bite plate facing two transducers positioned at equal angles from an imaginary line running through the center of the bite plate. An acoustic signal was then transmitted from one of the transducers and the dolphin was required to swim and touch the paddle on the same side as the emitting transducer. The angle between the transducers was varied in a modified staircase fashion. If the dolphin was correct for two consecutive trials, the transducers were moved an incremental distance closer together, decreasing the angle between the transducer by 0.5°. After each incorrect trial, the transducers were moved an incremental distance apart, increasing the angle between the transducers by 0.5°. This modified staircase procedure allowed threshold determination at the 70% level. The threshold angle was determined by averaging a minimum of 16 reversals. A randomized schedule for sound presentation through the right or left transducer was used.

The localization threshold determined in the horizontal and vertical planes as a function of frequency for narrow-band pure-tone signals is shown in Fig. 20.3. The MAA had a U-shaped pattern, with a large value of 3.6° at 6 kHz, decreasing to a minimum of 2.1° at 20 kHz and then slowly increasing in an irregular fashion to 3.8° at 100 kHz. MAAs for humans vary between 1° and 3.7°, with the minimum at a frequency of approximately 700 Hz [20.24]. The region where the MAA decreased to a minimum in Figs. 20.19 and 20.20 (about 20 kHz) may be close to the frequency at which the dolphin switches from using interaural time difference cues to interaural intensity difference cues. The MAAs for the

Amazon
EX1013 Page 814

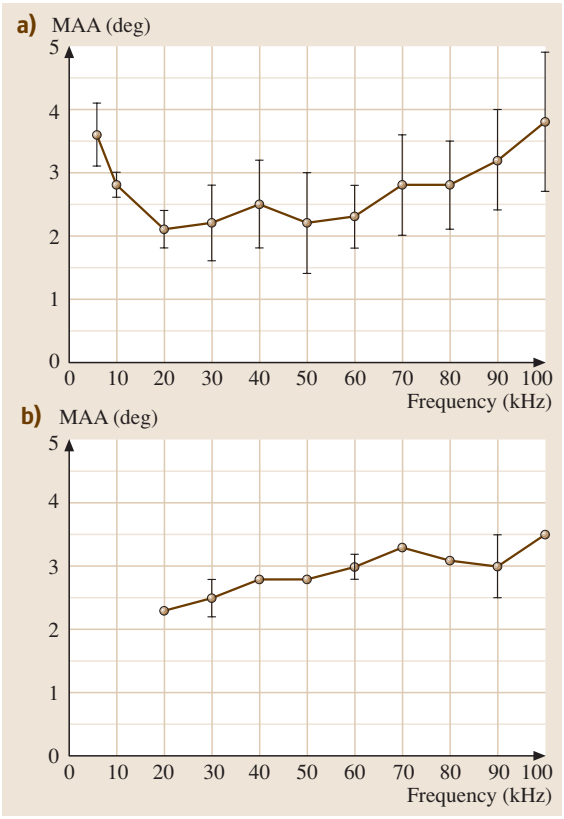


Fig. 20.3 **(a)** Localization threshold determined in the horizontal plane as a function of frequency. The mean ± one standard deviation is shown for seven determinations per frequency. The animal faced directly ahead at 0° azimuth. **(b)** Localization threshold determined in the vertical plane as a function of frequency. Standard deviation are indicated for 30, 60 and 90 kHz vertical data (seven sessions each). The dolphin's azimuth was 0° (after *Renauld* and *Popper*, [20.27]) Figs. 20.19 and 20.20

bottlenose dolphin were smaller than the MAAs (in the horizontal plane) of 3.5° at 3.5 kHz and 6° at 6 kHz for a harbor porpoise measured by *Dudok van Heel* [20.25], and 3° at 2 kHz, measured by *Andersen* [20.26] also for a harbor porpoise.

In order to measure the MAA in the vertical plane, the dolphin was trained to turn to its side (rotate its body 90° along its longitudinal axis) and to bite on a sheet of plexiglass which was used as the stationing device. The MAA in the vertical plane varied between 2.3° at 20 kHz to 3.5° at 100 kHz. These results indicate that the animal could localize in the vertical plane nearly as well as in the horizontal plane. These results were not expected since binaural affects, whether interaural time or intensity, should not be present in the vertical plane. However, the dolphin's ability to localize sounds in the vertical plane may be explained in part by the asymmetry in the receive beam patterns in the vertical plane discussed in the next section.

Renaud and *Popper* [20.23] also determined the MAA for a broadband transient signal or click signal, having a peak frequency of 64 kHz and presented to the dolphin at a repetition rate of 333 Hz. The MAA in the horizontal plane was found to be approximately 0.9° and 0.7° in the vertical plane. It is not surprising that a broadband signal should result in a lower MAA than a pure-tone signal of the same frequency as the peak frequency of the click. The short onset time and the broad frequency spectrum of a click signal should provide additional cues for localization.

Receiving Beam Patterns

Having narrow transmission and reception beams allows the dolphin to localize objects in a three-dimensional volume, spatially separate objects within a multi-object field, resolve features of extended or elongated objects, and to minimize the amount of interference received. The amount of ambient noise from an isotropic noise field and the amount of reverberation interference received is directly proportional to the width of the receiving beam. The effects of discrete or partially extended interfering noise or reverberant sources can be minimized by simply directing the beams away from the sources.

The receiving beam pattern of a dolphin was determined by measuring the masked hearing threshold as a function of the azimuth about the animal's head. The relative masked hearing threshold as a function of azimuth is equivalent to the received beam pattern since the receiving beam pattern is the spatial pattern of hearing sensitivity. *Au* and *Moore* [20.28] measured the dolphin's masked hearing threshold as the position of either the noise or signal sources varied in their angular position about the animal's head. The dolphin was required to voluntarily assume a stationary position on a bite plate constructed out of a polystyrene plastic material. The noise and signal transducers were positioned along an arc with the center of the arc located approximately at the pan bone of the animal's lower jaw. In order to measure the dolphin's receiving beam in the vertical plane, the animal was trained to turn onto its side before biting the specially designed vertical bite-plate stationing device. For the measurement in the vertical plane, the position of the signal transducer was fixed directly in

Amazon
EX1013 Page 815

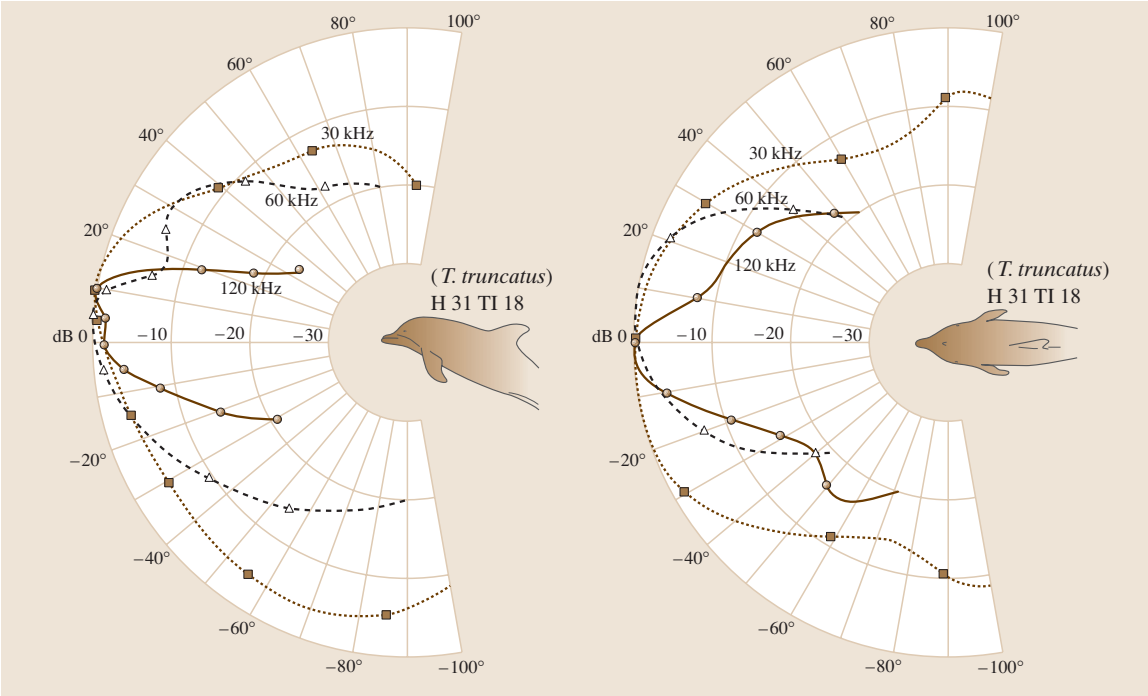


Fig. 20.4 Receive beam patterns in the vertical and horizontal planes for frequencies of 30, 60 and 120 kHz. The relative masked thresholds as a function of the elevation angle of the asking noise source are plotted for each signal frequency (after *Au* and *Moore*, [20.28])

line with the bite plate and its acoustic output was held constant. Masked thresholds were measured for different angular position of the noise transducer along the arc. The level of the noise was varied in order to obtain the masked threshold. A threshold estimate was considered complete when at least 20 reversals (10 per session) at a test angle had been obtained over at least two consecutive sessions, and if the average reversal values of the two sessions were within 3 dB. After a threshold estimate was achieved, the noise transducer was moved to a new azimuth over a set of randomly predesignated azimuths. As the azimuth about the dolphin's head increased, the hearing sensitivity of the dolphin tended to decrease, requiring higher levels of masking noise from a transducer located at that azimuth to mask the signal from a source located directly ahead of the animal.

The receiving beam patterns in both the vertical and horizontal plane are plotted for signal frequencies of 30, 60 and 120 kHz in Fig. 20.4. The radial axis of Fig. 20.4 represents the difference in dB between the noise level needed to mask the test signal at any azimuth and the minimum noise level needed to mask the test signal at the azimuth corresponding to the major axis of the vertical beam. The shape of the beams in Fig. 20.4 indicates that the patterns were dependent on frequency, becoming narrower, or more directional as the frequency increased. The beam of a planar hydrophone also becomes narrower as frequency increases. The 3 dB beam widths were approximately 30.4°, 22.7°, and 17.0° for frequencies of 30, 60, and 120 kHz, respectively. There was also an asymmetry between the portion of the beam above and below the dolphin's head. The shape of the beams dropped off more rapidly as the angle above the animal's head increased than for angles below the animal's head, indicating a more rapid decrease in the animal's hearing sensitivity for angles above the head than for angles below the head. If the dolphin receives sounds through the lower jaw, the more rapid reduction in hearing sensitivity for angles above the head may have been caused by shadowing of the received sound by the upper portion of the head structure including air in the nasal sacs [20.29]. There is a slight peculiarity in the 60 kHz beam which shows almost the same masked threshold values for 15° and 25° elevation angles.

The radial line passing through the angle of maximum sensitivity is commonly referred to as the major

Amazon
EX1013 Page 816

axis of the beam. The major axis of the vertical beams is elevated between 5° and 10° above the reference axis. The 30 and 120 kHz results show the major axis at 10° while the 60 kHz results showed the major axis at 5°. It will be shown in a later section that the major axis of the received beam in the vertical plane is elevated at approximately the same angle as the major axis of the transmitted beam in the vertical plane.

In the horizontal beam pattern measurement, two noise sources were fixed at azimuth angles of ±20°. The level of the noise sources was also fixed. The position of the signal transducer was varied from session to session. Masked thresholds were determined as a function of the azimuth of the signal transducer, by varying the signal level of the signal transducer in a staircase fashion. A threshold estimate was considered completed when at least 20 reversals at a test angle had been obtained over at least two consecutive sessions, if the average reversal values were within 3 dB of each other. After a threshold estimate was determined, the signal transducer was moved to a new azimuth over a set of randomly predesignated azimuths. Two noise sources were used in order to discourage the dolphin from internally steering its beam in the horizontal plane. If the animal could steer its beam, it would receive more noise from one of the two hydrophones, and therefore not experience any improvement in the signal-to-noise ratio. The masking noise from the two sources was uncorrelated but equal in amplitude. The radial axis represents the difference in dB between the signal level at the masked threshold for the various azimuths and the signal level of the masked threshold for 0° azimuth (along the major axis of the horizontal beam). The horizontal receiving beams were directed forward with the major axis being parallel to the longitudinal axis of the dolphin. The beams were nearly symmetrical about the major axis. Any asymmetry was within the margin of experimental error involved in estimating the relative thresholds. The horizontal beam patterns exhibited a similar frequency dependence as the vertical beam patterns, becoming narrower or more directional as the frequency increased. The 3 dB beam widths were 59.1°, 32.0°, and 13.7° for frequencies of 30, 60, and 120 kHz, respectively.

Zaytseva et al. [20.30] measured the horizontal beam pattern of a dolphin by measuring the masked hearing threshold as a function of azimuth. Their beam width of 8.2° for a frequency of 80 kHz was much narrower than the 13.7° for a frequency of 120 kHz. The difference in beam width is even larger if the results of *Au* and *Moore* [20.28] are linearly interpolated to 80 kHz. We calculated an interpolated beam width of 25.9° at 80 kHz, which was considerably greater than the 8.2° obtained by *Zaytseva* et al. [20.30]. The difference in beam width measured by Zaytseva et al. and Au and Moore may be attributed to the use of only one noise source by Zaytseva et al. compared to the two noise sources used by Au and Moore. With a single masking noise source in the horizontal plane, there is the possibility of the animal performing a spatial filtering operation by internally steering the axis of its beam in order to maximize the signal-to-noise ratio. Another possibility is that Zaytseva et al. did not use a fixed stationing device. Rather, the dolphin approached the signal hydrophone from a start line, always oriented in the direction of the signal hydrophone. The animal responded to the presence or absence of a signal by either swimming or not swimming to the hydrophone. In such a procedure, it is impossible to control the orientation of the animal's head with respect to the noise masker so that the dolphin could move its head to minimize the effects of the noise.

Directivity Index

The directivity index is a measure of the sharpness of the beam or major lobe of either a receiving or transmitting beam pattern. For a spherical coordinate system, the directivity index of a transducer is given by the equation [20.31]

$$\mathrm{DI} = 10\log\frac{4\pi}{\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\left(\frac{p(\theta,\phi)}{p_0}\right)^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi}\,. \qquad (20.1)$$

Although the expression for directivity index is relatively simple, using it to obtain numerical values can be quite involved unless transducers of relatively simple shapes (cylinders, lines and circular apertures) with symmetry about one axis is involved. Otherwise, the beam pattern needs to be measured as a function of both θ, and ϕ. This can be done by choosing various discrete values of θ and measuring the beam pattern as a function of ϕ, a tedious process. Equation (20.1) can then be evaluated by numerically evaluating the double integral with a digital computer. The directivity indices associated with the dolphin's beam patterns in Fig. 20.4 were estimated by *Au* and *Moore* [20.28] using (20.1) and a two-dimensional Simpson's 1/3-rule algorithm [20.32]. The results of the numerical evaluation of are plotted as a function of frequency in Fig. 20.5. DIs of 10.4, 15.3 and 20.6 dB were obtained for frequencies of 30, 60,

Amazon
EX1013 Page 817

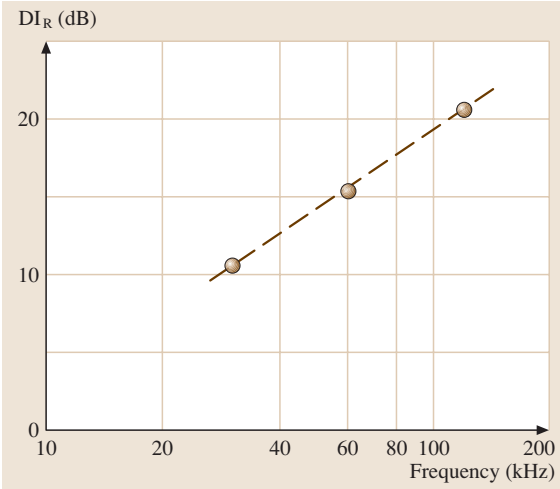


Fig. 20.5 Receiving directivity index as function of frequency for a *tursiops truncatus*

and 120 kHz, respectively. A linear curve fitted to the computed DIs in a least-square-error manner is also shown in the Fig. 20.5. The equation of the line is

$$\mathrm{DI(dB)} = 16.9 \log f(\mathrm{kHz}) - 14.5 \,. \tag{20.2}$$

The results of Fig. 20.5 indicate that the dolphin's receive directivity index increased with frequency in a manner similar to that of a linear transducer (Bobber, 1970). The expression in (20.2) is only valid for frequencies at which DI(dB) is greater or equal to 0.

Although the directivity index expressed by (20.1) is for a *tursiops*, it can also be used to estimate the directivity index of other dolphins by applying an appropriate correction factor, so that (20.2) can be rewritten as

$$\mathrm{DI(dB)} = 16.9 \log f(\mathrm{kHz}) - 14.5 + \mathrm{CF(dB)} \,, \tag{20.3}$$

where CF(dB) is a correction factor taking into account different head sizes. The directivity index of a planar circular plate is proportional to its diameter so if we let d_T be the diameter of the head of a *tursiops* at about the location of the blowhole and d_D be the diameter of the head of a different species of dolphin, then the correction factor can be expressed as

$$\mathrm{CF(dB)} = 20 \log(d_\mathrm{D}/d_\mathrm{T}) \,. \tag{20.4}$$

The correction factor will be positive for a dolphin with a larger head and negative for a dolphin with a smaller head than *tursiops*.

20.1.3 Hearing by Mysticetes

Our knowledge of the hearing characteristics of baleen whales is extremely limited. We do not know how they receive sounds, the frequency range of hearing, and the sensitivity of hearing at any frequency. Much of our understanding of hearing in baleen whales comes from anatomical studies of the ears of different species. Baleen whales have occluded external auditory canals that are filled with a homogeneous wax [20.33]. The lower jaws of mysticetes are designed for sieving or gulp feeding and have no evident connection to the temporal bones [20.33] making it very difficult to understand how sounds enter into the ears of these whales.

Various types of whales have been observed by many investigators to react strongly and drastically change their behavior in the presence of boats and low-flying aircraft, however, the sound pressure levels at the whales' locations are often difficult to define and measure. In some situations, the sound levels of the aversive sound could be estimated and a data point obtained. Bowhead whales were observed fleeing from a 13 m diesel-powered boat having a noise level at the location of the whales of about 84 dB re 1 μPa in the dominant 1/3-octave band, or about 6 dB above the ambient noise in that band [20.34]. Playback experiments with bowhead whales indicated that they took evasive action when the noise was about 110 dB re 1 μPa or 30 dB above the ambient noise in the same 1/3-octave band [20.35]. Playback experiments with migrating gray whales indicated that about 10% of the whales made avoidance behavior when the noise was about 110 dB in a 1/3 octave band, 50% at about 117 dB and 90% at about 122 dB or greater. These playback signals consisted of anthropogenic noise associated with the oil and gas industry.

Frankel [20.36] played back natural humpback whale sounds and a synthetic sound to humpback whales wintering in the waters of the Hawaiian islands. Twenty seven of the 1433 trials produced rapid approach response. Most of the responses were to the feeding call. Feeding call and social sounds produced changes in both separation and whale speed, indicating that these sounds can alter a whale's behavior. The humpback whales responded to sounds as low as 102–105 dB but the strongest responses occurred when the sounds were 111 to 114 dB re 1 μPa.

All of the playback experiments suggest that sounds must be between 85 and 120 dB before whales will react to them. These levels are very high compared to those that dolphins can hear and may suggest that it is very

Amazon
EX1013 Page 818

difficult to relate reaction to hearing sensitivity. Whales may not be reacting strongly unless the sounds are much higher than their hearing threshold.

Determining the hearing sensitivity or audiogram of baleen whales represents an extremely difficult challenge and will probably require the use of some sort of electrophysiological technique as suggested by *Ridgway* et al. [20.37]. Perhaps a technique measuring auditory-evoked potentials with beached whales may provide a way to estimate hearing sensitivity. Even with evoked potential measurements, there are many issues that have to be considered. For example, if an airborne source is used, the results cannot be translated directly to the underwater situation. If a sound source is placed on a whale's head, relating that to a whale receiving a plane wave will also not be simple. Measurement of evoked potentials themselves may not be simple because of the amount of flesh, muscles and blubber that the brain waves would have to travel through in order to reach the measurement electrodes. *Ridgway* and *Carder* [20.38] reported on some attempts to use the evoked potential method with a young gray whale held at Sea World in California. They also showed that the pygmy whale auditory system was most sensitive to very high frequencies (120–130 kHz) in the same range as their narrow-band echolocation pulses. A young sperm whale was most sensitive in the 10–20 kHz region of the spectrum [20.38].

20.2 Echolocation Signals

Echolocation is the process in which an organism projects acoustic signals and obtains a sense of its surroundings from the echoes it receives. In a general sense, any animal with a capability to hear sounds can echolocate by emitting sounds and listening to the echoes. A person in an empty room can gain an idea of the size and shape of the room by emitting sounds and listening to the echoes from the different walls. However, in this chapter, echolocation is used in a more specific sense in which an animal has a very specialized capability to determine the presence of objects considerably smaller than itself, discriminate between various objects, recognize specific objects and localize objects in three-dimensional space (determine range and azimuth). Dolphins and bats have this specialized capability of echolocation.

The echolocation system of a dolphin can be broken down into three subsystems: the transmission, reception and signal processing/decision subsystems. The reception system has to do with hearing and localization. The transmission system consist of the sound production mechanism, acoustic propagation from within the head of the dolphin to into the water, and the characteristics of the signals traveling in the surrounding environment. The third subsystem has to do with processing of auditory information by the peripheral and central nervous system. The capability of a dolphin to detect objects in noise and in clutter, and to discriminate between various objects depends to a large extent on the information-carrying capabilities of the emitted signals.

Dolphins most likely produce sounds within their nasal system and the signals are projected out through the melon. Although there has been a long-standing controversy on whether sounds are produced in the larynx or in the nasal system of odontocetes, almost all experimental data with dolphins indicate that sounds are produced in the nasal system [20.39]. For example, *Ridgway* and *Carder* [20.40] used catheters accepted into the nasal cavity by trained belugas. The belugas preformed an echolocation task in open water, detecting targets and reported the presence of targets by whistling. Air pressure within the nasal cavity was shown to be essential for echolocation and for whistling [20.40].

The melon immediately in front of the nasal plug may play a role in channeling sounds into the water, a notion first introduced by *Wood* [20.41]. *Norris* and *Harvey* [20.42] found a low-velocity core extending from just below the anterior surface towards the right nasal plug, and a graded outer shell of high-velocity tissue. Such a velocity gradient could channel signals originating in the nasal region in both the vertical and horizontal planes. Using both a two-dimensional [20.43] and a three-dimensional model [20.6] to study sound propagation in a dolphin's head, Aroyan has shown that echolocation signals most likely are generated in the nasal system and are channeled into the water by the melon. *Cranford* [20.44] has also collected evidence from nasal endoscopy of trained echolocating dolphins that suggest that echolocation signals are most likely produced in the nasal system at the location of the monkey-lips, dorsal bursae complex just beneath the blow hole.

Amazon
EX1013 Page 819

20.2.1 Echolocation Signals of Dolphins that also Whistle

Most dolphin species are able to produce whistle signals. Among some of the species in this category in which echolocation signals have been measured include the bottlenose dolphin (*tursiops sp*), beluga whale (*delphinapterus leucas*), killer whale (*orcinus orca*), false killer whale (*pseudorca crassidens*), Pacific white-sided dolphin (*lagenorhynchus obliquidens*), Amazon river dolphin (*inia geoffrensis*), Risso's dolphin (*grampus griseus*), tucuxi (*sotalia fluviatilis*), Atlantic spotted dolphin (*stenella frontalis*), Pacific spotted dolphin (*Stenella attenuata*), spinner dolphin (*Stenella longirostris*), pilot whale (*globicephala sp*), rough-toothed dolphin (*steno bredanesis*), Chinese river dolphin (*lipotes vexillifer*) and sperm whales (*physeter catodon*). However, most of the available data have been obtained for three species: the bottlenose dolphin, the beluga whale and the false killer whale.

Prior to 1973, most echolocation signals of *tursiops* were measured in relatively small tanks and the results provided a completely different picture of what we currently understand. It was not until the study of *Au* et al. [20.45] that certain features of biosonar signals used by *tursiops* and other dolphins in open waters were discovered. We discovered that the signals had peak frequencies between 120 and 130 kHz, over an octave higher than previously reported peak frequencies between 30 and 60 kHz [20.46]. We also measured an average peak-to-peak click source level on the order of 220 dB re 1 μPa at 1 m, which represents a level 30 to 50 dB higher than previously measured for *tursiops*. Examples of typical echolocation signals emitted by *tursiops truncatus* are shown in Fig. 20.6 for two situations. The top signal is typical of signals used in the open waters of Kaneohe Bay, Oahu, Hawaii, and the second signal represents typical signals for a *tursiops* in a tank. Signals measured in Kaneohe Bay regularly have duration between 40 and 70 μs, 4 to 10 positive excursion, peak frequencies between 110 and 130 kHz and peak-to-peak source levels between 210 and 228 dB re 1 μPa. The signals in Fig. 20.6 are not drawn to scale; if they were, the tank signal would resemble a flat line.

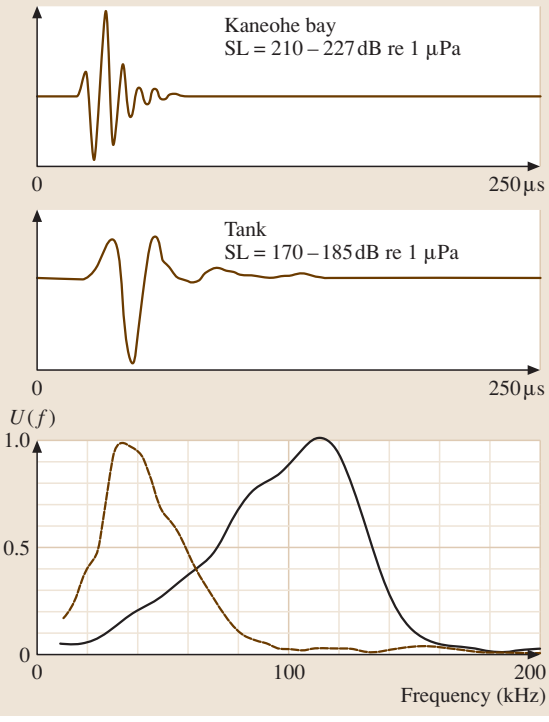


Fig. 20.6 Example of echolocation signals used by *tursiops truncatus* in Kaneohe Bay (after *Au* [20.45]), and in a tank

Au et al. [20.47] postulated that high-frequency echolocation signals were a byproduct of the animals producing high-intensity clicks to overcome snapping shrimp noise. In other words, dolphins can only emit high-level clicks (greater than 210 dB) if they use high frequencies. The effects of a noisy environment on the echolocation signals used by a beluga or white whale (*Delphinapterus leucas*) was vividly demonstrated by *Au* et al. [20.47]. The echolocation signal of a beluga was measured in San Diego Bay, California before the whale was moved to Kaneohe Bay. The ambient noise in Kaneohe Bay is between 15 to 20 dB greater than in San Diego Bay. The whale emitted echolocation signals with peak frequencies between 40 and 60 kHz and with a maximum averaged peak-to-peak source level of 202 dB re 1 μPa in San Diego Bay. In Kaneohe Bay, the whale shifted the peak frequency of its signals over an octave higher to 100 and 120 kHz. The source level also increased to over 210 dB re 1 μPa [20.47]. Examples of typical echolocation signals used by the whale in San Diego and in Kaneohe Bay are shown in Fig. 20.7a. Here, the signals are drawn to scale with respect to each other. Echolocation signals used by belugas in tanks also resemble the low-frequency signals shown in Fig. 20.2a [20.48, 49]. *Turl* et al. [20.50] measured the sonar signals of a beluga in a target-in-clutter detection task in San Diego Bay and found the animal used high-frequency (peak frequency above 100 kHz) and high-intensity (greater than 210 dB re 1 μPa) signals. Therefore low-amplitude clicks of the beluga had

Amazon
EX1013 Page 820

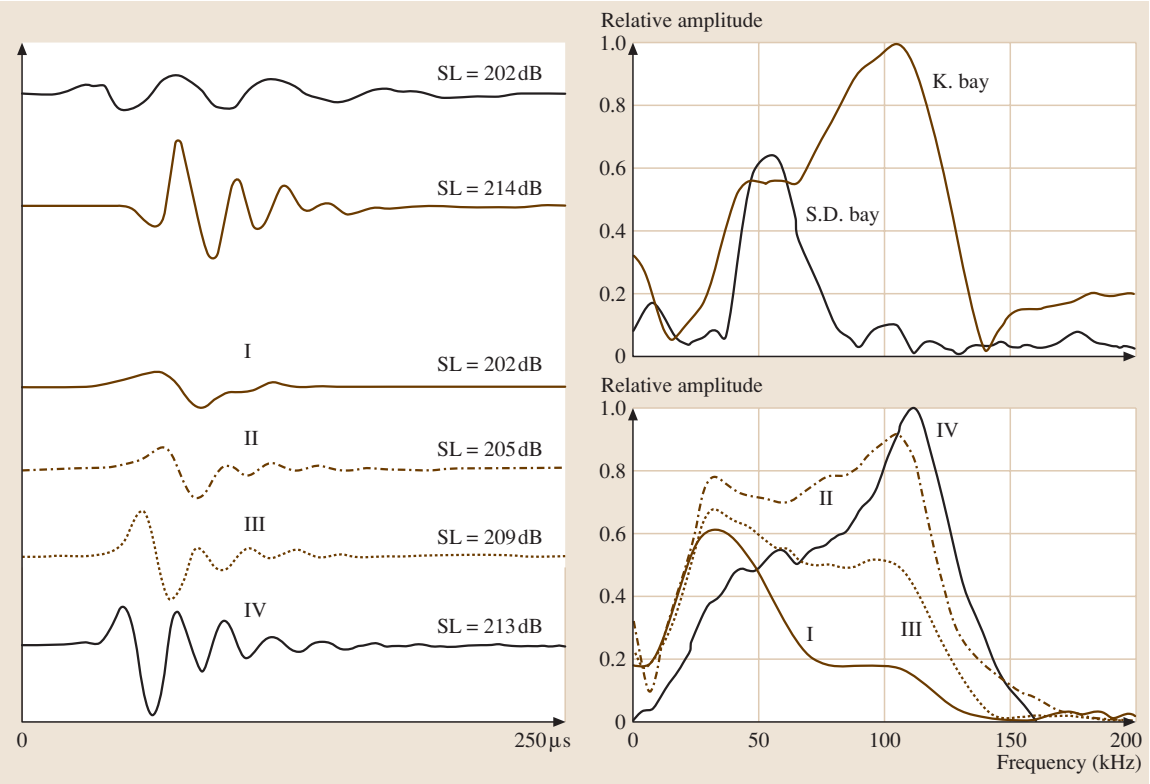


Fig. 20.7 **(a)** Example of beluga echolocation signals measured in San Diego Bay and in Kaneohe Bay (after *Au* et al. [20.47]); **(b)** Examples of *pseudorca* echolocation signals, Sl is the averaged peak-to-peak source level (after *Au* et al. [20.52])

low peak frequencies and the high-amplitude clicks had high peak frequencies. The data of *Moore* and *Pawloski* [20.51] for *Tursiops* also seem to support the notion that the shape of the signal in the frequency domain is related to the intensity of the signal.

Recent results with a false killer whale showed a clear relationship between the frequency content of echolocation signals and source level [20.52]. The *Pseudorca* emitted four basic types of signals, which are shown in Fig. 20.7b. The four signal types have spectra that are bimodal (having two peaks); the spectra in Fig. 20.2 are also bimodal. The type I signals were defined as those with the low-frequency peak (< 70 kHz) being the primary peak and the high-frequency peak being the secondary peak with its amplitude at least 3 dB below that of the primary peak. Type II signals were defined as those with a low-frequency primary peak and a high-frequency secondary peak having an amplitude within 3 dB of the primary peak. Type III signals were those with a high-frequency primary peak (>70 kHz), and a low-frequency secondary peak having an amplitude within 3 dB of the primary peak. Finally, type IV signals were those with a high-frequency primary peak having an amplitude that was at least 3 dB higher than that of the secondary low-frequency peak.

The data of *Thomas* et al. [20.53, 54] also indicated a similar relationship between intensity and the spectrum of the signal. The echolocation signals of a *pseudorca* measured in a tank had peak frequencies between 20 and 60 kHz and source levels of approximately 180 dB re 1 µPa [20.53]. Most of the sonar signals used by another *pseudorca* performing a detection task in the open waters of Kaneohe Bay had peak frequencies between 100 and 110 kHz and source levels between 220 and 225 dB re 1 µPa [20.54].

A bimodal spectrum is best described by its center frequency, which is defined as the centroid of the spectrum, and is the frequency which divides the spectrum into two parts of equal energy. From Fig. 20.7, we can see that, as the source level of the signal in-

Amazon
EX1013 Page 821