The sound field is produced by a circular disc of piezoelectric material glued to the bottom of the container. The bright spot in the upper part of the container is the light from the bubble. It is of blueish white color. The needle sticking out into the container in the upper left part of the figure is a platinum wire for producing a bubble via a current pulse. The bubble is then driven via acoustic radiation forces into its position where it is trapped almost forever. The bubble exerts large oscillations that can be photographed [8.118]. Figure 8.17 shows a sequence of the bubble oscillation of a light-emitting bubble. A full cycle is shown after which the oscillation repeats. Indeed, this repetition has been used to sample the series of photographs by shifting the instant of exposure by 500 ns from frame to frame with respect to the phase of the sound wave. The camera is looking directly into the illuminating light source (back illumination). The bubble then appears dark on a bright background because the light is deflected off its surface. The white spot in the middle of the bubble results from the light passing the spherical bubble undeflected.

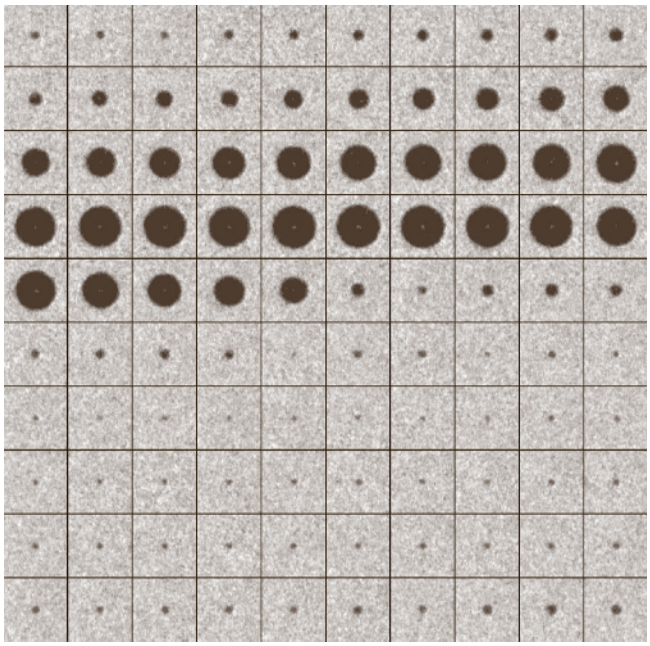


Fig. 8.17 Photographic series of a trapped, sonoluminescing bubble driven at 21.4 kHz and a pressure amplitude of 132 kPa. The interframe time is 500 ns. Picture size is 160 × 160 µm. (Courtesy of R. Geisler)

Upon the first large collapse a shock wave is radiated (Fig. 8.18). At the 30 ns interframe time the 12 frames given cover 330 ns, less than one interframe time Fig. 8.17. The shock wave becomes visible via deflection of the illuminating light. The bubble expands on a much slower time scale than the speed of the shock wave and is seen as a tiny black spot growing in the middle of the ring of the shock wave. Measurements of the shock velocity near the bubble at collapse (average from 6 to 73 µm) leads to about 2000 m/s, giving a shock pressure of about 5500 bar [8.119]. Higher-resolution measurements have increased these values to about 4000 m/s and 60 kbar [8.120].

The bubble oscillation as measured photographically can be compared with theoretical models for the motion of a bubble in a liquid. A comparison is shown in Fig. 8.19. The overall curve is reproduced well, although the steep collapse is not resolved experimentally. For details the reader is referred to [8.81].

There is the question whether shock waves are also emitted into the interior of the bubble [8.121–124]. Theoretically, shock waves occur in certain models, when the collapse is of sufficient intensity. Recently, the problem has been approached by molecular dynamics calculations [8.125–128]. As the number of molecules

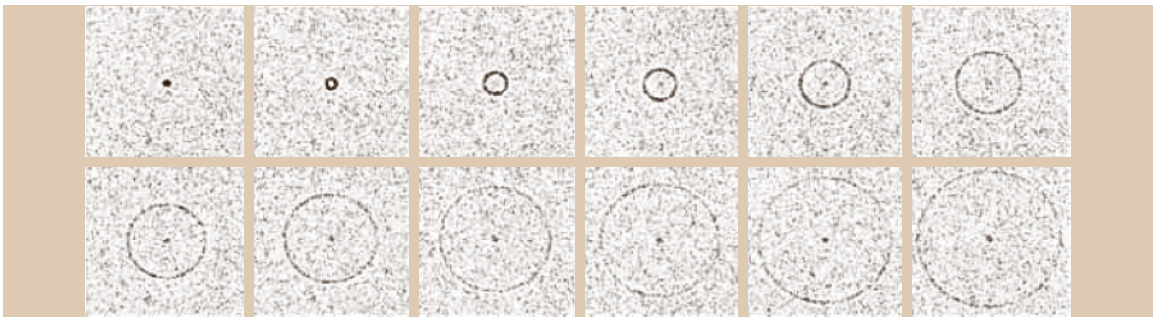


Fig. 8.18 Shock wave from a trapped, sonoluminescing bubble driven at 21.4 kHz and a pressure amplitude of 132 kPa. The interframe time is 30 ns. Picture size is 160 × 160 µm. (Courtesy of R. Geisler)

Amazon EX1014 Page 303

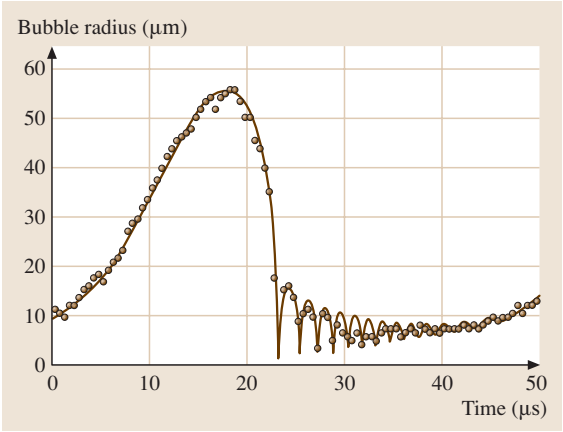


Fig. 8.19 Radius-time curve of a trapped bubble in a water-glycerine mixture derived from photographic observations. A numerically calculated curve (Gilmore model) is superimposed on the experimental data points (*open circles*). The calculation is based on the following parameters: driving frequency $f_0 = 21.4\,\text{kHz}$, ambient pressure $p_0 = 100\,\text{kPa}$, driving pressure $p_0 = 132\,\text{kPa}$, vapor pressure $p_v = 0$, equilibrium radius $R_n = 8.1\,\mu\text{m}$, density of the liquid $\rho = 1000\,\text{kg/m}^3$, viscosity $\mu = 0.0018\,\text{Ns/m}^2$ (measured) and surface tension $\sigma = 0.0725\,\text{N/m}$. The gas within the bubble is assumed to obey the adiabatic equation of state for an ideal gas with $\kappa = 1.2$. (Measured points courtesy of R. Geisler)

in a small bubble is relatively small, around 10^9 to 10^{10} molecules, this approach promises progress for the question of internal shock waves. At present molecular dynamics simulations are feasible with several million particles inside the bubble. Figure 8.20 gives a graphical view on the internal temperature distribution inside a collapsing sonoluminescence bubble with near-adiabatic conditions (reflection of the molecules at the inner boundary of the bubble) for six different times around maximum compression. The liquid

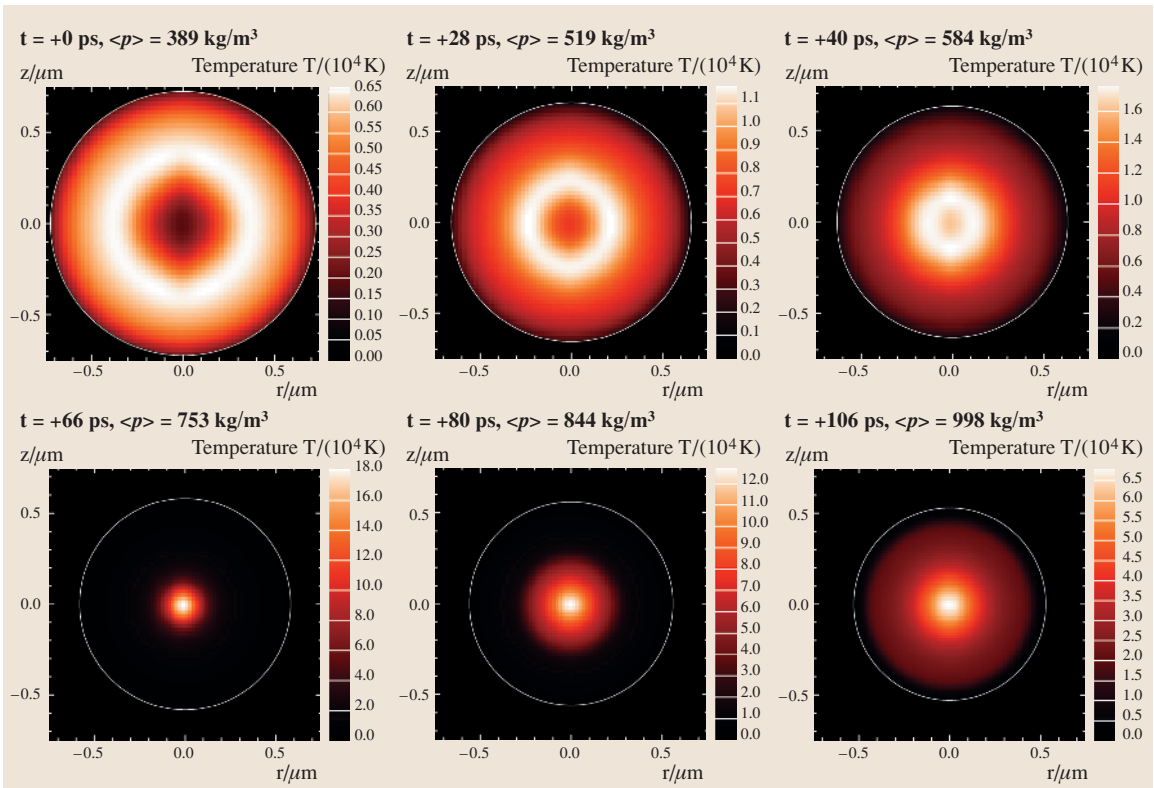


Fig. 8.20 Temperature distribution inside a collapsing sonoluminescence bubble filled with argon, one million particles, radius of the bubble at rest 4.5 μm, sound pressure amplitude 130 kPa, sound field frequency 26.5 kHz, total time covered 106 ps. (Courtesy of B. Metten [8.126])

Amazon EX1014 Page 304

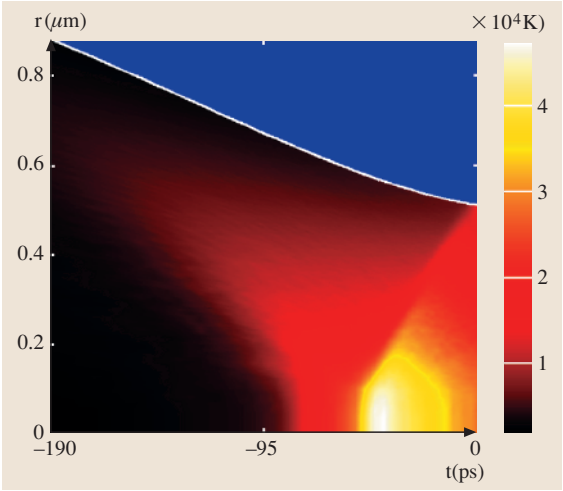


Fig. 8.21 Distribution of temperature inside an acoustically driven bubble in dependence on time. Molecular dynamics calculation with argon and water vapor including chemical reactions. (after [8.128])

motion and the inner molecular dynamics calculations are coupled via the gas pressure (at the interface), which is determined from the molecular dynamics of the 10^6 particles and inserted into the Rayleigh–Plesset equation giving the motion of the bubble wall. A strong focusing of energy inside the bubble is observed under these conditions. An inward-traveling compression wave focuses at the center and yields a core temperature of more than 10^5 K. This temperature is an overestimate as no energy-consuming effects are included.

The temperature variation inside the bubble with time can be visualized for spherical bubbles in a single picture in the following way (Fig. 8.21). The vertical axis represents the radial direction from the bubble center, the horizontal axis represents time and the color gives the temperature in Kelvin according to the color-coded scale (bar aside the figure). The upper solid line in the figure depicts the bubble wall; the blue part is water. In this case an argon bubble with water vapor is taken and chemical reactions are included. Then the temperature is lowered significantly by the endothermal processes of chemical species formation, for instance of OH^- radicals. The temperature drops to about 40 000 K in the center [8.128].

8.12 Acoustic Chaos

Nonlinearity in acoustics not only leads to shock waves and, as just demonstrated, to light emission, but to even deeper questions of nonlinear physics in general [8.4]. In his curiosity, man wants to know the future. Knowledge of the future may also be of help to master one's life by taking proper provisions. The question therefore arises: how far can we look into the future and where are the difficulties in doing so? This is the question of predictability or unpredictability. In the attempt to solve this question we find deterministic systems, stochastic systems and, nowadays, chaotic systems. The latter are special in that they combine deterministic laws that nevertheless give unpredictable output. They thus form a link between deterministic and stochastic systems.

8.12.1 Methods of Chaos Physics

A description of the methods developed to handle chaotic systems, in particular in acoustics, has been given in [8.32]. The main question in the context

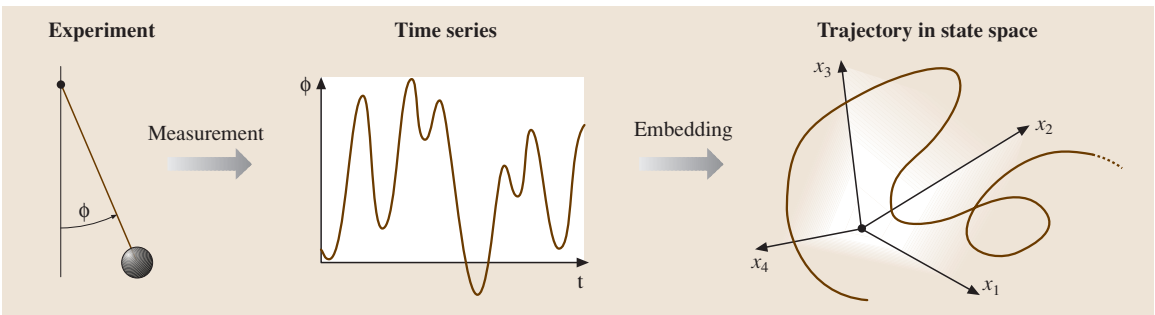


Fig. 8.22 Visualization of the embedding procedure

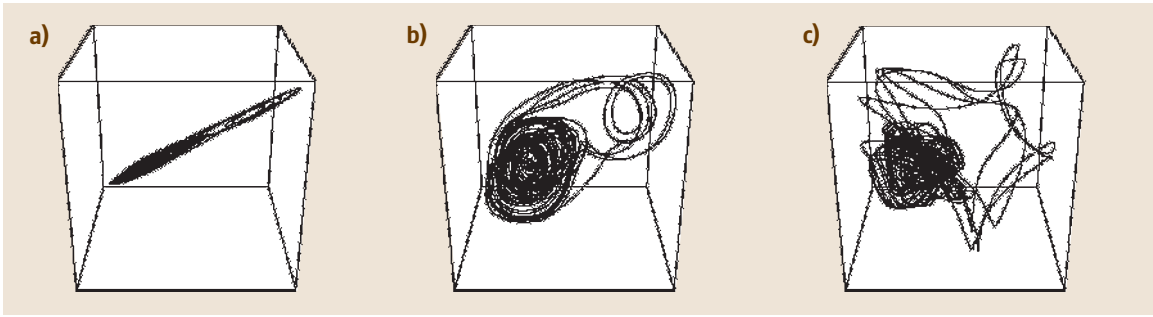


Fig. 8.23 Three embeddings with different delay times t_l for experimental data obtained from a chaotically oscillating, periodically driven pendulum. (Courtesy of U. Parlitz)

of experiments is how to connect the typically one-dimensional measurements (a time series) with the high-dimensional state space of theory (given by the number of variables describing the system). This is done by what is called *embedding* Fig. 8.22 [8.129, 130] and has led to a new field called *nonlinear time-series analysis* [8.131–133].

The embedding procedure runs as follows. Let $\{p(kt_s), k = 1, 2, \ldots, N\}$ be the measured time series, then an embedding into a n-dimensional state space is achieved by grouping n elements each of the time series to vectors $\boldsymbol{p}_k^{(n)} = [p(kt_s), p(kt_s + t_l), \ldots, p(kt_s + (n-1)t_l)]$, $k = 1, 2, \ldots, N-n$, in the n-dimensional state space. There t_s is the sampling interval at which samples of the variable p are taken. The delay time t_l is usually taken as a multiple of t_s, $t_l = lt_s$, $l = 1, 2, \ldots$, to avoid interpolation. The sampling time t_s, the delay time t_l and the embedding dimension n must all be chosen properly according to the problem under investigation. Figure 8.23 demonstrates the effect of different delay times with experimental data from a chaotically oscillating periodically driven pendulum. When the delay time t_l is too small then from one embedding vector to the next there is little difference and the embedded points all lie along the diagonal in the embedding space Fig. 8.23a. On the other hand, when the delay time t_l is too large, the reconstructed trajectory obtained by connecting successive vectors in the embedding space becomes folded and the reconstruction becomes fuzzy Fig. 8.23c. The embedding dimension n is usually not known beforehand when an unknown system is investigated for its properties. Therefore the embedding is done for increasing n until some criterion is fulfilled. The basic idea for a criterion is that a structure in the point set obtained by the embedding should appear. Then some law must be at work that generates it. The task after embedding is therefore to find and characterize the structure and to find the laws behind it. The characterization of the point set obtained can be done by determining its static structure (e.g., its dimension) and its dynamic structure (e.g., its expansion properties via the dynamics of nearby points).

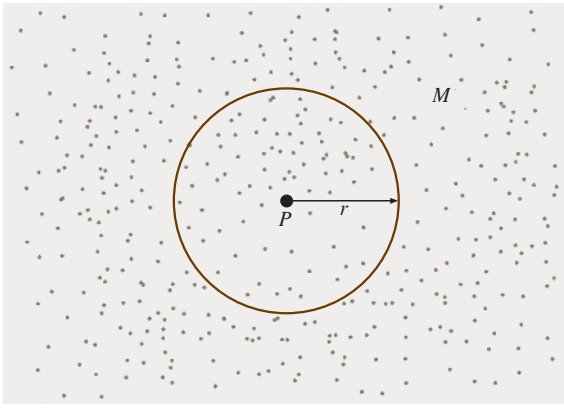


Fig. 8.24 Notations for the definition of the pointwise dimension

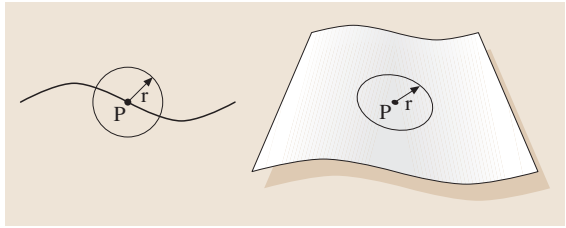


Fig. 8.25 Example of pointwise dimensions for one-dimensional (a curve) and two-dimensional (a surface) point sets

The dimension easiest to understand is the pointwise dimension. It is defined at a point P of a point set M by $N(r) \sim r^D$, $r \to 0$, where $N(r)$ is the number of



Fig. 8.26 The point set of a strange attractor of an experimental, driven pendulum. (Courtesy of M. Kaufmann)

points of the set M inside the ball of radius r around P (Fig. 8.24). Examples for a line and a surface give the right integer dimensions of one and two (Fig. 8.25), because for a line we have $N(r) \sim r^1$ for $r \to 0$ and thus $D = 1$, and for a surface we have $N(r) \sim r^2$ for $r \to 0$ and thus $D = 2$.

Chaotic or strange attractors usually give a fractal (noninteger) dimension. The pointwise dimension may not be the same in this case for each point P. The distribution of dimensions then gives a measure of the inhomogeneity of the point set. Figure 8.26 gives an example of a fractal point set. It has been obtained by sampling the angle and the angular velocity of a pendulum at a given phase of the periodic driving for parameters in a chaotic regime.

To investigate the dynamic properties of a (strange or chaotic) set the succession of the points must be retained. As chaotic dynamics is brought about by a stretching and folding mechanism and shows sensitive dependence on initial conditions (see, e.g., [8.32]) the behavior of two nearby points is of interest: do they separate or approach under the dynamics? This

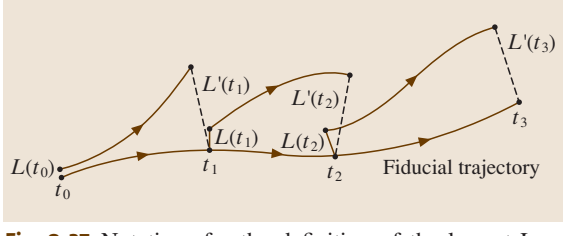


Fig. 8.27 Notations for the definition of the largest Lyapunov exponent

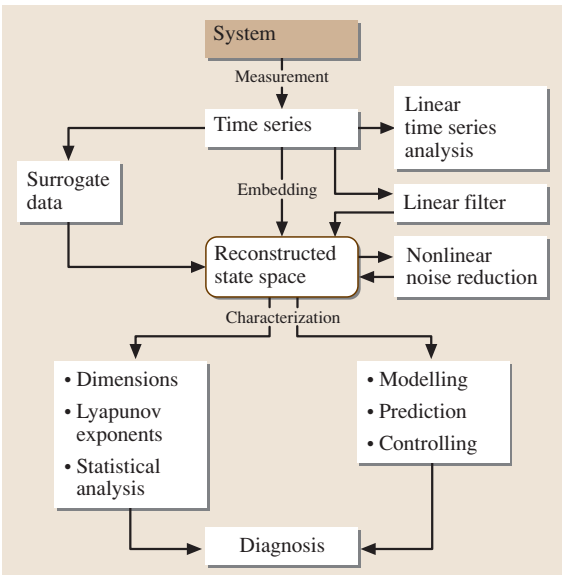


Fig. 8.28 Operations in nonlinear time-series analysis. (Courtesy of U. Parlitz)

behavior is quantified by the notion of the Lyapunov exponent. In Fig. 8.27 the calculation scheme [8.134] is depicted. It gives the maximum Lyapunov exponent via the formula

$$\lambda_{\max} = \frac{1}{t_m - t_0} \sum_{k=1}^{m} \log_2 \frac{L'(t_k)}{L(t_{k-1})} \quad \left[\frac{\text{bits}}{\text{s}}\right]. \qquad (8.224)$$

As space does not permit us to dig deeper into the operations possible with embedded data, only a graph depicting the possible operations in nonlinear time-series analysis is given in Fig. 8.28. Starting from the time series obtained by some measurement, the usual approach is to do a linear time-series analysis, for instance by calculating the Fourier transform or some correlation. The new way for chaotic systems proceeds via embedding to a reconstructed state space. There may be some filtering involved in between, but this is a dangerous operation as the results are often difficult to predict. For the surrogate data operation and nonlinear noise reduction the reader is referred to [8.131–133]. From the characterization operations we have mentioned here the dimension estimation and the largest Lyapunov exponent. Various statistical analyses can be done and the data can also be used for modeling, prediction and controlling the system. Overall, nonlinear time series analysis is a new diagnostic tool for describing (nonlinear) systems.

Amazon EX1014 Page 307

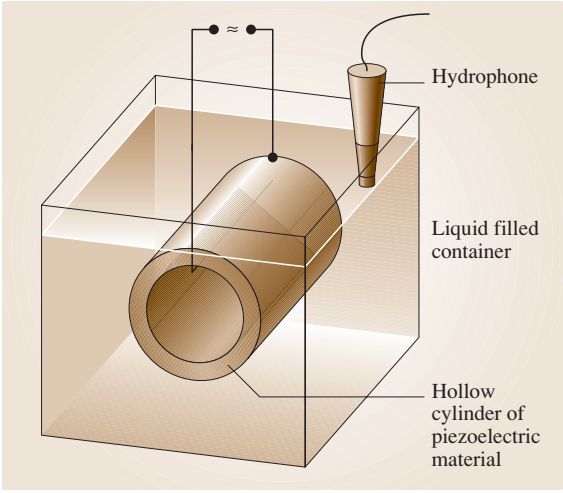


Fig. 8.29 Cylindrical transducer submerged in water for bubble oscillation, cavitation noise, and sonoluminescence studies

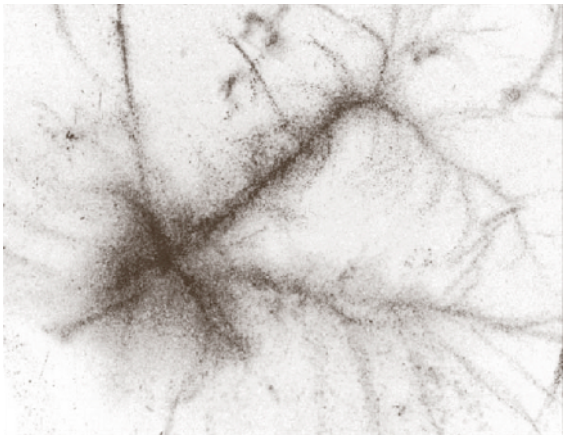
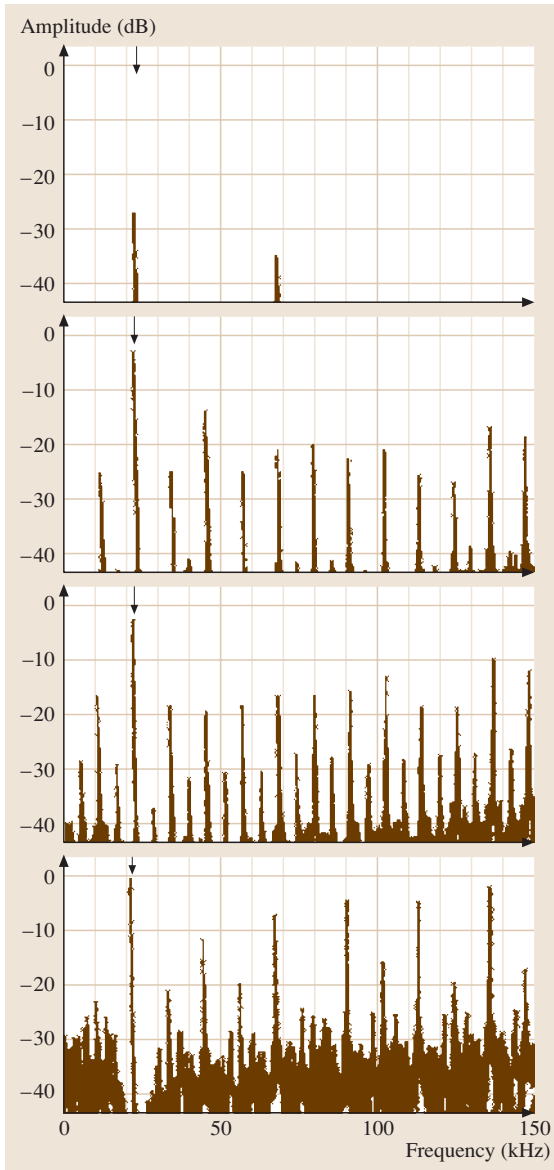


Fig. 8.30 Example of the filamentary bubble pattern (acoustic Lichtenberg figure) inside the cylinder in the projection along the axis of the transducer. (Courtesy of A. Billo)

Fig. 8.31 Power spectra of cavitation noise taken at a driving frequency of $f_0 = 23\,\text{kHz}$, marked in the figures by small triangle. The driving pressure was increased from one graph to the next, giving first a periodic, anharmonic signal (*top*), then subharmonic motion with lines at multiples of $f_0/2$ (period-2 motion, *top next*), then period-4 motion with lines at multiples of $f_0/4$ (*third plot*) and finally a broadband spectrum indicative of chaotic behavior (*bottom*)▶

8.12.2 Chaotic Sound Waves

One of the first examples of chaotic dynamics and where these methods were first applied was acoustic cavitation [8.4, 135, 136]. To investigate this phenomenon, the rupture of liquids by sound waves and the phenomena associated with it, an experimental arrangement as depicted in Fig. 8.29 was used. A cylindrical transducer of piezoelectric material is used to generate a strong sound field in water. A typical transducer was of 76 mm



Amazon EX1014 Page 308

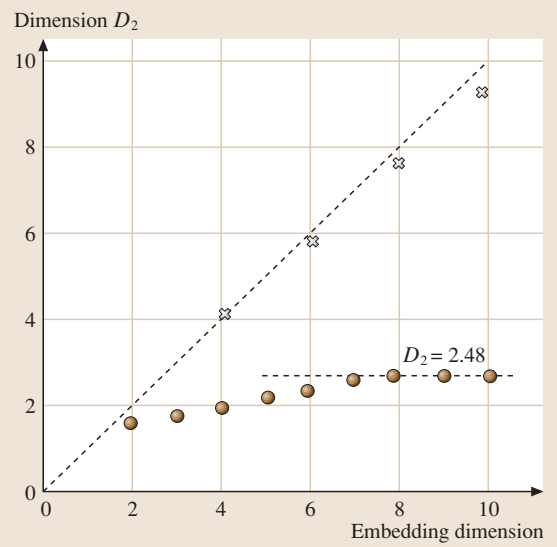


Fig. 8.32 Dimension estimation by embedding experimental acoustic cavitation sound pressure data into spaces of increasing dimension (*dots*). The *crosses* (×) are from an experiment, where noise from a noise generator has been embedded

length, 76 mm inner diameter, and 5 mm wall thickness submerged in a water-filled glass container. Cylinders of different size were available driven at different frequencies between 7 kHz and 23 kHz. A standing acoustic wave is set up inside the cylinder having its maximum pressure amplitude along the axis of the transducer. A hydrophone monitors the sound output of the liquid. Some electronics and a computer surround the experiment to drive the cylinder and to store the sampled pressure data from the hydrophone. When cavitation is fully developed, a dentritic filamentary bubble pattern is set up in the cylinder (Fig. 8.30).

When the sound field amplitude is raised the whole bubble field undergoes a period-doubling cascade to chaos [8.81, 137], a strong indication that chaotic dynamics really has been reached and that the broadband spectrum then encountered (Fig. 8.31) is not of stochastic origin. The cascade of period-doubled sound in the liquid has its origin in the cascade of period-doubled oscillations of the nonlinearly oscillating bubbles [8.81], and thus shows chaotic sound.

Embedding of the sound data into spaces of increasing dimension yields a structure with low fractal dimensions between two and three (Fig. 8.32). This is a quite surprising result in view of the thousands of oscillating bubbles of different sizes. Indeed, new results on the superposition of the sound output of a population of driven bubbles have indicated that the low dimensionality may be a result of the (weak) synchronization of their motion from the periodic driving and the scale of observation [8.138].

Chaotic sound waves also appear in other contexts [8.139]. In musical acoustics several instruments are susceptible to chaotic sound emission. Examples are the bowed string [8.140, 141], the clarinet [8.142] and gongs and cymbals [8.143].

The human speech production process is intrinsically nonlinear. To produce voice sounds the vocal cords are set into vibration through the air flow between them. The cords are a coupled system of nonlinear oscillators giving rise to bifurcations and chaos [8.144].

Sound can be generated by heat. When a gas-filled tube, closed at one end and open at the other is heated at the closed end and/or cooled at the open end, sound waves are emitted at a sufficiently high temperature difference (Sondhauss and Taconis oscillation). This phenomenon has been used to study its route to chaos [8.145]. It has been used to build a new type of refrigerator and heat pump [8.146].

References

8.1 M.F. Hamilton, D.T. Blackstock (Eds.): *Nonlinear Acoustics* (Academic, San Diego 1998)

8.2 K. Naugolnykh, L. Ostrovsky: *Nonlinear Wave Processes in Acoustics* (Cambridge Univ. Press, Cambridge 1998)

8.3 R.T. Beyer: *Nonlinear Acoustics* (Acoust. Soc. Am., Woodbury 1997)

8.4 W. Lauterborn, T. Kurz, U. Parlitz: Experimental nonlinear physics, Int. J. Bifurcation Chaos **7**, 2003–2033 (1997)

8.5 A.L. Thuras, R.T. Jenkins, H.T. O'Neil: Extraneous frequencies generated in air carrying intense sound waves, J. Acoust. Soc. Am. **6**, 173–180 (1935)

8.6 D.T. Blackstock, M.J. Crocker, D.G. Crighton, E.C. Everbach, M.A. Breazeale, A.B. Coppens, A.A. Atchley, M.F. Hamilton, W.E. Zorumski, W. Lauterborn, K.S. Suslick, L.A. Crum: Part II, Nonlinear Acoustics and Cavitation. In: *Encyclopedia of Acoustics*, Vol. 281, ed. by M.J. Crocker (Wiley, New York 1997) pp. 191–281

Amazon EX1014 Page 309

8.7 N.S. Bakhvalov, Y.M. Zhileikin, E.A. Zabolotskaya: *Nonlinear Theory of Sound Beams* (American Institute of Physics, Melville 1987)
8.8 B.K. Novikov, O.V. Rudenko, V.I. Timoshenko: *Nonlinear Underwater Acoustics* (American Institute of Physics, New York 1987)
8.9 O.V. Rudenko, S.I. Soluyan: *Theoretical Foundations of Nonlinear Acoustics* (Consultants Bureau, London 1977)
8.10 G.B. Whitham: *Linear and Nonlinear Waves* (Wiley, London 1974)
8.11 L.D. Rozenberg (Ed.): *High-Intensity Ultrasonic Fields* (Plenum, New York 1971)
8.12 M.F. Hamilton, D.T. Blackstock (Eds.): *Frontiers of Nonlinear Acoustics* (Elsevier, London 1990)
8.13 H. Hobaek (Ed.): *Advances in Nonlinear Acoustics* (World Scientific, Singapore 1993)
8.14 R.J. Wei (Ed.): *Nonlinear Acoustics in Perspective* (Nanjing Univ. Press, Nanjing 1996)
8.15 W. Lauterborn, T. Kurz (Eds.): *Nonlinear Acoustics at the Turn of the Millennium* (American Institute of Physics, Melville 2000)
8.16 O.V. Rudenko, O.A. Sapozhnikov (Eds.): *Nonlinear Acoustics at the Beginning of the 21st Century* (Faculty of Physics, Moscow State University, Moscow 2002)
8.17 S. Makarov, M. Ochmann: Nonlinear and thermoviscous phenomena in acoustics, Part I, Acustica **82**, 579–606 (1996)
8.18 S. Makarov, M. Ochmann: Nonlinear and thermoviscous phenomena in acoustics, Part II, Acustica **83**, 197–222 (1997)
8.19 S. Makarov, M. Ochmann: Nonlinear and thermoviscous phenomena in acoustics, Part III, Acustica **83**, 827–864 (1997)
8.20 C.E. Brennen: *Cavitation and Bubble Dynamics* (Oxford Univ. Press, Oxford 1995)
8.21 T.G. Leighton: *The Acoustic Bubble* (Academic, London 1994)
8.22 J.R. Blake, J.M. Boulton-Stone, N.H. Thomas (Eds.): *Bubble Dynamics and Interface Phenomena* (Kluwer, Dordrecht 1994)
8.23 F.R. Young: *Cavitation* (McGraw-Hill, London 1989)
8.24 L. van Wijngaarden (Ed.): *Mechanics and Physics of Bubbles in Liquids* (Martinus Nijhoff, The Hague 1982)
8.25 W. Lauterborn (Ed.): *Cavitation and Inhomogeneities in Underwater Acoustics* (Springer, Berlin, Heidelberg 1980)
8.26 E.A. Neppiras: Acoustic cavitation, Phys. Rep. **61**, 159–251 (1980)
8.27 H.G. Flynn: Physics of acoustic cavitation. In: *Physical Acoustics*, ed. by W.P. Mason (Academic, New York 1964) pp. 57–172
8.28 F.R. Young: *Sonoluminescence* (CRC, Boca Raton 2005)
8.29 T.J. Mason, J.P. Lorimer: *Applied Sonochemistry* (Wiley-VCH, Weinheim 2002)
8.30 K.S. Suslick (Ed.): *Ultrasound: Its Chemical, Physical and Biological Effects* (VCH, New York 1988)
8.31 W. Lauterborn: Acoustic chaos. In: *Encyclopedia of Physical Science and Technology*, 3rd edn., ed. by R.A. Meyers (Academic, San Diego 2002) pp. 117–127
8.32 W. Lauterborn, U. Parlitz: Methods of chaos physics and their application to acoustics, J. Acoust. Soc. Am. **84**, 1975–1993 (1988)
8.33 F.E. Fox, W.A. Wallace: Absorption of finite amplitude sound waves, J. Acoust. Soc. Am. **26**, 994–1006 (1954)
8.34 R.T. Beyer: Parameter of nonlinearity in fluids, J. Acoust. Soc. Am. **32**, 719–721 (1960)
8.35 A.B. Coppens, R.T. Beyer, M.B. Seiden, J. Donohue, F. Guepin, R.H. Hodson, C. Townsend: Parameter of nonlinearity. II, J. Acoust. Soc. Am. **38**, 797–804 (1965)
8.36 R.T. Beyer: *Nonlinear Acoustics* (Naval Ship Systems Command, Washington 1974), Table 3-1
8.37 E.C. Everbach: *Tissue Composition Determination via Measurement of the Acoustic Nonlinearity Parameter B/A, Ph.D. Dissertation* (Yale Univ., New Haven 1989) p. 66
8.38 Z. Zhu, M.S. Roos, W.N. Cobb, K. Jensen: Determination of the acoustic nonlinearity parameter B/A from phase measurements, J. Acoust. Soc. Am. **74**, 1518–1521 (1983)
8.39 X. Gong, Z. Zhu, T. Shi, J. Huang: Determination of the acoustic nonlinearity parameter in biological media using FAIS and ITD methods, J. Acoust. Soc. Am. **86**, 1–5 (1989)
8.40 W.K. Law, L.A. Frizell, F. Dunn: Determination of the nonlinearity parameter B/A of biological media, Ultrasound Med. Biol. **11**, 307–318 (1985)
8.41 J. Zhang, F. Dunn: A small volume thermodynamic system for B/A measurement, J. Acoust. Soc. Am. **89**, 73–79 (1991)
8.42 F. Plantier, J.L. Daridon, B. Lagourette: Measurement of the B/A nonlinearity parameter under high pressure: Application to water, J. Acoust. Soc. Am. **111**, 707–715 (2002)
8.43 E.C. Everbach: Parameters of nonlinearity of acoustic media. In: *Encyclopedia of Acoustics*, ed. by M.J. Crocker (Wiley, New York 1997) pp. 219–226
8.44 I. Rudnick: On the attenuation of finite amplitude waves in a liquid, J. Acoust. Soc. Am. **30**, 564–567 (1958)
8.45 J. Banchet, J.D.N. Cheeke: Measurements of the acoustic nonlinearity parameter B/A in solvents: Dependence on chain length and sound velocity, J. Acoust. Soc. Am. **108**, 2754–2758 (2000)
8.46 A.B. Coppens, R.T. Beyer, M.B. Seiden, J. Donohue, F. Guepin, R.H. Hodson, C. Townsend: Parameter of nonlinearity in fluids II, J. Acoust. Soc. Am. **38**, 797–804 (1965)

Amazon EX1014 Page 310

8.47 O. Nomoto: Nonlinear parameter of the 'Rao Liquid', J. Phys. Soc. Jpn. **21**, 569–571 (1966)
8.48 R.T. Beyer: Parameter of nonlinearity in fluids, J. Acoust. Soc. Am. **32**, 719–721 (1960)
8.49 K.L. Narayana, K.M. Swamy: Acoustic nonlinear parameter (B/A) in n-pentane, Acustica **49**, 336–339 (1981)
8.50 S.K. Kor, U.S. Tandon: Scattering of sound by sound from Beyers (B/A) parameters, Acustica **28**, 129–130 (1973)
8.51 K.M. Swamy, K.L. Narayana, P.S. Swamy: A study of (B/A) in liquified gases as a function of temperature and pressure from ultrasonic velocity measurements, Acustica **32**, 339–341 (1975)
8.52 H.A. Kashkooli, P.J. Dolan Jr., C.W. Smith: Measurement of the acoustic nonlinearity parameter in water, methanol, liquid nitrogen, and liquid helium-II by two different methods: A comparison, J. Acoust. Soc. Am. **82**, 2086–2089 (1987)
8.53 R.T. Beyer: The parameter B/A. In: *Nonlinear Acoustics*, ed. by M.F. Hamilton, D.T. Blackstock (Academic, San Diego 1998) pp. 25–39
8.54 L. Bjørnø: Acoustic nonlinearity of bubbly liquids, Appl. Sci. Res. **38**, 291–296 (1982)
8.55 J. Wu, Z. Zhu: Measurement of the effective nonlinearity parameter B/A of water containing trapped cylindrical bubbles, J. Acoust. Soc. Am. **89**, 2634–2639 (1991)
8.56 M.P. Hagelberg, G. Holton, S. Kao: Calculation of B/A for water from measurements of ultrasonic velocity versus temperature and pressure to 10 000 kg/cm^2, J. Acoust. Soc. Am. **41**, 564–567 (1967)
8.57 B. Riemann: Ueber die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite (On the propagation of plane waves of finite amplitude), Abhandl. Ges. Wiss. Göttingen **8**, 43–65 (1860), in German
8.58 S. Earnshaw: On the mathematical theory of sound, Philos. Trans. R. Soc. London **150**, 133–148 (1860)
8.59 D.T. Blackstock: Propagation of plane sound waves of finite amplitude in nondissipative fluids, J. Acoust. Soc. Am. **34**, 9–30 (1962)
8.60 M.F. Hamilton, D.T. Blackstock: On the coefficient of nonlinearity β in nonlinear acoustics, J. Acoust. Soc. Am. **83**, 74–77 (1988), Erratum, ibid, p. 1976
8.61 D.G. Crighton: Propagation of finite-amplitude waves in fluids. In: *Encyclopedia of Acoustics*, ed. by M.J. Crocker (Wiley, New York 1997) pp. 203–218
8.62 B.D. Cook: New procedure for computing finite-amplitude distortion, J. Acoust. Soc. Am. **34**, 941–946 (1962), see also the footnote on page 312 of 8.1
8.63 W. Keck, R.T. Beyer: Frequency spectrum of finite amplitude ultrasonic waves in liquids, Phys. Fluids **3**, 346–352 (1960)
8.64 L.E. Hargrove: Fourier series for the finite amplitude sound waveform in a dissipationless medium, J. Acoust. Soc. Am. **32**, 511–512 (1960)
8.65 E. Fubini Ghiron: Anomalie nella propagazione di onde acustiche di grande ampiezza (Anomalies in the propagation of acoustic waves of large amplitude), Alta Frequenza **4**, 530–581 (1935), in Italian
8.66 D.T. Blackstock: Nonlinear Acoustics (Theoretical). In: *American Institute of Physics Handbook*, 3rd edn., ed. by D.E. Gray (McGraw Hill, New York 1972) pp. 3/183–3/205
8.67 A. Hirschberg, J. Gilbert, R. Msallam, A.P.J. Wijnands: Shock waves in trombones, J. Acoust. Soc. Am. **99**, 1754–1758 (1996)
8.68 M.J. Lighthill: Viscosity effects in sound waves of finite amplitude. In: *Surveys in Mechanics*, ed. by G.K. Batchelor, R.M. Davies (Cambridge Univ. Press, Cambridge 1956) pp. 249–350
8.69 J. Lighthill: *Waves in Fluids* (Cambridge Univ. Press, Cambridge 1980)
8.70 J.S. Mendousse: Nonlinear dissipative distortion of progressive sound waves at moderate amplitudes, J. Acoust. Soc. Am. **25**, 51–54 (1953)
8.71 Z.A. Gol'dberg: Second approximation acoustic equations and the propagation of plane waves of finite amplitude, Sov. Phys.-Acoust. **2**, 346–350 (1956)
8.72 D.T. Blackstock: Thermoviscous attenuation of plane, periodic, finite amplitude sound waves, J. Acoust. Soc. Am. **36**, 534–542 (1964)
8.73 E. Hopf: The partial differential equation $u_t + uu_x = \mu u_{xx}$, Comm. Pure Appl. Math. **3**, 201–230 (1950)
8.74 J.D. Cole: On a quasi-linear parabolic equation occurring in aerodynamics, Quart. Appl. Math. **9**, 225–236 (1951)
8.75 M. Abramowitz, I.A. Stegun (Eds.): *Handbook of Mathematical Functions* (Dover, New York 1970)
8.76 L.S. Gradsteyn, I.M. Ryshik: *Table of Integrals, Series and Products* (Academic, Orlando 1980)
8.77 D.T. Blackstock: Connection between the Fay and Fubini solutions for plane sound waves of finite amplitude, J. Acoust. Soc. Am. **39**, 1019–1026 (1966)
8.78 R.D. Fay: Plane sound waves of finite amplitude, J. Acoust. Soc. Am. **3**, 222–241 (1931)
8.79 R.V. Khokhlov, S.I. Soluyan: Propagation of acoustic waves of moderate amplitude through dissipative and relaxing media, Acustica **14**, 241–246 (1964)
8.80 I.S. Akhatov, R.I. Nigmatulin, R.T. Lahey Jr.: The analysis of linear and nonlinear bubble cluster dynamics, Multiphase Sci. Technol. **17**, 225–256 (2005)
8.81 W. Lauterborn, T. Kurz, R. Mettin, C.D. Ohl: Experimental and theoretical bubble dynamics, Adv. Chem. Phys. **110**, 295–380 (1999)
8.82 C. Chaussy (Ed.): *Extracorporeal Shock Wave Lithotripsy* (Karger, New York 1986)
8.83 G.A. Askar'yan: Self-focusing of powerful sound during the production of bubbles, J. Exp. Theor. Phys. Lett. **13**, 283–284 (1971)
8.84 P. Ciuti, G. Jernetti, M.S. Sagoo: Optical visualization of non-linear acoustic propagation in cavitating liquids, Ultrasonics **18**, 111–114 (1980)

Amazon EX1014 Page 311

8.85 Yu.A. Kobelev, L.A. Ostrovsky: Nonlinear acoustic phenomena due to bubble drift in a gas-liquid mixture, J. Acoust. Soc. Am. **85**, 621–627 (1989)

8.86 S.L. Lopatnikov: Acoustic phase echo in liquid with gas bubbles, Sov. Tech. Phys. Lett. **6**, 270–271 (1980)

8.87 I.Sh. Akhatov, V.A. Baikov: Propagation of sound perturbations in heterogeneous gas-liquid systems, J. Eng. Phys. **9**, 276–280 (1986)

8.88 I.Sh. Akhatov, V.A. Baikov, R.A. Baikov: Propagation of nonlinear waves in gas-liquid media with a gas content variable in space, Fluid Dyn. **7**, 161–164 (1986)

8.89 E. Zabolotskaya: Nonlinear waves in liquid with gas bubbles, Trudi IOFAN (Institute of General Physics of the Academy of Sciences, Moscow) **18**, 121–155 (1989), in Russian

8.90 I.Akhatov, U. Parlitz, W. Lauterborn: Pattern formation in acoustic cavitation, J. Acoust. Soc. Am. **96**, 3627–3635 (1994)

8.91 I.Akhatov, U. Parlitz, W. Lauterborn: Towards a theory of self-organization phenomena in bubble-liquid mixtures, Phys. Rev. E **54**, 4990–5003 (1996)

8.92 O.A. Druzhinin, L.A. Ostrovsky, A. Prosperetti: Low-frequency acoustic wave generation in a resonant bubble-layer, J. Acoust. Soc. Am. **100**, 3570–3580 (1996)

8.93 L.A. Ostrovsky, A.M. Sutin, I.A. Soustova, A.I. Matveyev, A.I. Potapov: Nonlinear, low-frequency sound generation in a bubble layer: Theory and laboratory experiment, J. Acoust. Soc. Am. **104**, 722–726 (1998)

8.94 R.I. Nigmatulin: *Dynamics of Multiphase Systems* (Hemisphere, New York 1991)

8.95 V.E. Nakoryakov, V.G. Pokusaev, I.R. Shreiber: *Propagation of Waves in Gas- and Vapour-liquid Media* (Institute of Thermophysics, Novosibirsk 1983)

8.96 L. van Wijngaarden: One-dimensional flow of liquids containing small gas bubbles, Ann. Rev. Fluid Mech. **4**, 369–396 (1972)

8.97 U. Parlitz, R. Mettin, S. Luther, I. Akhatov, M. Voss, W. Lauterborn: Spatiotemporal dynamics of acoustic cavitation bubble clouds, Philos. Trans. R. Soc. London Ser. A **357**, 313–334 (1999)

8.98 J.G. McDaniel, I. Akhatov, R.G. Holt: Inviscid dynamics of a wet foam drop with monodisperse bubble size distribution, Phys. Fluids **14**, 1886–1984 (2002)

8.99 A. Prosperetti, A. Lezzi: Bubble dynamics in a compressible liquid. Part 1. First order theory, J. Fluid Mech. **168**, 457–478 (1986)

8.100 R.I. Nigmatulin, I.S. Akhatov, N.K. Vakhitova, R.T. Lahey Jr.: On the forced oscillations of a small gas bubble in a spherical liquid-filled flask, J. Fluid Mech. **414**, 47–73 (2000)

8.101 A. Jeffrey, T. Kawahara: *Asymptotic Methods in Nonlinear Wave Theory* (Pitman, London 1982)

8.102 L. van Wijngaarden: On the equations of motion for mixtures of fluid and gas bubbles, J. Fluid Mech. **33**, 465–474 (1968)

8.103 N.A. Gumerov: Propagation of long waves of finite amplitude in a liquid with polydispersed gas bubbles, J. Appl. Mech. Tech. Phys. **1**, 79–85 (1992)

8.104 A. Hasegawa: *Plasma Instabilities and Nonlinear Effects* (Springer, Berlin, Heidelberg 1975)

8.105 N.A. Gumerov: The weak non-linear fluctuations in the radius of a condensed drop in an acoustic field, J. Appl. Math. Mech. (PMM U.S.S.R.) **53**, 203–211 (1989)

8.106 N.A. Gumerov: Weakly non-linear oscillations of the radius of a vapour bubble in an acoustic field, J. Appl. Math. Mech. **55**, 205–211 (1991)

8.107 N.A. Gumerov: On quasi-monochromatic weakly nonlinear waves in a low-dissipative bubbly liquid, J. Appl. Math. Mech. **56**, 50–59 (1992)

8.108 S. Gavrilyuk: Large amplitude oscillations and their "thermodynamics" for continua with "memory", Euro. J. Mech. B/Fluids **13**, 753–764 (1994)

8.109 S.L. Gavrilyuk, V.M. Teshukov: Generalized vorticity for bubbly liquid and dispersive shallow water equations, Continuum Mech. Thermodyn. **13**, 365–382 (2001)

8.110 L.F. McGoldrick: Resonant interactions among capillary-gravity waves, J. Fluid Mech. **21**, 305–331 (1965)

8.111 D.J. Benney: A general theory for interactions between long and short waves, Studies Appl. Math. **56**, 81–94 (1977)

8.112 I.Sh. Akhatov, D.B. Khismatulin: Effect of dissipation on the interaction between long and short waves in bubbly liquids, Fluid Dyn. **35**, 573–583 (2000)

8.113 I.Sh. Akhatov, D.B. Khismatulin: Two-dimensional mechanisms of interaction between ultrasound and sound in bubbly liquids: Interaction equations, Acoust. Phys. **47**, 10–15 (2001)

8.114 I.Sh. Akhatov, D.B. Khismatulin: Mechanisms of interaction between ultrasound and sound in liquids with bubbles: Singular focusing, Acoust. Phys. **47**, 140–144 (2001)

8.115 D.B. Khismatulin, I.Sh. Akhatov: Sound – ultrasound interaction in bubbly fluids: Theory and possible applications, Phys. Fluids **13**, 3582–3598 (2001)

8.116 V.E. Zakharov: Collapse of Langmuir waves, Sov. Phys. J. Exp. Theor. Phys. **72**, 908–914 (1972)

8.117 D.F. Gaitan, L.A. Crum, C.C. Church, R.A. Roy: An experimental investigation of acoustic cavitation and sonoluminescence from a single bubble, J. Acoust. Soc. Am. **91**, 3166–3183 (1992)

8.118 Y. Tian, J.A. Ketterling, R.E. Apfel: Direct observations of microbubble oscillations, J. Acoust. Soc. Am. **100**, 3976–3978 (1996)

8.119 J. Holzfuss, M. Rüggeberg, A. Billo: Shock wave emission of a sonoluminescing bubble, Phys. Rev. Lett. **81**, 5434–5437 (1998)

8.120 R. Pecha, B. Gompf: Microimplosions: Cavitation collapse and shock wave emission on a nanosecond time scale, Phys. Rev. Lett. **84**, 1328–1330 (2000)

Amazon EX1014 Page 312

8.121 E. Heim: Über das Zustandekommen der Sonoluminеszenz (On the origin of sonoluminescence). In: *Proceedings of the Third International Congress on Acoustics, Stuttgart, 1959*, ed. by L. Cremer (Elsevier, Amsterdam 1961) pp. 343–346, in German

8.122 C.C. Wu, P.H. Roberts: A model of sonoluminescence, Proc. R. Soc. Lond. A **445**, 323–349 (1994)

8.123 W.C. Moss, D.B. Clarke, D.A. Young: Calculated pulse widths and spectra of a single sonoluminescing bubble, Science **276**, 1398–1401 (1997)

8.124 V.Q. Vuong, A.J. Szeri, D.A. Young: Shock formation within sonoluminescence bubbles, Phys. Fluids **11**, 10–17 (1999)

8.125 B. Metten, W. Lauterborn: Molecular dynamics approach to single-bubble sonoluminescence. In: *Nonlinear Acoustics at the Turn of the Millennium*, ed. by W. Lauterborn, T. Kurz (Am. Inst. Physics, Melville 2000) pp. 429–432

8.126 B. Metten: *Molekulardynamik-Simulationen zur Sonolumineszenz*, *Dissertation* (Georg-August Universität, Göttingen 2000)

8.127 S.J. Ruuth, S. Putterman, B. Merriman: Molecular dynamics simulation of the response of a gas to a spherical piston: Implication for sonoluminescence, Phys. Rev. E **66**, 1–14 (2002), 036310

8.128 W. Lauterborn, T. Kurz, B. Metten, R. Geisler, D. Schanz: Molecular dynamics approach to sonoluminescent bubbles. In: *Theoretical and Computational Acoustics 2003*, ed. by A. Tolstoy, Yu-Chiung Teng, E.C. Shang (World Scientific, New Jersey 2004) pp. 233–243

8.129 N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw: Geometry from a time series, Phys. Rev. Lett. **45**, 712–716 (1980)

8.130 F. Takens: Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence*, ed. by D.A. Rand, L.S. Young (Springer, Berlin, Heidelberg 1981) pp. 366–381

8.131 H.D.I. Abarbanel: *Analysis of Observed Chaotic Data* (Springer, Berlin, New York 1996)

8.132 H. Kantz, T. Schreiber: *Nonlinear Time Series Analysis* (Cambridge Univ. Press, Cambridge 1997)

8.133 U. Parlitz: Nonlinear time serie analysis. In: *Nonlinear Modeling – Advanced Black Box Techniques*, ed. by J.A.K. Suykens, J. Vandewalle (Kluwer Academic, Dordrecht 1998) pp. 209–239

8.134 A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano: Determining Lyapunov exponents from time series, Physica D **16**, 285–317 (1985)

8.135 W. Lauterborn, E. Cramer: Subharmonic route to chaos observed in acoustics, Phys. Rev. Lett. **47**, 1445–1448 (1981)

8.136 W. Lauterborn, J. Holzfuss: Acoustic chaos, Int. J. Bifurcation Chaos **1**, 13–26 (1991)

8.137 W. Lauterborn, A. Koch: Holographic observation of period-doubled and chaotic bubble oscillations in acoustic cavitation, Phys. Rev. A **35**, 1774–1976 (1987)

8.138 S. Luther, M. Sushchik, U. Parlitz, I. Akhatov, W. Lauterborn: Is cavitation noise governed by a low-dimensional chaotic attractor?. In: *Nonlinear Acoustics at the Turn of the Millennium*, ed. by W. Lauterborn, T. Kurz (Am. Inst. Physics, Melville 2000) pp. 355–358

8.139 W. Lauterborn: Nonlinear dynamics in acoustics, Acustica acta acustica Suppl. 1 **82**, S46–S55 (1996)

8.140 M.E. McIntyre, R.T. Schumacher, J. Woodhouse: On the oscillation of musical instruments, J. Acoust. Soc. Am. **74**, 1325–1345 (1983)

8.141 N.B. Tuffilaro: Nonlinear and chaotic string vibrations, Am. J. Phys. **57**, 408–414 (1989)

8.142 C. Maganza, R. Caussé, Laloë: Bifurcations, period doublings and chaos in clarinet-like systems, Europhys. Lett. **1**, 295–302 (1986)

8.143 K.A. Legge, N.H. Fletcher: Nonlinearity, chaos, and the sound of shallow gongs, J. Acoust. Soc. Am. **86**, 2439–2443 (1989)

8.144 H. Herzel: Bifurcation and chaos in voice signals, Appl. Mech. Rev. **46**, 399–413 (1993)

8.145 T. Yazaki: Experimental observation of thermoacoustic turbulence and universal properties at the quasiperiodic transition to chaos, Phys. Rev. **E48**, 1806–1818 (1993)

8.146 G.W. Swift: Thermoacoustic engines, J. Acoust. Soc. Am. **84**, 1145–1180 (1988)

Amazon EX1014 Page 313

Part C Architect

Part C Architectural Acoustics

9 Acoustics in Halls for Speech and Music
Anders Christian Gade, Lyngby, Denmark

10 Concert Hall Acoustics Based on Subjective Preference Theory
Yoichi Ando, Makizono, Kirishima, Japan

11 Building Acoustics
James Cowan, White River Junction, USA

Amazon EX1014 Page 314

9. Acoustics in Halls for Speech and Music

This chapter deals specifically with concepts, tools, and architectural variables of importance when designing auditoria for speech and music. The focus will be on cultivating the useful components of the sound in the room rather than on avoiding noise from outside or from installations, which is dealt with in Chap. 11. The chapter starts by presenting the subjective aspects of the room acoustic experience according to consensus at the time of writing. Then follows a description of their objective counterparts, the objective room acoustic parameters, among which the classical *reverberation time* measure is only one of many, but still of fundamental value. After explanations on how these parameters can be measured and predicted during the design phase, the remainder of the chapter deals with how the acoustic properties can be controlled by the architectural design of auditoria. This is done by presenting the influence of individual design elements as well as brief descriptions of halls designed for specific purposes, such as drama, opera, and symphonic concerts. Finally, some important aspects of loudspeaker installations in auditoria are briefly touched upon.

9.1 **Room Acoustic Concepts** 302

9.2 **Subjective Room Acoustics** 303
- 9.2.1 The Impulse Response 303
- 9.2.2 Subjective Room Acoustic Experiment Techniques 303
- 9.2.3 Subjective Effects of Audible Reflections 305

9.3 **Subjective and Objective Room Acoustic Parameters** 306
- 9.3.1 Reverberation Time 306
- 9.3.2 Clarity 308
- 9.3.3 Sound Strength 308
- 9.3.4 Measures of Spaciousness 309
- 9.3.5 Parameters Relating to Timbre or Tonal Color 310
- 9.3.6 Measures of Conditions for Performers 310
- 9.3.7 Speech Intelligibility 311
- 9.3.8 Isn't One Objective Parameter Enough? 312
- 9.3.9 Recommended Values of Objective Parameters 313

9.4 **Measurement of Objective Parameters** ... 314
- 9.4.1 The Schroeder Method for the Measurement of Decay Curves 314
- 9.4.2 Frequency Range of Measurements 314
- 9.4.3 Sound Sources 315
- 9.4.4 Microphones 315
- 9.4.5 Signal Storage and Processing 315

9.5 **Prediction of Room Acoustic Parameters** 316
- 9.5.1 Prediction of Reverberation Time by Means of Classical Reverberation Theory 316
- 9.5.2 Prediction of Reverberation in Coupled Rooms 318
- 9.5.3 Absorption Data for Seats and Audiences 319
- 9.5.4 Prediction by Computer Simulations 320
- 9.5.5 Scale Model Predictions 321
- 9.5.6 Prediction from Empirical Data ... 322

9.6 **Geometric Design Considerations** 323
- 9.6.1 General Room Shape and Seating Layout 323
- 9.6.2 Seating Arrangement in Section .. 326
- 9.6.3 Balcony Design 327
- 9.6.4 Volume and Ceiling Height 328
- 9.6.5 Main Dimensions and Risks of Echoes 329
- 9.6.6 Room Shape Details Causing Risks of Focusing and Flutter 329
- 9.6.7 Cultivating Early Reflections 330
- 9.6.8 Suspended Reflectors 331
- 9.6.9 Sound-Diffusing Surfaces 333

9.7 **Room Acoustic Design of Auditoria for Specific Purposes** 334
- 9.7.1 Speech Auditoria, Drama Theaters and Lecture Halls 334

Amazon EX1014 Page 315

9.7.2 Opera Halls 335
9.7.3 Concert Halls for Classical Music .. 338
9.7.4 Multipurpose Halls 342
9.7.5 Halls for Rhythmic Music 344
9.7.6 Worship Spaces/Churches 346
9.8 **Sound Systems for Auditoria** 346
9.8.1 PA Systems 346
9.8.2 Reverberation-Enhancement Systems 348
References 349

Current knowledge about room acoustic design is based on several centuries of practical (trial-and-error) experience plus a single century of scientific research. Over the last couple of decades scientific knowledge has matured to a level where it can explain to a usable degree:

1. the subjective aspects related to how we perceive the acoustics of rooms
2. how these subjective aspects are governed by objective properties in the sound fields
3. how these objective properties of the sound field are governed by the physical variables in the architectural design

This chapter is organized in the same manner. The various subjective acoustic aspects will form the basis for our discussion and be a guide through most aspects of relevance in room acoustic design. However, some trends in current design practice based on the experience and intuition of individual acoustical designers will also be commented on. It is hoped that this approach will also have the advantage of stimulating the reader's ability to judge a room from her/his own listening experience, an ability which is important not only for designers of concert halls, but also for those responsible for creating and enjoying the sound in these halls: musicians, sound-system designers, recording engineers and concert enthusiasts.

9.1 Room Acoustic Concepts

In order to help the reader maintain a clear perspective all the way through this chapter, Fig. 9.1 illustrates the universe of architectural acoustics.

In the upper half of the figure, we have the phenomena experienced in the real world. Going from left to right, we have the auditoria, in which we can experience objective sound fields causing subjective impressions of the acoustic conditions. In all of these three domains, we find a huge number of degrees of freedom: halls can differ in a myriad of ways (from overall dimensions to detailed shaping of small details like door handles), the sound fields which we describe by the impulse response concept – as explained in a moment – contain a wealth of individual sound components (reflections), each being a function of time, level, direction and frequency. Also, every individual may have his/her own way of expressing what the room does to the sound heard.

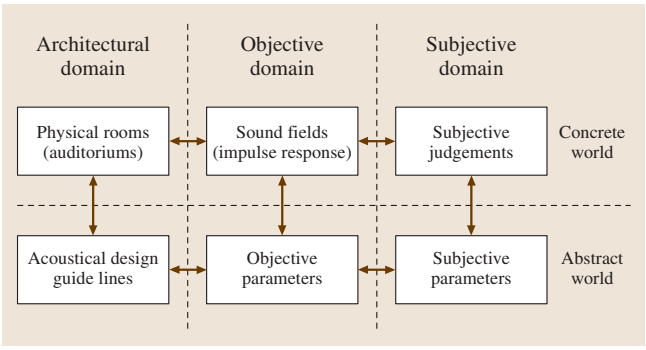


Fig. 9.1 Overview of concepts related to subjective room acoustics (after [9.1])

In other words, like in many other aspects of life, the real world is so complex that we need to simplify the problem – reduce the degrees of freedom – through definition of abstract, well-formulated and meaningful concepts and parameters in all three domains (the lower row of boxes). First, we must try to define a vocabulary to describe the subjective acoustic impression: we will isolate a set of subjective acoustic parameters that are valid to most people's listening experience (as indicated by the box in the lower-right corner), then we will try to deduce those properties from the sound fields that are responsible for our experience of each of the subjective aspects: we will define a set of objective, measurable parameters that correlate well with the subjective parameters, and finally – in order to be able to guide the architect – we must find out which aspects of the design govern the important objective and in turn the subjective

Amazon EX1014 Page 316

parameters, so we can assist in building halls meeting the specific needs of their users.

A rough historical overview: systematic search for objective acoustic parameters peaked in the 1960s and 1970s, while the relationships between the objective parameters and design variables were the topic of many research efforts in the 1980s and 1990s. However, we cannot yet claim to have the answers to all questions regarding the acoustic design of auditoria. Hopefully, there will always be room for intuition and individual judgment, not only in room acoustic design; but also as a source of inspiration for continued research efforts. Still it should be clear that room acoustic design today is based on solid research; it is not a snake-oil business.

9.2 Subjective Room Acoustics

9.2.1 The Impulse Response

The basic source of information regarding the audible properties of the sound field in a room is the impulse response signal. Actually, this signal – when recorded with a multichannel technique preserving the information about direction of incidence – can be shown to contain all information about the acoustics of a room between two specific source and receiver positions.

Consider a short sound impulse being emitted by the source on the stage, as shown in Fig. 9.2. A spherical wave propagates away from the source in all directions and the sound first heard in the listener position originates from that part of the wave that has propagated directly from the source to the receiver, called *the direct sound*. This is shown on the left in the lower part of Fig. 9.2, which shows the received signal versus time at a given position in the room.

This component is soon followed by other parts of the wave that have been reflected one or more times by the room boundaries or by objects in the room before reaching the receiver, called *the early reflections*. Besides arriving later than the direct sound, normally these reflections are also weaker because the intensity is reduced as the area of the spherical wavefront increases with time (spherical distance attenuation) and because a fraction of the energy is being absorbed each time the wave hits a more or less sound-absorbing room boundary or object in the room.

The sound wave will continue to be reflected and to pass the receiver position until all the energy has been absorbed by the boundaries/objects or by the air. The density of these later reflections increases with time (proportional to t^2), but the attenuation due to absorption at the room boundaries ensures that eventually all sound dies out. This decay is often heard as reverberation in the room, as Sabine did, when he carried out his famous experiment in the Fogg Art Museum more than 100 years ago [9.2]. This event marked the start of subjective room acoustics as a science.

9.2.2 Subjective Room Acoustic Experiment Techniques

With each of the reflections specified in time, level, spectrum and direction of incidence, the impulse response contains a huge amount of information. Now, the question is how we can reduce this to what is necessary for an explanation of why we perceive the acoustics in different rooms as being different.

Experiments trying to answer this question have mainly been carried out in the second half of the 20th century when electro-acoustic means for recording, measurement and simulation of sound fields had become available. Such experiments take place in the four rightmost boxes in Fig. 9.1: a number of impulse responses – or rather, music or speech convolved (filtered) through these responses – are presented to a number of subjects. The subjective evaluations are collected either as simple preferences between the sound fields presented in pairs or as scalings along specific subjective parameters suggested by the experimenter. The experimenter will also choose a number of different objective parameters, and calculate their values for each of the impulse responses presented. The level of correlation between these objective values and the objective scalings will then indicate which of the parameters and therefore which properties of the impulse responses are responsible for the subjective evaluations.

As can be imagined, the results of such experiments will be strongly dependent on the range of variation and degree of realism of the impulse responses presented – as well as by the sound quality of the equipment used for the presentation. Besides, the results will always be limited by the imagination of the experimenter regarding his/her:

1. suggestions of proper semantic descriptors for the subjects' evaluation (except in cases where the subjects are only asked to give preferences) and

2. suggestions of calculation of parameters from the impulse response.

In other words, before the experiments, the experimenter must have some good hypotheses regarding the contents of the lower-mid and lower-right boxes in Fig. 9.1.

Over the years, a variety of different techniques have been applied for presenting room sound fields to test subjects:

1. Collection of opinions about existing halls recalled from the memory of qualified listeners
2. Listening tests in different halls or in a hall with variable acoustics
3. Listening tests with presentation of recordings from existing halls or of *dry* speech/music convolved with impulse responses recorded in existing halls
4. Listening tests with sound fields synthesized in anechoic rooms (via multiple loudspeakers in an anechoic room) or impulse responses generated/modified by computer convolved with *dry* speech/music and presented over headphones.

It should be mentioned here that our acoustic memory is very short. Therefore, unless the subjects are highly experienced in evaluating concert hall acoustics, comparison of acoustic experiences separated by days,

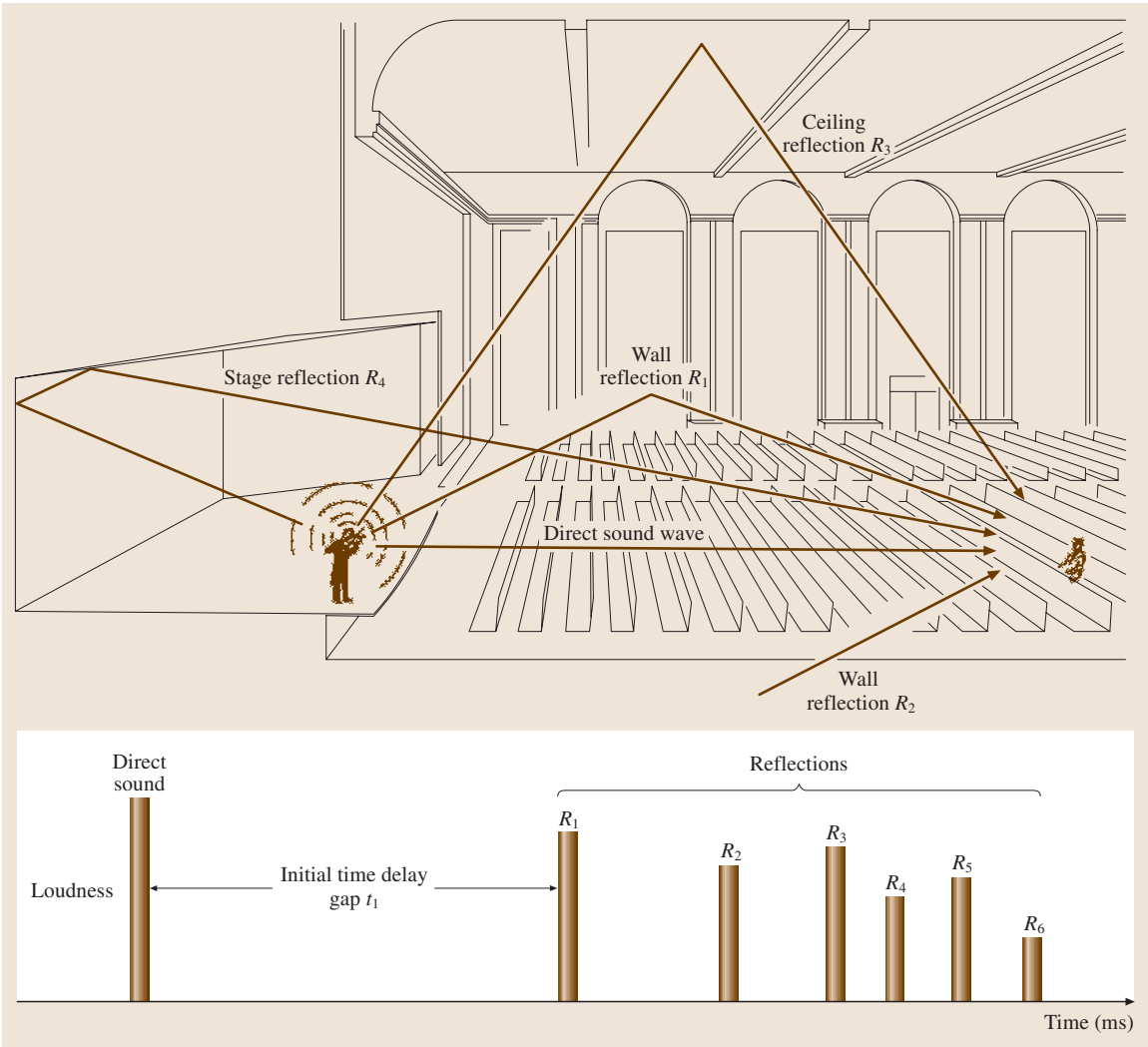


Fig. 9.2 Diagram illustrating the generation of an impulse response in a room (after [9.3])

weeks or even years is less reliable than comparison in the lab, where different recordings can be presented immediately one after the other. On the other hand, it is often difficult to get really qualified people to come to the lab to participate in scientific tests. Therefore, results from interview surveys can still be highly relevant.

These different techniques for presenting the sound fields all have different limitations. Experiments in the lab involving electro-acoustic reproduction or the generation of sound fields may lack fidelity, while listening in a real hall to a real orchestra will normally lack flexibility and control of the independent variables. Another important difference between these two methods is the absence of realistic visual cues in the lab. Enjoyment of a performance is a holistic experience, a fact which is now given increased scientific attention. In all cases elaborate statistical analysis of the results are needed in order to separate the significant facts from the error variance present in any experiment involving subjective judgements. Since the 1960s, multidimensional statistical tool such as factor analysis and multidimensional scaling have been very useful in distinguishing between the different subjective parameters which are present – consciously or not – in our evaluation. In any case, published results are more likely to be close to the truth, if they have been verified by several experiments – and experimenters – using different experimental approaches.

9.2.3 Subjective Effects of Audible Reflections

As the impulse response consists of the direct sound followed by a series of reflections, the simplest case possible would be to have only the direct sound plus a single reflection (and perhaps some late, artificial reverberation). Many experiments in simulated sound fields have been conducted using such a simple setup, and in spite of the obvious lack of realism, we can still learn a lot about the subjective effects of reflections from these experiments.

First of all, due to the psycho-acoustic phenomenon called forward masking, the reflection will be inaudible if it arrives very soon after the direct sound and/or its level is very low relative to the direct sound. Thus, there exists a level threshold of audibility depending on delay and direction of incidence relative to the direct sound. Only if the level of the reflection is above this threshold will the reflection have an audible effect, which again depends on its level, delay and direction of incidence. The possible effects are (at least):

- Increased level (energy addition)
- Increased clarity, if it arrives within 50–80 ms after the direct sound
- Increased spaciousness, if the reflection direction in the horizontal plane is different from that of the direct as this causes the signals received at the two ears to be different (If the angle between the direct and reflected sounds differs in the vertical plane only, the effect is rather a change in *timbre* or coloration of the sound.)
- Echo, typically observed for delays beyond 50 ms and at high reflection levels. If the delay is very long, say 200 ms, then the echo may even be detected at a much lower level;
- Coloration. If the delay is short, say below 30 ms, and the level is high, phase addition of the direct sound and the reflection will create a *comb filter* effect, which severely deteriorates the original frequency spectrum
- Change in localization direction, in cases where the reflection is louder than the direct sound. This may happen either due to amplification of the reflection via a concave surface or due to excess attenuation of the direct sound, for instance by an orchestra pit rail.

Audibility thresholds, echo risk and other effects also depend on spectral and temporal properties of the signal itself. Thus, with speech and fast, staccato played music (not to speak of artificial, impulsive *click* sounds), echoes are much easier to detect than when the signal

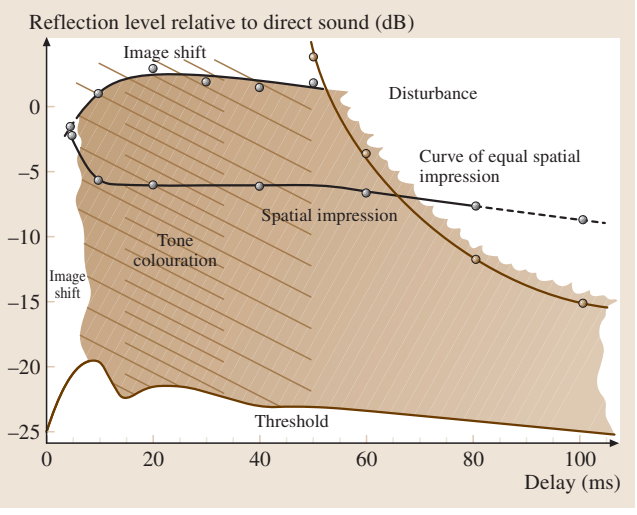


Fig. 9.3 Various audible effects of a single reflection arriving from the side. The signal used was music (after [9.4])

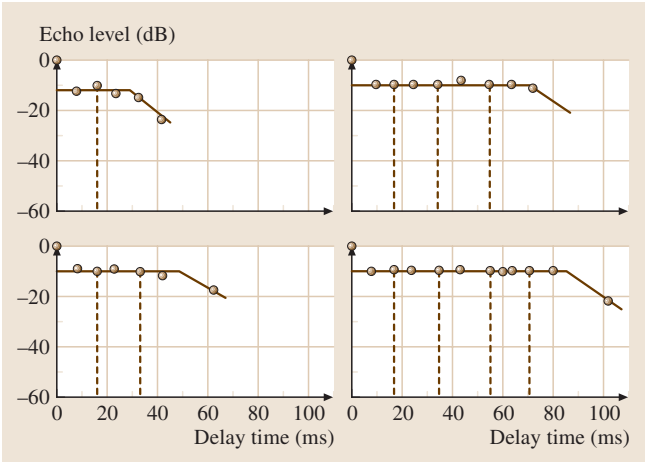


Fig. 9.4 Thresholds of perception of a new reflection (*circles* and *horizontal lines*) being added to impulse responses already containing one, two, three, or four other reflections with delay and relative levels, as indicated by the *vertical lines* (after [9.5])

is slow, legato music. On the other hand, these slower signals are more sensitive for the detection of coloration effects.

The temporal properties of sounds are sometimes described by the autocorrelation function. According to Ando (see Chap. 10 in this book) not only the thresholds of audibility but also the preferred delay of early reflections and the preferred reverberation time depend strongly on the autocorrelation of the signal. However, such a strong effect on concert hall preference – derived from listening experiments in rather simplified simulated sound fields in the lab – is not in accordance with everyday listening experiences in concert halls, in which we gladly accept listening to different music pieces of different tempo from different seats (different reflection delays) in the same hall (fixed reverberation time).

Figure 9.3 illustrates the various audible effects of a single reflection arriving from a 40° angle relative to a frontal direct sound. In this experiment music was used as the stimulus signal. Below the lower threshold (reflection level below about −20 dB and almost independent of the delay), the reflection is inaudible. In the upper-right region it causes a distinct echo, while in the shaded area we experience *spatial impression* or a broadening of the area from which the sound seem to originate. For delays lower than about 30 ms, this single reflection also causes an unpleasant change in *timbre* or *tonal color*, which is due to the combination of the direct and reflected signal forming a comb filter giving systematic changes in the spectrum. If the level of the reflection is increased to be louder than the direct sound (above 0 dB) while the delay is still below say 50 ms, we experience an *image shift* or a change in localization from the direction of the direct sound towards the direction from which the reflection is coming.

In practice, of course, the impulse response from a room contains many reflections, and we therefore need to know how the threshold of perceptibility of the incoming reflection changes in cases where the impulse response contains other reflections already. As seen from Fig. 9.4, the threshold level increases substantially in the delay range already occupied by other reflections. This phenomenon is primarily due to masking.

From this we can conclude that many of the details in a complex impulse response will be masked and only some of the dominant components or some of its overall characteristics seriously influence our perception of the acoustics. This is the reason why it is possible to create rather simple objective room acoustic parameters, as listed in the following section, which still describe the main features of the room acoustic experience. Thus, we seem to have some success describing the complex real world (the upper-middle and right boxes in Fig. 9.1) by simpler, abstract concepts. Also, it will be seen that the subjective effects caused by a single reflection remain important in the evaluation of more-realistic and complicated impulse responses.

9.3 Subjective and Objective Room Acoustic Parameters

From a consensus of numerous subjective experiments in real rooms and in simulated sound fields (representing all of the previously mentioned subjective research techniques), we now have a number of subjective parameters and corresponding objective measures available. Most of these are generally recognized as relevant descriptors of major aspects in our experience of the acoustics in rooms. This situation has promoted standardization of measurement methods and many of the objective parameters are now described in an appendix to the International Organization for Standardization (ISO) standard [9.6]. In order to maintain a clear distinction between the subjective and objective parameters in the following, the names for the subjective parameters will be printed in italics.

Amazon EX1014 Page 320

9.3.1 Reverberation Time

Reverberance is probably the best known of all subjective room acoustic aspects. When a room creates too much *reverberance*, speech loses intelligibility because important details (consonants) are masked by louder, lingering speech sounds (the vowels). For many forms of music, however, *reverberance* can add an attractive fullness to the sound by bonding adjacent notes together and blending the sounds from the different instruments/voices in an ensemble.

The reverberation time T which is the traditional objective measure of this quality, was invented 100 years ago by W. C. Sabine. T is defined as the time it takes for the sound level in the room to decrease by 60 dB after a continuous sound source has been shut off. In practice, the evaluation is limited to a smaller interval of the decay curve, from -5 dB to -35 dB (or -5 dB to -25 dB) below the start value; but still relating to a 60 dB decay (Fig. 9.5), i. e.:

$$T = 60\,\mathrm{dB}\,\frac{(t_{-35}) - (t_{-5})}{(-5\,\mathrm{dB}) - (-35\,\mathrm{dB})}. \tag{9.1}$$

In this equation, t_{-x} denotes the time when the decay has decreased to XdB below its start value, or, if we let $R(t)$ represent the squared value of the decaying sound pressure and shut off the sound source at time $t = 0$:

$$10\log_{10}\left(\frac{R(t_{-X})}{R(0)}\right) = -X\,\mathrm{dB}\,. \tag{9.2}$$

With the fluctuations always present in decay curves, T should rather be determined from the decay rate, A dB/s, as found from a least-squares regression line (determined from the relevant interval of the decay curve). Hereby we get for T:

$$T = \frac{60\,\mathrm{dB}}{A\,\frac{\mathrm{dB}}{\mathrm{s}}} = \frac{60}{A}\mathrm{s}\,. \tag{9.3}$$

Ways to obtain the decay curve from the impulse response will be further explained in the section on measurement techniques.

Due to masking, the entire decay process is only perceivable during breaks in the speech or music. Besides, the rate of decay is often different in the beginning and further down the decay curve. During running music or speech, the later, weaker part of the reverberation will be masked by the next syllable or musical note. Therefore an alternative measure, early decay time (EDT) has turned out to be better correlated with the *reverberance* perceived during running speech and music. This parameter, like T, also measures the rate of the decay; but now evaluated from the initial part, the interval between 0 and -10 dB, only. Thus,

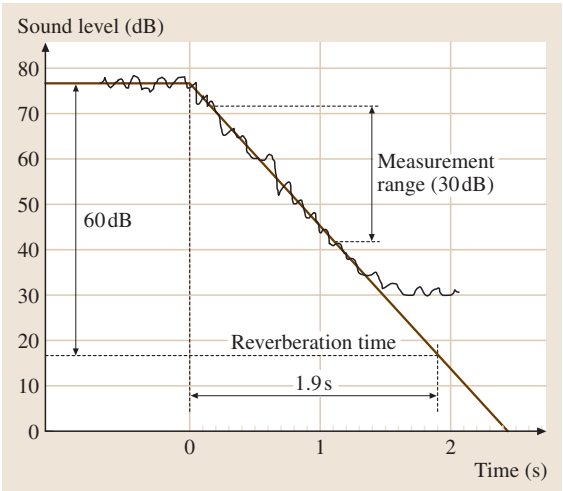


Fig. 9.5 The definition of reverberation time (after [9.4])

$$\mathrm{EDT} = 6(t_{-10}) \quad \text{or} \quad \mathrm{EDT} = \frac{60}{A_{(0\,\mathrm{dB}\to-10\,\mathrm{dB})}}\mathrm{s}\,. \tag{9.4}$$

The detailed behavior of the early part of the reverberation curve is influenced by the relative levels and distribution in time of the early reflections, which in turn vary depending on the positions of the source and receiver in the room. Likewise, the value of EDT is often found to vary throughout a hall, which is seldom the case with T.

In spite of the fact that EDT is a better descriptor of *reverberance* than T, T is still regarded the basic and most important objective parameter. This is mainly due to the general relationship between T and many of the other room acoustic parameters and because a lot of room acoustic theory relates to this concept, not least diffuse field theory, which is the basis for measurements of sound power, sound absorption, and sound insulation. T is also important by being referred to in legislation regarding room acoustic conditions in buildings.

Talking about diffuse field theory, it is often of relevance to compare the measured values of certain objective parameters with their expected values according to diffuse field theory and the measured or calculated reverberation time. As diffuse field theory predicts the decay to be purely exponential, the distribution in time of the impulse response squared should

Amazon EX1014 Page 321

follow the function:

$$h^2(t) = A \exp\left(\frac{-13.8}{T} t\right) , \tag{9.5}$$

in which the constant −13.8 is determined by the requirement that for $t = T$:

$$10 \log_{10}\left(\exp\left(\frac{-13.8}{T} t\right)\right) = 60\,\text{dB} . \tag{9.6}$$

With an exponential decay, the decay curve in dB becomes a straight line. Consequently, the expected value of EDT, EDT_{exp}, equals T.

When evaluating measurement results, it is also relevant to compare differences with the smallest change that can be detected subjectively. For EDT, this so-called subjective difference limen is about 5% [9.7].

9.3.2 Clarity

Clarity describes the degree to which every detail of the performance can be perceived as opposed to everything being blurred together by later-arriving, reverberant sound components. Thus, *clarity* is to a large extent a property complementary to *reverberance*.

When reflections are delayed by no more than 50–80 ms relative to the direct sound, the ear will integrate these contributions and the direct sound together, which means that we mainly perceive the effect as if the clear, original sound has been amplified relative to the later, reverberant energy. Thus, an objective parameter that compares the ratio between energy in the impulse response before and after 80 ms has been found to be a reasonably good descriptor of *clarity*

$$C = 10 \log_{10}\left[\int_0^{80\,\text{ms}} h^2(t)\,\text{d}t \Big/ \int_{80\,\text{ms}}^{\infty} h^2(t)\,\text{d}t\right] . \tag{9.7}$$

The higher the value of C, the more the early sound dominates, and the higher the impression of *clarity*.

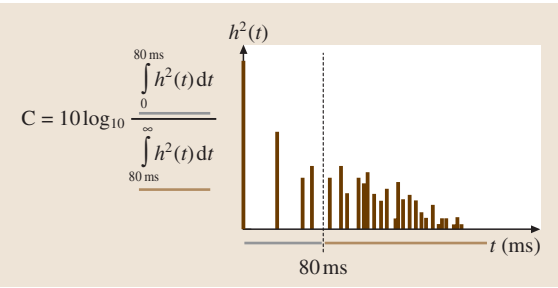


Fig. 9.6 The definition of C: the ratio between early and late energy in the impulse response

With exponential decay, the expected value of C becomes a function of T alone:

$$C_{\text{exp}} = 10 \log_{10}\left[\exp\left(\frac{1.104}{T}\right) - 1\right] \text{dB} . \tag{9.8}$$

The subjective difference limen for C (the equal best difference perceivable) is about 0.5 dB.

The definition of C is illustrated in Fig. 9.6.

Another parameter, which is also used to describe the balance between early and late sound or the balance between *clarity* and *reverberance*, is the center time t_s, which describes the center of gravity of the squared impulse response:

$$t_s = \int_0^{\infty} t h^2(t)\,\text{d}t \Big/ \int_0^{\infty} h^2(t)\,\text{d}t . \tag{9.9}$$

A low value of t_s corresponds to a clear sound, whereas higher values indicate dominance of the late, reverberant energy. The main advantage of t_s is that it does not contain a sharp time limit between early and late energy, since this sharp distinction is not justified by our knowledge about the functioning of our hearing system. The subjective difference limen for t_s is about 10 ms, and the expected diffuse field value is simply given by:

$$t_{s,\text{exp}} = \frac{T}{13.8} \tag{9.10}$$

9.3.3 Sound Strength

The influence of the room on the perceived *loudness* is another important aspect of room acoustics. A relevant measurement of this property is simply the difference in dB between the level of a continuous, calibrated sound source measured in the room and the level the same source generates at 10 m distance in anechoic surroundings. This objective measure called the (relative) strength G can also be obtained from impulse response recordings from the ratio between the total energy of the impulse response and the energy of the direct sound with the latter being recorded at a fixed distance (10 m) from the impulsive sound source:

$$G = 10 \log_{10} \frac{\int_0^{\infty} h^2(t)\,\text{d}t}{\int_0^{t_{\text{dir}}} h^2_{10\,\text{m}}(t)\,\text{d}t} . \tag{9.11}$$

Here the upper integration limit in the denominator t_{dir} should be limited to the duration of the direct sound pulse (which in practice will depend on the bandwidth

selected). A distance different from 10 m can be used, if a correction for the distance attenuation is applied as well.

The expected value of G according to diffuse field theory becomes a function of T as well as of the room volume, V:

$$G_{\mathrm{exp}} = 10\log_{10}\left(\frac{T}{V}\right) + 45\,\mathrm{dB}\,. \qquad (9.12)$$

The subjective difference limen for G is about 1.0 dB. The definition of G is illustrated in Fig. 9.7.

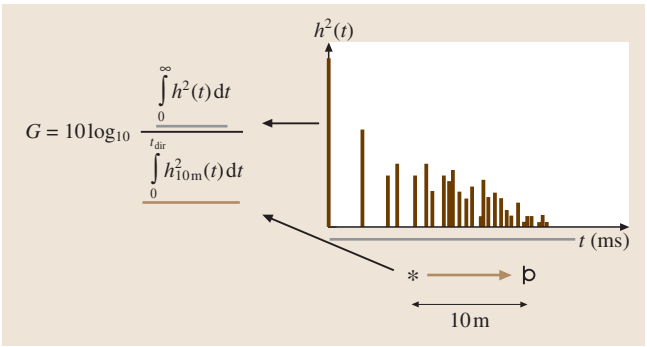


Fig. 9.7 The definition of strength G: the total energy in the impulse response measured relative to the direct sound level at 10 m distance from the source

9.3.4 Measures of Spaciousness

Spaciousness is the feeling that the sound is arriving from many different directions in contrast to a monophonic impression of all sound reaching the listener through a narrow opening. It is now clear that there are two aspects of *spaciousness*, both of which are attractive, particularly when listening to music:

- *apparent source width* (ASW): the impression that the sound image is wider than the visual, physical extent of the source(s) on stage. This should not be confused with localization errors, which of course should be avoided.

and

- *listener envelopment* (LEV): the impression of being inside and surrounded by the reverberant sound field in the room.

Both aspects have been found to be dependent on the direction of incidence of the impulse response reflections. When a larger portion of the early reflection energy (up to about 80 ms) arrives from lateral directions (from the sides) the ASW increases. When the level of the late, lateral reflections is high, strong LEV results.

The lateral components of the impulse response energy can be recorded using a figure-of-eight microphone with the sensitive axis held horizontal and perpendicular to the direction towards the sound source (so that the source lies in the deaf plane of the microphone). For measurement of the lateral energy fraction (LEF), the early part (up to 80 ms) of this lateral sound energy is compared with the energy of the direct sound plus all early reflections picked up by an ordinary omnidirectional microphone:

$$\mathrm{LEF} = \int_{t=5\,\mathrm{ms}}^{t=80\,\mathrm{ms}} h_1^2(t)\,dt \Big/ \int_{t=0\,\mathrm{ms}}^{t=80\,\mathrm{ms}} h^2(t)\,dt\,, \qquad (9.13)$$

where h_1 is the impulse response pressure recorded with a figure-of-eight microphone, whereas h is captured through the (usual) omnidirectional microphone.

It is mainly the energy at low and mid frequencies that contribute to the *spaciousness*. Consequently, LEF is normally averaged over the four octaves 125–1000 Hz. The higher the value of LEF, the wider the ASW. The literature contains many data on LEF values in different, existing concert halls.

LEF does not have an expected value related to T. In a completely diffuse field, LEF would be constant with a value of 0.33, which is higher than that normally found in real halls. The subjective difference limen for LEF is about 5%.

The definition of LEF is illustrated in Fig. 9.8.

The ASW aspect of *spaciousness* is not only dependent on the ratio between early lateral and total early

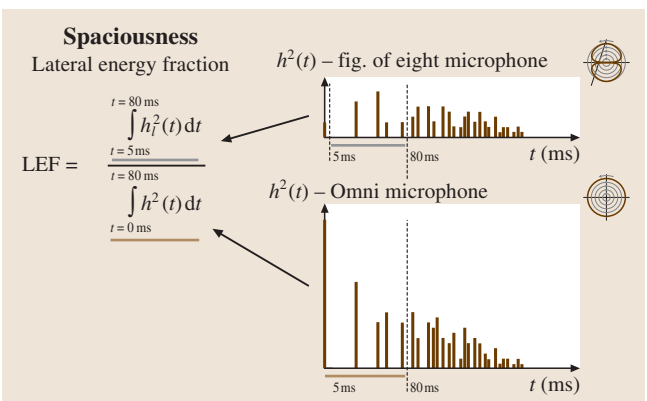


Fig. 9.8 The definition of lateral energy fraction LEF: the ratio between early reflection energy arriving from the sides and total early energy

sound; but also on the total level of the sound. The higher the G value (and the louder the sound source), the broader the acoustic image of the source. However, at the time of writing, there is no solid way of incorporating the influence of level into the objective measure.

Listener envelopment seems to be determined mainly by the spatial distribution and the level of the late reflections (arriving after 80 ms). A parameter called *late lateral strength* (LG) relating the late lateral energy to the direct sound at 10 m distance from the source has been proposed to measure this quality [9.8]:

$$\mathrm{LG} = \int_{t=80\,\mathrm{ms}}^{t=\infty} h_1^2(t)\,\mathrm{d}t \Big/ \int_{t=0\,\mathrm{ms}}^{t=t_{\mathrm{dir}}} h_{10\,\mathrm{m}}^2(t)\,\mathrm{d}t\,. \qquad (9.14)$$

LG is likely to be included in the ISO 3382 standard (from 2007); but so far LG has only been found correlating with subjective responses in laboratory experiments using synthetic sound fields. Therefore, it may be wise not to put too much emphasis on this parameter until its value has also been confirmed by experiments in real halls.

In contrast to reflections arriving in the median plane, lateral reflections will cause the instantaneous pressure at the two ears of the listener to be different. Such a dissimilarity of the signals at the two ears can also be measured by means of the interaural cross-correlation coefficient IACC:

$$\mathrm{IACC}_{t_1,t_2} = \max \left| \frac{\int_{t_1}^{t_2} h_{\mathrm{L}}(t) h_{\mathrm{R}}(t+\tau)\,\mathrm{d}t}{\sqrt{\int_{t_1}^{t_2} h_{\mathrm{L}}^2(t)\,\mathrm{d}t \int_{t_1}^{t_2} h_{\mathrm{R}}^2(t)\,\mathrm{d}t}} \right| \qquad (9.15)$$

in which t_1 and t_2 define the time interval of the impulse response within which the correlation is calculated, h_{L} and h_{R} represent the impulse response measured at the entrance of the left and right ear, respectively, and τ is the interval within which we search for the maximum of the correlation. Normally, the range of τ is chosen between -1 and $+1$ ms, covering roughly the time it takes for sound to travel the distance from one ear to the other. As the correlation is normalized by the product of the energy of h_{L} and h_{R}, the IACC will take a value between 0 and 1. Results are often reported as $1-\mathrm{IACC}$ in order to obtain a value increasing with increasing dissimilarity, corresponding to an increasing impression of *spaciousness*. If the time interval for the calculation (t_1, t_2) is (0 ms, 100 ms), then the IACC will measure the ASW, while a later interval $(t_1, t_2) = (100\,\mathrm{ms}, 1000\,\mathrm{ms})$ may be used to describe the LEV. The literature on measured values in halls mainly contain data on IACC related to the 0–100 ms time interval.

Although the LEF and IACC parameters relate to the same subjective quality, they are not highly correlated in practice. Another puzzling fact is that LEF and IACC emphasize different frequency regions being of importance. LEF is primarily measured in the four lowest octaves, 125 Hz, 250 Hz, 500 Hz and 1000 Hz while IACC should rather be measured in the octave bands above 500 Hz. IACC values would always be high in the lower octaves, because the distance between the ears (< 30 cm) is small compared to 1/4 of the wave length (≈ 70 cm at 125 Hz).

9.3.5 Parameters Relating to Timbre or Tonal Color

Timbre describes the degree to which the room influences the frequency balance between high, middle and low frequencies, i. e. whether the sound is *harsh*, *bright*, *hollow*, *warm*, or whatever other adjective one would use to describe *tonal color*. Traditionally, a graph of the frequency variation of T (per 1/1 or 1/3 octave) has been used to indicate this quality; but a convenient single-number parameter intended to measure the warmth of the hall has been suggested [9.9]: the bass ratio (BR) given by:

$$\mathrm{BR} = \frac{T_{125\,\mathrm{Hz}} + T_{250\,\mathrm{Hz}}}{T_{500\,\mathrm{Hz}} + T_{1000\,\mathrm{Hz}}}\,. \qquad (9.16)$$

Likewise, a *treble ratio* (TR) can be formed as:

$$\mathrm{TR} = \frac{T_{2000\,\mathrm{Hz}} + T_{4000\,\mathrm{Hz}}}{T_{500\,\mathrm{Hz}} + T_{1000\,\mathrm{Hz}}}\,. \qquad (9.17)$$

However, in some halls, a lack of bass sound is experienced in spite of high T values at low frequencies. Therefore EDT or perhaps G versus frequency would be a better – and intuitively more logical – parameter for measurement of *timbre*. Likewise, BR or TR could be based on G rather than T values.

Besides the subjective parameters mentioned above, a quality called *intimacy* is also regarded as being important when listening in auditoria. *Beranek* [9.9] originally suggested this quality to be related to the initial time delay gap ITDG; but this has not been experimentally verified (At the time of writing, Beranek no longer advocates this idea but mentions the observation that, if the ITDG exceeds 45 ms, the hall has no intimacy [9.10]). It is likely that *intimacy* is related to a combination of some of the other parameters already mentioned, such as a high sound level combined with a clear and envelop-

ing sound, as will naturally be experienced fairly close to the source or in small rooms.

9.3.6 Measures of Conditions for Performers

In rooms for performance of music it is also relevant to consider the acoustic conditions for the musicians, partly because it is important to ensure that the musicians are given the best working conditions possible and partly because the product as heard by the audience will also be better if the conditions are optimal for the performers.

Musicians will be concerned about *reverberance* and *timbre* as mentioned above; but also about at least two aspects which are unique for their situation: *ease of ensemble* and *support* [9.11] and [9.12].

Ease of ensemble relate to how well musicians can hear – and play together with – their colleagues. If ensemble is difficult to achieve, the result – as perceived by musicians and listeners alike – might be less precision in rhythm and intonation and a lack of balance between the various instruments. *Ease of ensemble* has been found to be related to the amount of early reflection energy being distributed on the stage. So far, the only widely recognized objective parameter suggested for objective measurement of this quality is *early support* $\mathrm{ST}_{\mathrm{early}}$.

$\mathrm{ST}_{\mathrm{early}}$ is calculated from the ratio between the early reflection energy and the direct sound in an impulse response recorded on the stage with a distance of only one meter between the source and receiver:

$$\mathrm{ST}_{\mathrm{early}} = 10\log_{10}\frac{\int_{20\,\mathrm{ms}}^{100\,\mathrm{ms}} h_{1\,\mathrm{m}}^2(t)\,\mathrm{d}t}{\int_{0\,\mathrm{ms}}^{t_{\mathrm{dir}}} h_{1\,\mathrm{m}}^2(t)\,\mathrm{d}t}\,. \tag{9.18}$$

Support relates to the degree to which the room supports the musicians' efforts to create the tone on their own instruments, whether they find it easy to play or whether they have to force their instruments to fill the room. Having to force the instrument leads to playing fatigue and inferior quality of the sound.

Support is also related to the amount of reflected energy on the stage measured using only a one meter source/microphone distance. For measurement of *support*, however, one needs to consider also the later reflections. This is because for many types of instruments, (especially strings), the early reflections will be masked by the strong direct sound from the instrument itself. Consequently, it seems relevant to define a measure

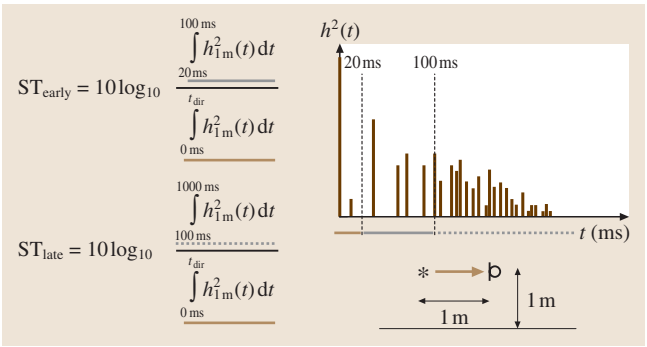


Fig. 9.9 The definition of early support $\mathrm{ST}_{\mathrm{early}}$ and late support $\mathrm{ST}_{\mathrm{late}}$, the ratio between early and late reflection energy respectively and the direct sound in the impulse response. Both are measured at 1 m distance from the source

such as *late support*, $\mathrm{ST}_{\mathrm{late}}$:

$$\mathrm{ST}_{\mathrm{late}} = 10\log_{10}\frac{\int_{100\,\mathrm{ms}}^{1000\,\mathrm{ms}} h_{1\,\mathrm{m}}^2(t)\,\mathrm{d}t}{\int_{0\,\mathrm{ms}}^{t_{\mathrm{dir}}} h_{1\,\mathrm{m}}^2(t)\,\mathrm{d}t} \tag{9.19}$$

which relates to late response (after 100 ms) from the room to the sound emitted.

$\mathrm{ST}_{\mathrm{early}}$ and $\mathrm{ST}_{\mathrm{late}}$ are typically measured in the four octaves 150 Hz, 500 Hz, 1000 Hz, 2000 Hz. The lowest octave is omitted primarily because it is not possible to isolate the direct sound from the early reflections in a narrow band measurement. The definitions of early and late support are illustrated in Fig. 9.9.

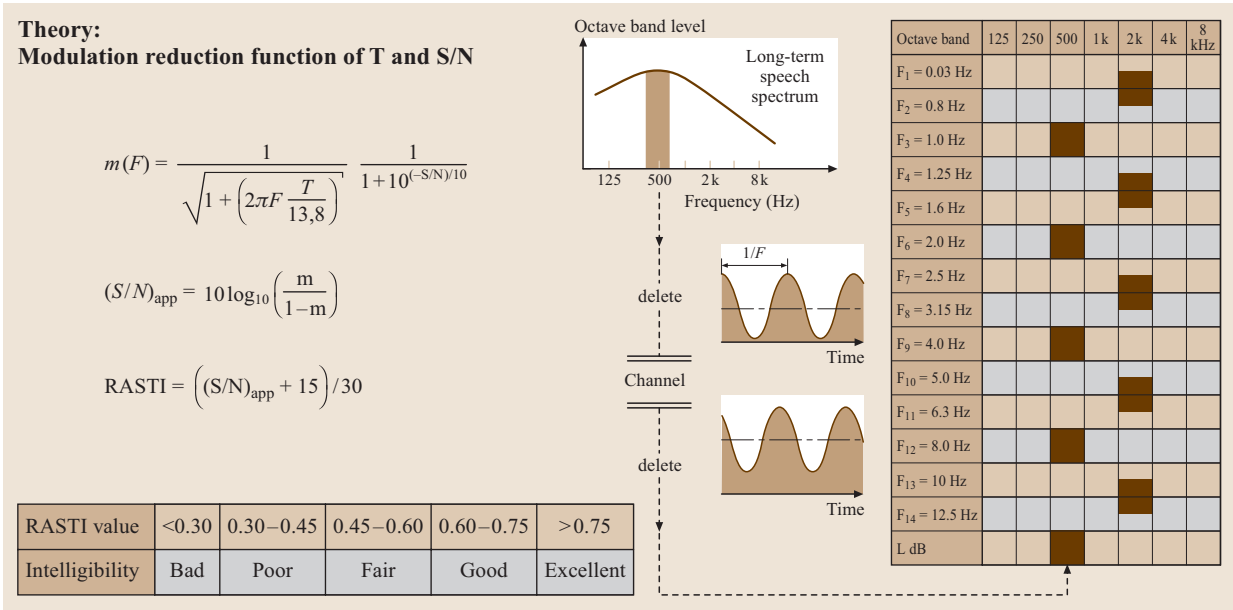
Also the support parameters are planned to be included in the ISO 3382 standard (from 2007); but like LG the experiences and amount of data available from existing halls are still rather limited.

The amount of *reverberance* on the stage may be measured by means of EDT, (with a source–receiver distance of, say, 5 m or more).

9.3.7 Speech Intelligibility

All the objective parameters mentioned above (except the basic T parameter), are mainly relevant in larger auditoria intended for performance of music. In auditoria used for speech, such as lecture halls or theaters, the influence of the acoustics on intelligibility is a major issue.

Currently, the most common way to assess objectively *speech intelligibility* in rooms is by measurement of the speech transmission index STI.



RASTI value	<0.30	0.30–0.45	0.45–0.60	0.60–0.75	>0.75
Intelligibility	Bad	Poor	Fair	Good	Excellent

Fig. 9.10 Illustration of the theory and principle in the measurement of STI or RASTI. The scale for the evaluation of RASTI values is shown at the bottom (after [9.13])

As illustrated in Fig. 9.10, this measure is based on the idea that speech can be regarded as an amplitude-modulated signal in which the degree of modulation carries the speech information. If the transmission path adds noise or reverberation to the signal, the degree of modulation in the signal will be reduced, resulting in reduced intelligibility.

The modulation transfer is tested by emitting noise in seven octave bands, each modulated with 14 different modulation frequencies as listed in the table in Fig. 9.10 and then calculating the ratio between the original and the received degree of modulation, the modulation reduction factor, in each of these 98 combinations. A weighted average of the modulation reduction factor then results in a number between 0 and 1, corresponding to very poor and excellent conditions respectively.

A faster measurement method using only two carrier bands of noises and four plus five modulation frequencies (indicated by the dark squares in Fig. 9.10) is called rapid STI (RASTI). The STI/RASTI method is described in an International Electrotechnical Commission (IEC) standard: IEC 286-16.

Although the original method of STI or RASTI measurement employs modulated noise signals, it is also possible to calculate the modulation reduction factor from the impulse response. Thus, the modulation reduction factor versus modulation frequency F, which is called the modulation transfer function (MTF), can be found as the Fourier transform of the squared impulse response normalized by the total impulse response energy.

According to *Bradley* [9.14], a simpler parameter, called the *useful-to-detrimental sound ratio* U_{80} is equally suitable for measurement of speech intelligibility. U_{80} is simply a modified version of the clarity parameter defined in Sect. 9.3.2, in which a correction is made according to the ratio between the levels of speech and background noise.

9.3.8 Isn't One Objective Parameter Enough?

It is important to notice that all the parameters mentioned above – apart from T – may show large differences between different seats within the same hall. Actually, differences within a single hall are sometimes as large (both objectively and subjectively) as the differences between two different halls.

It should also be mentioned that the objective parameters related to the three subjectively different aspects: *reverberance/clarity*, *loudness* and *spaciousness* show low mutual correlation when measured in different seats or in different halls. In other words, they behave as orthogonal factors and do not monitor the same properties

Amazon EX1014 Page 326

of the sound fields. Consequently, it is obviously necessary to apply one objective parameter for each of these subjective aspects.

It is also clear that different listeners will judge the importance of the various subjective parameters differently. While some may base their judgment primarily on *loudness* or *reverberance* others may be more concerned about *spaciousness* or *timbre*.

For these reasons, one should be very sceptical when people attempt to rank concert halls along a single, universal, one-dimensional quality scale or speak about the best hall in the world. Fortunately, the diversity of room acoustic design and the complexity of our perception does not justify such simplifications.

9.3.9 Recommended Values of Objective Parameters

In view of the remarks above, it may be regarded risky to recommend values for the various objective acoustic parameters. On the other hand, there is a long tradition of suggesting optimal values for reverberation time T as a function of hall volume and type of performance. Besides, most of the other parameters will seldom deviate drastically from the expected value determined by T. As we shall see later, this deviation is primarily influenced by how we shape the room to control early reflections. Designing for strong early reflections can increase *clarity*, *sound strength* and *spaciousness*. Current taste regarding concert hall acoustics for classical music seems to be in favor of high values for these factors in addition to strong *reverberance*. This could lead to suggested ranges for the various parameters in small and large halls, as shown in Table 9.1. These values relate to empty halls with well-upholstered seating, assuming that, when fully occupied with musicians and audience, the T values will drop by no more than say 0.2 s. The reason for relating the values to the unoccupied condition is that it is seldom possible to make acoustic measurements in the occupied state and almost all reference data existing regarding parameters other than T are from unoccupied halls. On the other hand, suggesting well-upholstered seats is a sound recommendation in most cases, which will ensure only small changes depending on the degree of occupancy. This will also justify extrapolation of the unoccupied values of the other parameters to the occupied condition, as mentioned later in this section.

The values listed in the table were arrived at by first choosing values for T ensuring high *reverberance*, while the correlative values for EDT, G and C stem from the diffuse field expected values; but these have been slightly changed to promote high *clarity* and high *levels*. As this is particularly important in larger halls, it has been suggested to use an EDT value 0.1 s lower than T ($=\mathrm{EDT}_{\mathrm{exp}}$) and C values 1 dB higher than C_{exp} in a 2500 m^3 hall, but 0.2 s less than T and 2 dB higher than C_{exp}, respectively, in the 25 000 m^2 hall. Likewise, it is suggested to use G values 2 dB less than G_{exp} in the small halls; but only 1 dB less than or even equal to G_{exp} in large halls. For halls of size between the 2500 and 25 000 m^3 listed in the table, one may of course interpolate to taste. Please note that, in general, G is found to be 2–3 dB lower that G_{exp} (Sect. 9.5.6) so the G values suggested here are actually 1–2 dB higher than those found in "normal" halls.

In the occupied condition, one may expect EDT, C and G to be reduced by amounts equal to the difference between the expected values calculated using T for the empty and occupied conditions respectively. In any case,

Table 9.1 Suggested position-averaged values of objective room acoustic parameters in unoccupied concert halls for classical music. It is assumed that the seats are highly absorptive, so that T will be reduced by no more than 0.2 s when the hall is fully occupied. 2.4 s should be regarded as an upper limit mainly suitable for terraced arena or DRS halls with large volumes per seat (Sect. 9.7.3), whereas 2.0 s will be more suitable for shoe-box-shaped halls, which are often found to be equally reverberant with lower T values and lower volumes per seat

Parameter	Symbol	Chamber music	Symphony
Hall size	V/N	2500 m^3/300 seats	25 000 m^3/2000 seats
Reverberation time	T	1.5 s	2.0–2.4 s
Early decay time	EDT	1.4 s	2.2 s
Strength	G	10 dB	3 dB
Clarity	C	3 dB	−1 dB
Lateral energy fraction	LEF	0.15–0.20	0.20–0.25
Interaural cross correlation	1 − IACC	0.6	0.7
Early support	$\mathrm{ST}_{\mathrm{early}}$	−10 dB	−14 dB

if the empty–occupied difference in T is limited to 0.2 s, as suggested here, then the difference in these parameters will be fairly limited, with changes in both C and G being less than one dB - given that the audience does not introduce additional grazing incidence attenuation of direct sound and early reflections (Sect. 9.6.2).

For opera, one may aim towards the same relationships between the diffuse field expected and the desired values of EDT, C and G, as mentioned above for symphonic halls; but the goal for the reverberation time will normally be set somewhat lower, 1.4–1.8 s in order to obtain a certain intelligibility of the libretto and perhaps to make breaks in the music sound more dramatic (Sect. 9.7.2).

For rhythmic music, T values of 0.8–1.5 s, depending on room size, are appropriate, and certainly the value should not increase towards low frequencies in this case. In very large rock venues such as sports arenas, it may actually be very difficult to get T below 3–5 s; but even then the conditions can be satisfactory with a properly designed sound system. The reason is that in such large room volumes the level of the reverberant sound can often be controlled, if highly directive loudspeakers are being used. In these spaces, the biggest challenge is often to avoid highly delayed echoes, which are very annoying. For amplified music, only recommendations for T are relevant, because the acoustic aspects related to the other parameters will be determined primarily by the sound system (Sect. 9.7.5).

In order to ensure adequate speech intelligibility in lecture halls and theaters, values for STI/RASTI should be at least 0.6. In reverberant houses of worship values higher than 0.55 are hard to achieve even with a well-designed loudspeaker system.

9.4 Measurement of Objective Parameters

Whenever the acoustic specifications of a building are of importance we need to be able to predict the acoustic conditions during the design process as well as document the conditions after completion. Therefore, in this and the following sections, techniques for measurement and prediction of room acoustic conditions will be presented.

In Sect. 9.2.1 it was claimed that the impulse response contains all relevant information about the acoustic conditions in a room. If this is correct, it must also be true that all relevant acoustic parameters can be derived from impulse response measurements. Fortunately, this has also turned out to be the truth.

9.4.1 The Schroeder Method for the Measurement of Decay Curves

Originally, reverberation decays were recorded from the decay of the sound level following the termination of a stationary sound source. However, *Schroeder* [9.15] has shown that the reverberation curve can be measured with increased precision by *backwards* integration of the impulse response, $h(t)$, as follows:

$$R(t) = \int_t^\infty h^2(t)\,\mathrm{d}t = \int_0^\infty h^2(t)\,\mathrm{d}t - \int_0^t h^2(t)\,\mathrm{d}t \quad (9.20)$$

in which $R(t)$ is equivalent to the decay of the squared pressure decay. The method is called backwards integration because the fixed upper integration limit, *infinite time*, is not known when we start the recording. Therefore, in the days of magnetic tape recorders, the integration was done by playing the tape backwards. In this context *infinite time* means the duration of the recording, which should be comparable with the reverberation time. Traditional noise decays contain substantial, random fluctuations because of interference by the different eigenmodes present within the frequency range of the measurement (normally 1/3 or 1/1 octaves). These fluctuations will be different each time the measurement is carried out because the modes will be excited with random phase by the random noise (and the random time of the noise interruption). When, instead, the room is exited by a repeatable impulse, the response will be repeatable as well without such random fluctuations. In fact, Schroeder showed that the ensemble average of interrupted noise decays (recorded in the same source and receiver positions) will converge towards $R(t)$ as the number of averages goes towards infinity.

In (9.20), the right-hand side indicates that the impulse response decay can be derived by subtracting the running integration (from time zero to t) from the total energy. Contrary to the registration of noise decays, this means that the entire impulse response must be stored before the decay function is calculated. However, with modern digital measurement equipment, this is not a problem. From the $R(t)$ data, T is calculated by linear regression as explained in Sect. 9.3.1. As the EDT is evaluated from a smaller interval of the decay curve than T, the EDT is more sensitive to random decay fluctuations than is T. Therefore, EDT should never

be derived from interrupted noise decays – only from integrated impulse responses.

9.4.2 Frequency Range of Measurements

T may be measured per 1/3 octave; but regarding the other objective parameters mentioned in this chapter, it makes no sense to use a frequency resolution higher than 1/1 octave. The reason is that the combination of narrow frequency intervals and narrow time windows in the other parameters will lead to extreme fluctuations with position of the measured values due to the acoustic relation of uncertainty. Such fluctuations have no parallel in our subjective experience. *Bradley* [9.16] has shown that, with 1/1 octave resolution, a change in microphone position of only 30 cm can already cause fluctuations in the clarity C and strength G of about 0.5–1 dB, which is equal to the subjective difference limen. Higher fluctuations would not make sense, as normally we do not experience an audible change when moving our head 30 cm in a concert hall.

The frequency range in which room acoustic measurements are made is normally the six octaves from 125 to 4000 Hz, but it may be extended to 63–8000 Hz when possible. Particularly for reverberation measurements in halls for amplified music, the 63 Hz octave is important.

9.4.3 Sound Sources

With the possibility of deriving all of the objective parameters described above from impulse response measurements, techniques for creating impulses are described briefly in the following.

Basically, the measurements require an omnidirectional sound source emitting short sound pulses (or other signals that can be processed to become impulses) covering the frequency interval of interest, an omnidirectional microphone and a medium for the storage of the electrical signal generated by the microphone.

Impulsive sources such as pistols firing blank cartridges, bursting balloons, or paper bags and electrical spark sources may be used; but these sources are primarily used for noncritical measurements, as their directivity and acoustic power are not easy to control.

For measurements requiring higher precision and repeatability, omnidirectional loudspeakers are preferable, as the power and directivity characteristics of loudspeakers are more stable and well defined. Omnidirectional loudspeakers are often built as six units placed into the sides of a cube, 12 units placed into a dodecahedron or 20 units in an icosahedron, see Fig. 9.11. The requirements regarding omnidirectivity of the source are described in the ISO 3382 standard.

Loudspeakers, however, cannot always emit sufficient acoustic power if an impulsive signal is applied directly. Therefore, special signals of longer duration such as maximum-length sequences or tone sweeps have been developed, which have the interesting property that their autocorrelation functions are band-limited delta pulses. This means that the loudspeaker can emit a large amount of energy without challenging its limited peak-power capability while impulse responses with high time resolution can still be obtained afterwards through post-processing (a correlation/convolution process) of the recorded noise or sweep responses.

In certain cases, other source directivity patterns than omnidirectional can be of interest. For example, a directivity like that of a human talker may be relevant if the measurement is related to the intelligibility of speech or to measurements on the stage in an opera house. A suitable talker/singer directivity can be obtained by using a small loudspeaker unit (membrane diameter about three to four inches) mounted in a closed box of size comparable to that of the human head. Using only one of the loudspeakers in a cube or dodecahedron speaker of limited size is another possibility.

9.4.4 Microphones

For most of the parameters mentioned, the microphone should be omnidirectional, while omni plus figure-of-eight directivity and dummy heads are used when the impulse responses are to be used for the calculation of LEF and IACC.

Three-dimensional intensity probes, sound field microphones or microphone arrays have also been suggested in attempts to obtain more-complete information about the amplitude and direction of incidence of individual reflections [9.17] or for wave field analysis [9.18]. However, such measurements are mainly for research or diagnostic purposes. As such, they do not fulfill the goal set up in Sect. 9.1 of reducing the information to what is known to be subjectively relevant. A selection of relevant microphones are shown in Fig. 9.12.

9.4.5 Signal Storage and Processing

For the storage of the impulse responses the following choices appear.

- The hard disk on a personal computer (PC) equipped with a sound card or special analog-to-digital (A/D) converter card

- Real-time (1/3 and 1/1 octave) analyzers or sound level meters with sufficient memory for storage of a series of short time-averaged root-mean-square (RMS) values
- Fast Fourier transform (FFT) analysers capable of recording adequately long records
- Tape recorders, preferably a digital audio tape (DAT) recorder, or even MP3 recorders given that the data reduction does not deteriorate the signal.

Of these, real-time analyzers and sound level meters will process the signal through an RMS detector before storage, which means that not the full signal, but only energy with a limited time resolution will be recorded.

Today, calculation of the objective parameters will always be carried out by a PC or by a computer built into a dedicated instrument. A large number of PC-based systems and software packages particularly suited for room acoustic measurements are available. Some systems come with a specific analog-to-digital (A/D), digital-to-analog (D/A) and signal-processing card, which must be installed in or connected to the PC, while others can use the sound card already available in most PCs as well as external sound cards for hobby music use.

9.5 Prediction of Room Acoustic Parameters

Beyond theoretical calculation of the reverberation time, the prediction techniques presented in this section range from computer simulations and scale model measurements to simple linear models based on empirical data collected from existing halls.

In all cases, the main objective is the prediction of the acoustic conditions described in terms of the acoustic parameters presented in Sect. 9.3. Just as we can record the sound in a hall, both scale models and numerical computer simulations can also provide *auralizations*, which are synthesized audible signals that allow the client and architect to listen to and participate in the acoustic evaluation of proposed design alternatives. Such audible simulations will often have much greater impact and be more convincing than the acoustician's verbal interpretation of the numerical results. Therefore, the acoustician must be very careful to judge the fidelity and the degree of realism before presenting auralizations.

9.5.1 Prediction of Reverberation Time by Means of Classical Reverberation Theory

Reverberation time as defined in Sect. 9.3.1 is still the most important objective parameter for the characterization of the acoustics of a room. Consequently, the prediction of T is a basic tool in room acoustic design. Calculation of reverberation time according to the Sabine equation was the first attempt in this direction and today is still the most commonly used. However, during the 100 years since its invention, several other reverberation time formulae have been suggested (by Eyring, Fitzroy, Millerton and Sette, Metha and Mulholland Kuttruff and others), all with the purpose of correcting some obvious shortcomings of the Sabine method.

The Sabine equation has been described in Sect. 11.1.4. In larger rooms, the absorption in the air plays a major role at high frequencies. If we include this factor, the Sabine equation reads:

$$T = \frac{0.161V}{S\alpha^* + 4mV}\,, \tag{9.21}$$

where V is the room volume, S is the total area of the room surfaces, α^* is the area-weighted average absorption coefficient of the room surfaces

$$\alpha^* = \overline{\alpha} = \frac{1}{S}\sum_i S_i\alpha_i \tag{9.22}$$

and m accounts for the air absorption, which is a function of the relative humidity and increases rapidly with fre-



Fig. 9.11 Two types of omnidirectional sound sources for room acoustic measurements: blank pistol (in the *raised hand of the seated person*) and an omnidirectional loudspeaker (icosahedron with 20 units)

Table 9.2 Air absorption coefficient m for different frequencies and relative humidity; valid for an air temperature of 20 °C (after [9.19])

Relative humidity (%)	0.5 kHz	1 kHz	2 kHz	4 kHz	8 kHz	Air absorption
40	0.4	1.1	2.6	7.2	23.7	$10^{-3}\,\mathrm{m}^{-1}$
50	0.4	1.0	2.4	6.1	19.2	$10^{-3}\,\mathrm{m}^{-1}$
60	0.4	0.9	2.3	5.6	16.2	$10^{-3}\,\mathrm{m}^{-1}$
70	0.4	0.9	2.1	5.3	14.3	$10^{-3}\,\mathrm{m}^{-1}$
80	0.3	0.8	2.0	5.1	13.3	$10^{-3}\,\mathrm{m}^{-1}$

quency [9.19]. Values of m versus frequency and relative humidity are listed in Table 9.2.

The Sabine equation assumes the absorption to be evenly distributed on all surfaces, and that the sound field is *diffuse*, which means that in each point of the room:

1. the energy density is the same, and
2. there is equal probability of the sound propagating in all directions.

Among the alternative reverberation formulae, only the Eyring method will be described here, as it may give more-accurate predictions in cases where the room is highly absorptive.

In contrast to Sabine, the Eyring theory assumes that the sound field is composed of plane sound waves that lose a fraction of their energy equal to the absorption coefficient of the surface whenever the wave hits a surface in the room. Thus, after n reflections, the energy in all wavefronts and the average energy density in the room has been reduced to $(1-\overline{\alpha})^n$ times the original value. The average distance of propagation during this process is $l_\mathrm{m} n = ct$, in which l_m is the *mean free path*, equal to the average distance traveled by the wave between two reflections, and t is the time elapsed since the wave propagation started. *Kosten* [9.20] has shown, that regardless of the room shape, l_m equals $4V/S$, if all directions of propagation are equally probable. These assumptions lead to the Eyring formula, which can be expressed by substituting α^* in (9.21) by

$$\alpha^* = \alpha_\mathrm{e} = \ln\left(\frac{1}{1-\overline{\alpha}}\right) \tag{9.23}$$

For low values of $\overline{\alpha}$, $\overline{\alpha}$ and α_e are almost identical, causing the Sabine and the Eyring formulae to give nearly identical results; but the difference becomes noticeable when $\overline{\alpha}$ exceeds about 0.3. In general, α_e will always be larger than $\overline{\alpha}$, leading to $T_\mathrm{Sabine} > T_\mathrm{Eyring}$. In the extreme case of the mean absorption approaching 1, α_e converges towards ∞, whereby $T_\mathrm{Eyring} \to 0$ as one would expect for a totally absorbing room. However, T_Sabine converges towards the finite positive value 0.16 V/S, which of course is unrealistic.

Both the Sabine and the Eyring theories suffer from at least two shortcomings. Neither of them consider the actual distribution of the free paths (see the comments on the *mean free path* above), which is highly dependent on the room shape, nor the often very uneven distribution of the absorption in the room, which causes different directions of sound propagation not to be equally probable. Typical situations where these factors become important are rooms with highly absorbing acoustic ceilings and hard walls and floors – or the reverse situation of an auditorium, where almost all absorption is found in the seating. In these situations the energy is certainly not evenly distributed, there is a considerable distribution of the path lengths, and there is a much higher probability for sound energy to travel towards the absorbing surface than away from it.

A common example in which the Sabine and Eyring theories fall short is the simple case of a rectangular room with plane walls, absorbing ceiling and low height compared to length and width. In such a room the main part of the late reverberant energy will be stored in

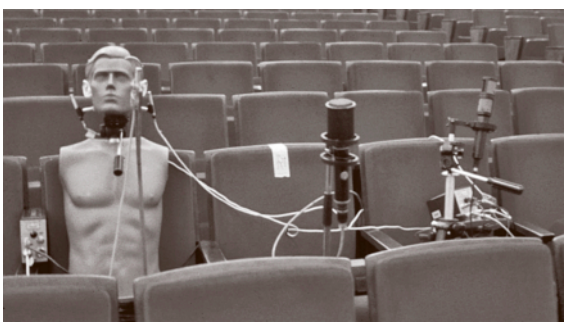


Fig. 9.12 A selection of microphones types used for recording impulse responses in rooms. *From left*: artificial head (with both internal and external microphones), *sound field microphone* (with four capsules for 3-D recordings of reflections) and a two-channel (stereo) microphone with omnidirectional and figure-of-eight capsules

Amazon EX1014 Page 331

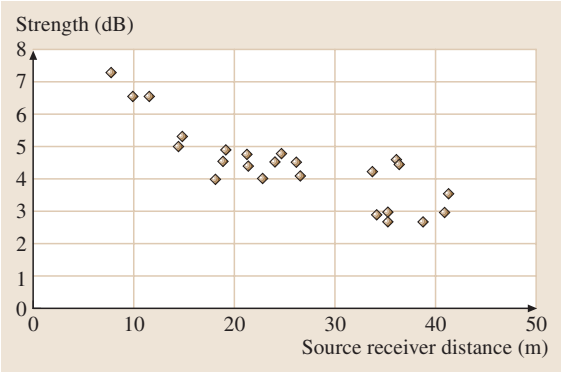


Fig. 9.13 Strength versus distance measured in the Boston Symphony Hall; values averaged over the four octaves 250–2000 Hz. The reverberation distance in this hall is about 5 m

one- and two-dimensional horizontal room modes with long free paths, whereas the three-dimensional modes with shorter free paths will soon have lost their energy by interacting with the absorbing ceiling. Hereby, different modes end up having very different decay rates, leading to bent or double-sloped decay curves and to measured T values often longer than calculated. In fact, only the unnatural situation with an even absorption distribution and all free paths of equal length will result in a straight-line decay and a well-defined T value.

From practical measurements of the distribution of G values, it is also clear that the assumption of an even distribution of the energy density is not met in large auditoria. Contrary to the Sabine theory, which predicts a constant level beyond the reverberation distance (Fig. 11.6), we observe that the level continues to decrease with distance from the source – even far beyond the reverberation distance, where the contribution of the direct sound is negligible. An example of this is seen in Fig. 9.13. (Based on empirical data, simple equations predicting this attenuation of G with distance as a function of room geometry have been derived as described in Sect. 9.5.6.)

Still, rather than relying on another of the reverberation formulas to provide a more correct result, today it may be wiser just to apply the Sabine and perhaps the Eyring equations, keeping in mind their limitations. Thus, *Beranek* [9.21] has suggested that, for Sabine calculations in concert halls, one should apply different sets of seat absorption values depending on the room shape. These values should be measured in nearly the same type of room as the hall in question.

When higher prediction accuracy is required, it is recommended to carry out simulations in a computer model as described in Sect. 9.5.4. However, also with this calculation approach, the question about which absorption values to use is relevant.

9.5.2 Prediction of Reverberation in Coupled Rooms

Auditoria with coupled spaces form another situation in which the decay curve may end up being nonlinear and the energy density will be different in different areas of the room. Typical of such cases are open-plan office floors coupled to an atrium covering several storeys, seating areas in auditoria subdivided by (deep) balconies, theaters with proscenium openings between the auditorium and the stage house, churches subdivided into main and side naves and concert halls with reverberation chambers. Diffuse field theory can be used to illuminate important properties of the reverberation conditions in coupled rooms.

We consider a sound source placed in a room, room 1, with volume V_1 and physical absorption area A_{10}. This room is coupled by an opening with area S to another room, room 2, with volume V_2 and absorption area A_{20}. From diffuse field theory (the energy balance equations) we find that the apparent absorption area of room 1, A_1, depends on the ratio between the absorption area of the attached room A_{20} and the area of the opening S:

$$A_1 = A_{10} + \frac{A_{20}S}{A_{20}+S}\,. \tag{9.24}$$

It is seen that for small openings, $S \ll A_{20}$, $A_1 \approx A_{10} + S$, and for large openings, $S \gg A_{20}$, $A_1 \approx A_{10} + A_{20}$, as we would expect.

As for the reverberation curve, a double slope may be observed, particularly if both source and receiver are placed in the less reverberant of the two rooms. The two reverberation times T_I and T_II defining the double slope can be described in terms of the reverberation times for each of the two rooms separately if we consider the opening to be totally absorbing:

$$T_1 = \frac{0.161V_1}{A_{10}+S} \quad \text{and} \quad T_2 = \frac{0.161V_2}{A_{20}+S}\,. \tag{9.25}$$

However, for our purpose the math gets simpler if we apply damping coefficients: $\delta_1 = 6.9/T_1$ and $\delta_2 = 6.9/T_2$ instead of the reverberation times. Then, with coupling, the two damping coefficients, δ_I and δ_II, corresponding to the slope in the first and second part of the decay,

Amazon EX1014 Page 332

respectively, can be calculated from:

$$\delta_{\mathrm{I,II}} = \frac{\delta_1 + \delta_2}{2} \pm \sqrt{\frac{(\delta_1 - \delta_2)^2}{4} + \frac{S^2 \delta_1 \delta_2}{(A_{10} + S)\,(A_{20} + S)}} \tag{9.26}$$

upon which

$$T_{\mathrm{I}} = 6.9/\delta_{\mathrm{I}} \quad \text{and} \quad T_{\mathrm{II}} = 6.9/\delta_{\mathrm{II}}\,. \tag{9.27}$$

Another important feature of the double-slope decay is how far down the knee point appears. If the opening area is large and/or the source or receiver is placed close to the opening, then a substantial exchange of energy can occur and the knee point may appear so early that even the EDT can be influenced by the decay from the coupled room with longer T. If the opening area is small, then the knee point appears late and far down on the curve, in which case the influence of the coupled room can only be perceived after final cords. This is the case with some of the existing concert halls equipped with coupled reverberation chambers. Further descriptions of coupled spaces can be found in [9.5].

9.5.3 Absorption Data for Seats and Audiences

Regardless of which reverberation formula is used, the availability of reliable absorption values is fundamental for the accurate prediction of T as well as of most other room acoustic parameters. Beyond the absorption values of various building materials (as listed in Table 11.1), realistic figures for the absorption of the chairs and seated audience are of special concern, because these elements normally account for most of the absorption in auditoria. The absorption of chairs – both with and without seated people – depends strongly on the degree of upholstery. Average values of absorption for three different degrees of chair upholstery have been published by *Beranek* [9.22] and quoted in Table 9.3. As mentioned earlier, chair absorption values for use specifically in Sabine calculations and for different types of hall geometries can be found in [9.21].

Sometimes audience absorption data are quoted as absorption area per chair or per person. However, as demonstrated by *Beranek* [9.3], the use of absorption coefficients multiplied by the area covered by the seats plus an exposed perimeter correction is more representative than absorption area per person, because audience absorption tends to vary proportionally with the area covered, while the absorption coefficient does not change much if the density of chairs is changed within that area.

Minimizing the total sound absorption area is often attempted in auditoria for classical music in order to obtain a strong, reverberant sound field. One way to achieve this is not to make the row-to-row distance and the width of the chairs larger than absolutely necessary. Another factor is the seat design itself. Here one should aim at a compromise between minimum absorption and still a minimum difference between the absorption of the occupied and empty seat. This is facilitated by back rests not being higher than the shoulders of the seated person and only the surfaces covered by the person being upholstered. The rear side of the back rest should be made from a hard material like varnished wood. On the other hand, a minimum difference between the occupied and empty chair implies that the upholstery on the seat and back rest should be fairly thick (e.g., 80 mm on the

Table 9.3 Absorption coefficients of seating areas in concert halls for three different degrees of upholstery; both empty and occupied (after [9.22])

Octave centre frequency (Hz)	125	250	500	1000	2000	4000
Heavily upholstered chairs Unoccupied	0.70	0.76	0.81	0.84	0.84	0.81
Heavily upholstered chairs Occupied	0.72	0.80	0.86	0.89	0.90	0.90
Medium upholstered chairs Unoccupied	0.54	0.62	0.68	0.70	0.68	0.66
Medium upholstered chairs Occupied	0.62	0.72	0.80	0.83	0.84	0.85
Lightly upholstered chairs Unoccupied	0.36	0.47	0.57	0.62	0.62	0.60
Lightly upholstered chairs Occupied	0.51	0.64	0.75	0.80	0.82	0.83

Amazon EX1014 Page 333

seat and 50 mm on the back rest), the bottom of the tip-up seats should be perforated or otherwise absorptive, and when the seat is in its vertical position, there should be a wide opening between the seat and the back rest so that the sound can still access the upholstered areas.

9.5.4 Prediction by Computer Simulations

Computer simulations take into account the geometry and actual distribution of the absorption materials in a room as well as the actual source and receiver positions. The results provided are values for all relevant acoustic parameters as well as the generation of audio samples for subjective judgments. Computer simulation is useful not only for the design of concert halls and theaters, but also for large spaces such as factories, open-plan offices, atria, traffic terminals, in which both reverberation, intelligibility and noise-mapping predictions are of interest. Besides, they can be used in *archaeological acoustics* for virtual reconstruction of the acoustics of ancient buildings such as Roman theaters (http://server.oersted.dtu.dk/www/oldat/erato/).

The first computer models appeared in the 1960s, and today a number of commercially available software packages are in regular use by acoustic consultants as well as by researchers all over the world.

In computer simulations the room geometry is represented by a three-dimensional (3-D) computer-aided design (CAD) model. Thus, the geometry information can often be imported from a 3-D model file already cre-

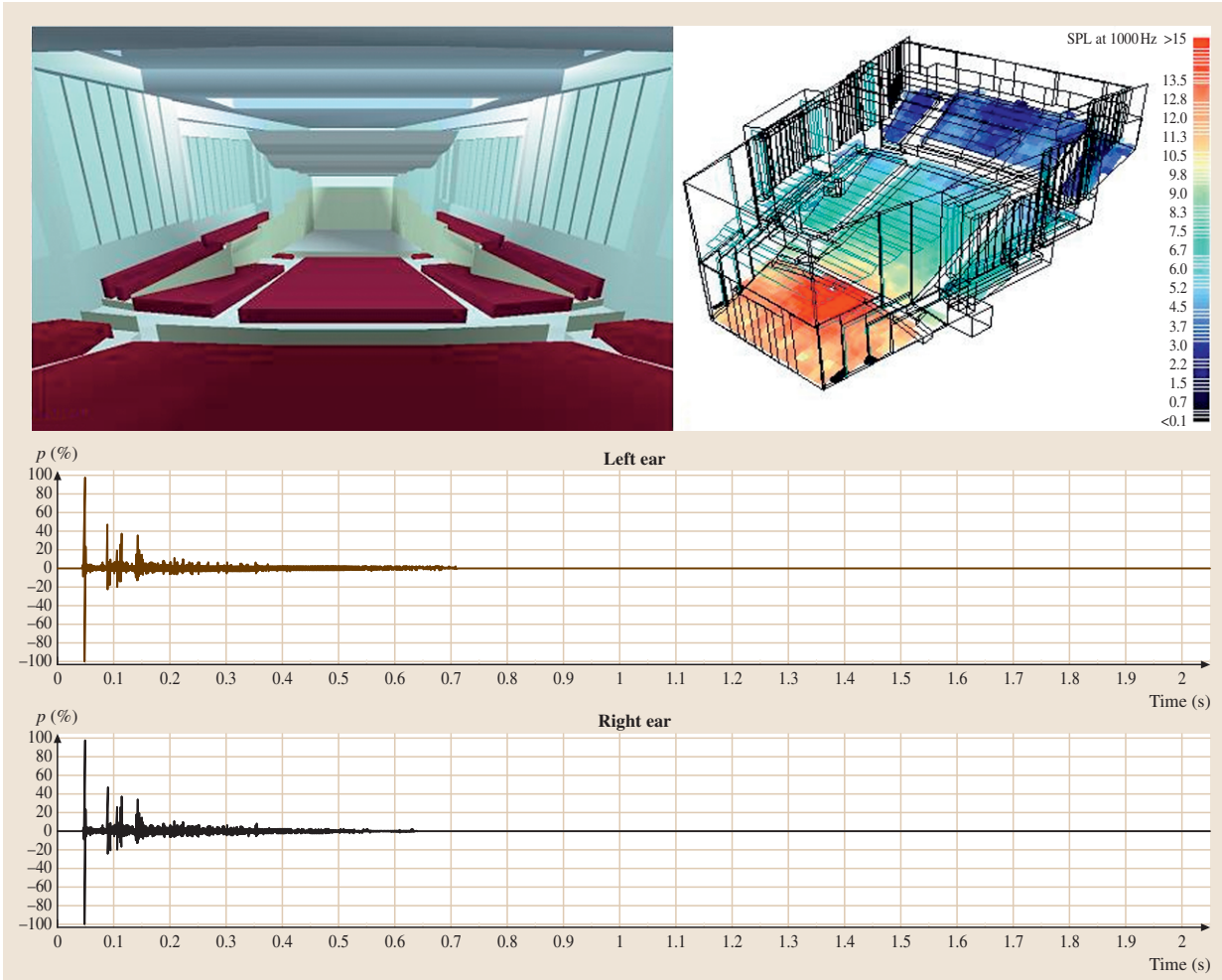


Fig. 9.14 Examples of output from a room acoustic simulation software package (ODEON); see text for explanation

Amazon EX1014 Page 334

ated by the architects. However, one needs to break up eventual curved surfaces into plane elements and often it is also advisable to simplify details in the geometry provided by the architect. Alternatively, the acoustician can build the room model by writing a file of coordinate points or mathematical expressions for all surface corners. Of course this will be more time consuming in cases of complex room geometry, but some programs offer intelligent programming routines for the modeling of geometric elements.

When the geometry has been completed, the absorption values for each octave band, the scatter, and the eventual degree of acoustic transparency have to be assigned to each surface. In addition, the source and receiver characteristics (power and directivity versus frequency, position and orientation) must also be entered.

In the calculation of sound propagation, most models disregard phase properties and use an energy approximation. The sound propagation is studied by means of rays (up to millions depending on the room complexity) being traced from the source position through reflections until their energy has been reduced below the dynamic range of interest (often determined by T or the signal-to-noise ratio in auralizations). Reflection directions are chosen so that angle of reflection equals angle of incidence or with more or less random angles depending on the scatter value attributed to the surface. For the calculation of early reflections, some models also apply image source theory, which actually makes phase considerations possible if the complex impedances of the various surface materials and not just their energy absorption coefficients are known. However, the late reverberation part is normally calculated by some kind of ray method, as image source calculations become impractical after a few orders of reflection. Recently, however, attempts to carry out complete calculations inspired by finite-element (FEM) or boundary-element (BEM) methods have been reported even for rather complex rooms [9.23], whereby phase and diffraction phenomena can be handled.

As a result of tracing the history of all rays emitted in the room model, source and receiver specific impulse responses are produced with detailed information about the direction of incidence, delay and level versus frequency of each reflection component. From these, all the parameters mentioned in Sect. 9.3 can be calculated. Most programs can generate a grid of receiver point over selected audience areas and code the numeric results using a color scale to facilitate a fast overview of the detailed position results for any of the parameters. Also, automatic plots of relevant statistics regarding the distribution with frequency and position are provided.

Impulse responses suitable for use in *auralization* can be generated by applying available head-related transfer functions, which have been recorded as a function of direction of sound incidence on an artificial head. Such transfer functions not only represent the directivity of the human receiver, but are also important for correct perception of direction of incidence of sound field components when listening through headphones. Alternatively, the directional information of the individual reflections making up the impulse response can be coded into a surround-sound format for listening via a multiple-loudspeaker setup.

Figure 9.14 shows a view of a room model as well as two binaural impulse responses and a color mapping of a calculated parameter (generated by the ODEON software).

9.5.5 Scale Model Predictions

Acoustic scale models have been used since the 1930s [9.24] to study the behavior of sound in rooms. With proper care in detailing the geometry and the choice of materials for building the model, this technique can provide quite accurate predictions of impulse responses and acoustic parameters. Actually, scale modeling is still regarded as more reliable than computer models in cases where the room contains a lot of irregular surfaces or objects that diffuse or diffract sound waves. However, the use of physical scale models is limited to larger, acoustically demanding projects, as building and testing the model is far more time consuming and costly than computer modeling.

The diffusion and diffraction caused by an object or surface in a room depends on the ratio between the linear dimensions and the wavelength of the sound: l/λ. Therefore, in a $1:M$ scale model, the diffraction/diffusion conditions will be modeled correctly if the wavelengths of the measurement signals in the model are chosen such that:

$$\frac{l}{\lambda} = \frac{l_{\mathrm{M}}}{\lambda_{\mathrm{M}}} = \frac{l/M}{\lambda/M} \quad \Rightarrow \quad f_{\mathrm{M}} = fM\,. \tag{9.28}$$

This means that we should increase the frequency used in the model measurements by the factor M. Fortunately, sound sources and microphones exist that are capable of handling frequencies up to around 100 kHz, whereby the frequency range up to 10 kHz can be handled in a 1 : 10 scale model or 2 kHz in a 1 : 50 scale model. The sources may be small loudspeaker units (piezoelectric or

electro magnetic) and the microphones 1/4" or even 1/8" condenser types.

Ideally, the absorption coefficient α_M of the materials chosen for building each surface in the hall model should fulfil the equation:

$$\alpha_M(fM) = \alpha(f) \,. \tag{9.29}$$

In practice, however, we need only to distinguish between mainly reflecting and mainly absorptive materials. The reflecting surfaces are quite easy, as these can be made of any hard material (plywood, plaster) with a couple of layers of varnish if necessary. It is far more important to build the seats/audience so that both head diffraction and absorption is correct for the scale chosen. In the end it is often necessary to fine-tune the reverberation time in the model by applying absorption to some secondary surfaces (surfaces that do not generate primary early reflections).

Another important issue is to compensate for the excessive air absorption at high frequencies in the model. This can be achieved either by drying the air in the model, by exchanging the air for nitrogen, or simply as a calculated compensation since the attenuation is a simple linear function of time once the, frequency, humidity and temperature have been specified. However, in the case of compensation the signal-to-noise ratio at high frequencies will be severely limited, which reduces the decay range for the estimation of T and the sound quality if auralization is attempted. Figures 9.15 and 9.16 show an example of a 1:10 scale model and its audience respectively.

9.5.6 Prediction from Empirical Data

In the 1980s, when most of the current objective acoustic parameters had been defined, several researchers started making systematic measurement surveys of concert, opera and multipurpose halls. Many of these measured data were published in [9.3]. Combining acoustic data with data on the geometry of the halls has made it possible to derive some simple rules of thumb regarding how, and how much, acoustic parameters are affected by the choice of gross room shape and dimensions.

Through statistical multiple linear-regression analysis of data from more than 50 halls, a number of relationships between acoustical and geometrical properties have been found [9.25]. A number of the linear regression models derived are shown in Table 9.4. In these models, the variations in the acoustic parameters are described as functions of the expected value according to diffuse field theory (G_{exp} and C_{exp}, Sect. 9.3) plus various geometrical variables such as: average room height H, average width W, room volume V, number of seats N, number of rear balcony levels 'no. rear balc.' etc. When the models were derived, the measured reverberation time (T) was used to calculate the expected values of strength G_{exp} and clarity C_{exp}. When using these models for prediction of the values in a hall not yet built, the expected values could be based on a Sabine calculation of T instead. It should be mentioned that, apart from $\Delta G(10\,\text{m})$, the attenuation in strength per 10 m increase in source–receiver distance, the predicted values should be considered as representing the position average of values to be found in the hall.

The three rightmost columns in Table 9.4 contain information about how well each prediction formula matched the measured data. When comparing the amounts of variance explained by the different models, it is seen that the diffuse field prediction is responsible for the largest portion of the variance. This may be interpreted as hall volume and total absorption area being the most important factors governing the behavior

Table 9.4 Regression models for the relationships between room acoustic parameters and room design variables as derived from a database containing data from more than 50 halls (after [9.25])

Room acoustic parameter	Regression models: $f(\text{PAR}_{expected}, \text{geometry})$	Correlation coefficient	% of variance	STD residuals
C (dB)	$-0.1+1.0C_{exp}$	0.76	58	1.0 dB
	$-1.4+0.95C_{exp}+0.47W/H+0.031$ floor slope	0.83	68	0.9 dB
	$-1.2+1.03C_{exp}+0.43W/H+0.013$ wall angle	0.84	70	0.9 dB
	$-1.77+1.10C_{exp}+0.055W+0.027$ stage ceil. angle	0.86	74	0.8 dB
G (dB)	$-2.0+0.94G_{exp}$	0.91	83	0.9 dB
	$-5.61+1.06G_{exp}+0.17V/N+0.04$ distance	0.94	89	0.9 dB
$\Delta G(10\,\text{m})$ (dB)	$-1.85+0.42$ no. rear balc.	0.50	25	0.7 dB
	$-1.41+0.35$ no. rear balc. -3.93 distance$/(HW)$	0.55	31	0.6 dB
LEF (–)	$0.39-0.0061$ width	0.70	49	0.05
	$0.37-0.0051$ width -0.00069 wall angle	0.72	53	0.05

of C and G. However, as all the independent variables listed gave a significant improvement of the model accuracy, this also demonstrated that consideration of the geometrical factors can improve the prediction accuracy. Equally importantly, the acoustic effects of changes in certain design variables can be quantified quickly.

Prediction of Clarity

The first C-model illustrates that absorption and volume, as reflected in T, are responsible for the main part of the variation in C. The other three models all illustrate the positive, but sometimes unwanted, effect of average hall width on clarity. Moreover, it is seen that introducing a moderate 15° slope of the main audience floor (without changing the other variables: average width-to-height ratio, volume and absorption area) causes C to increase by about 0.5 dB on average. Similarly, changing the basic design from rectangular to a 70° fan shape makes C increase by about 1 dB. The last C-model illustrates the effect of tilting the angle of the stage ceiling towards the audience. Changing the slope from horizontal to 20° results in about 0.5 dB higher C values.

Predictions of Strength

The G model only considering G_{exp} is rather accurate. The constants in both G-models illustrate the fact that G is always a few dB lower than G_{exp}, as also predicted by *Barron's revised theory* [9.26].

The second G-model contains an independent variable: the ratio between the volume and the number of seats. The positive influence of this ratio is not immediately evident because V/N, which is strongly correlated with T, is expected to be incorporated into the variable G_{exp} already (although this result has also been confirmed by other researchers [9.27]).

As mentioned earlier, G shows a steady and significant decrease with distance in most halls, as was also found by *Barron* [9.26]. This phenomenon is described quantitatively by the two models for estimation of the rate of decrease in G per 10 m source–receiver distance $\Delta G(10\,\mathrm{m})$.

Both listed $\Delta G(10\,\mathrm{m})$ models indicate a reduced distance attenuation when the number of rear balconies is increased. This may be related to the fact that the level is increased in the more-distant seats when these are placed on a balcony, whereby they are closer to the stage and to the reflecting ceiling.

In the second $\Delta G(10\,\mathrm{m})$ model, distance/HW appears as an independent variable. This variable equals the distance from the stage to the rearmost seat divided by the product of the average room height H and average hall width W. As the coefficient to this variable is negative, the natural result appears to be that attenuation with distance will increase if the hall is long, narrow or has a low ceiling.

Predictions of Lateral Energy Fraction

At the bottom of Table 9.4 are listed two models for LEF as a function of hall geometry only, as diffuse field theory has no effect on LEF variation. The effects of the width and angle between the side walls are understandable.

Simple prediction formulae as listed in Table 9.4 are particularly useful in the very early phases of the design process, in which it is natural for the architect to produce and test many different sketches in short succession, leaving no time to carry out computer simulation of each proposal. The importance of the knowledge embedded in these rule-of-thumb equations is highlighted by the fact that many aspects of the acoustics of a new hall are settled when one of these sketches is selected for the further development of the project.

9.6 Geometric Design Considerations

With the connections between room acoustic parameters and hall geometry described in the previous section, we have made the first approach to the third question set up in the introduction to this chapter, which is of real interest to the acoustic designer: how do we control the acoustic parameters by the architectural design variables? Referring again to Fig. 9.1, we are beginning to fill the lower-left box with architectural parameters and establish the relationship between these and the objective parameters. The major geometric factors of importance in auditorium design will be dealt with in the present section: seating layout in plan (determining the gross plan shape of the room) and in section (determining sight lines), use of balconies, choice of wall structure, room height, ceiling shape and use of free-hanging reflectors.

9.6.1 General Room Shape and Seating Layout

When people stop to watch or listen to a spontaneous performance, for instance in an open square in the city, the way they arrange themselves depends on the type



Fig. 9.15 A 1:10 scale model of the Danish Broadcasting concert hall in Ørestad, Copenhagen. The model is fitted with audience and orchestra (acoustic design by Nagata Acoustics, Japan)

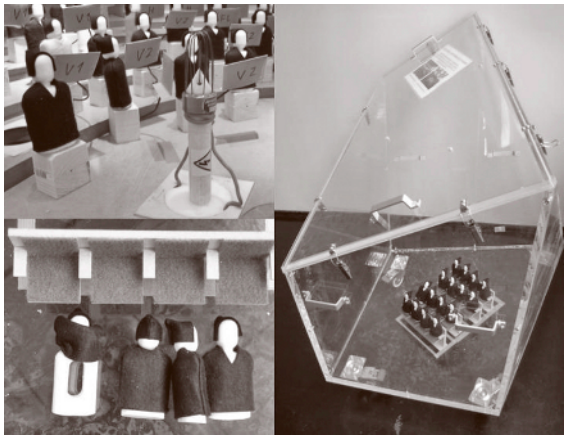


Fig. 9.16 Model audience used in a 1:10 scale model of the Danish Broadcasting concert hall shown in Fig. 9.15. *Upper-left corner*: Close view of model orchestra musicians and spark sound source. *Lower left*: close-up view of model chairs and polystyrene audience with hollow chest. Dress and hair is made from wool felt. *Right*: model audience and chairs placed in model reverberation chamber for absorption testing (reverberation room designed by Nagata Acoustics)

of performance. It will be governed by the fact that any newcomer will look for the best position available relative to his/her need to see or hear properly. The choice will be a compromise between choosing a position close to the performer(s) and next to other members of the audience or farther away but close to an eventual main center line for vision and sound radiation. In Fig. 9.17,

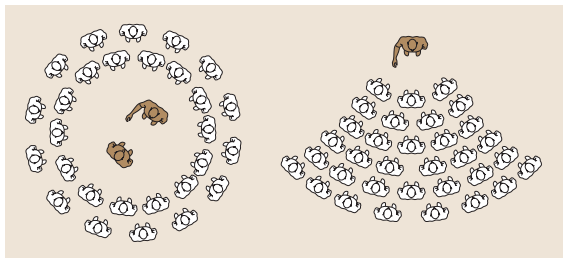


Fig. 9.17 *Organic* formations of audience depending on the type of performance

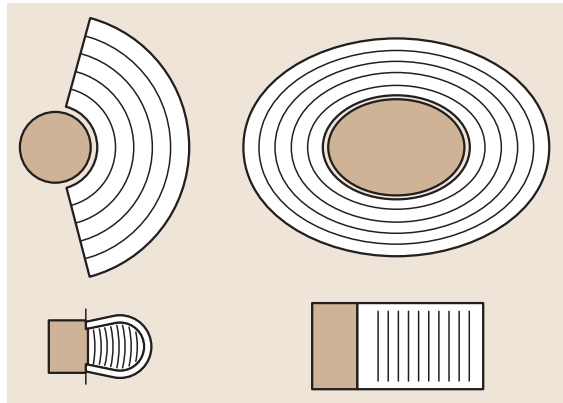


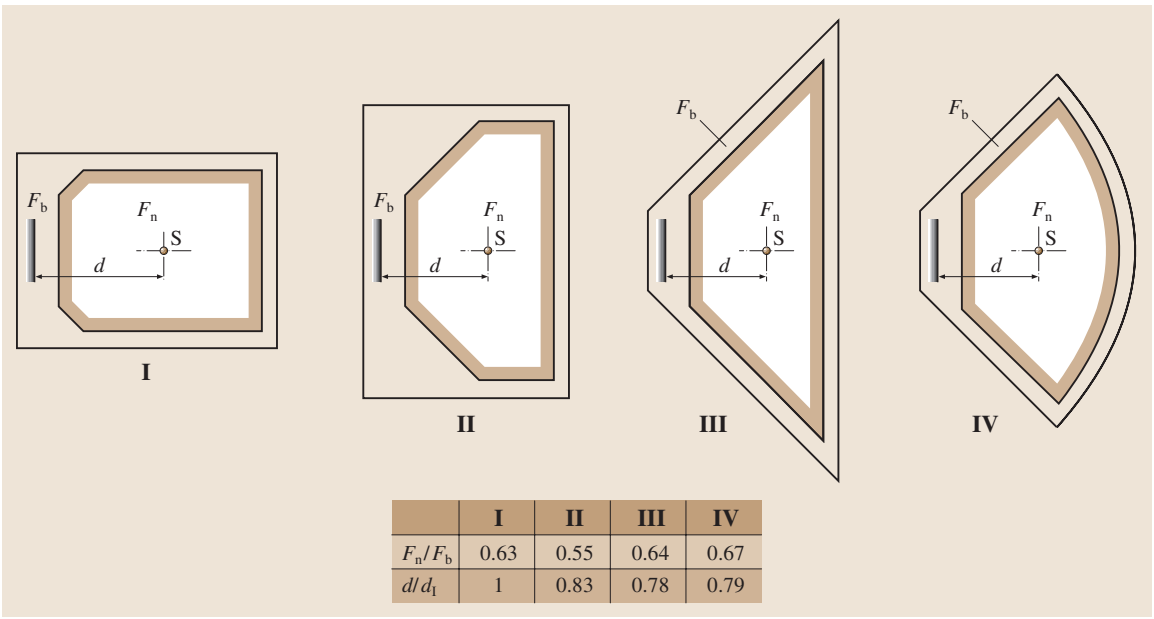
Fig. 9.18 Basic room shapes (after [9.4])

such organic audience arrangements are shown for two different types of performances:

1. action or dialogue theater such as dancing, fighting, debating, circus performance, for which the visual and acoustic emission are more or less omnidirectional and
2. monologue or proscenium stage theater performance and concerts with acoustic instruments – all of which have a more-limited visual and/or acoustic directivity.

When we set up walls around the gathered people, we arrive at two classical plan shapes of auditoria developed early during our cultural history. These are shown at the top of Fig. 9.18: the fan-shaped Greek/Roman theater (left), and the amphitheater, circus or arena (right). Both were originally open theaters; but the shapes were maintained in roofed buildings. In the bottom of Fig. 9.18, later, basic forms are shown: the horseshoe, the Italian opera plan and the rectangular concert hall. The latter form was originally a result of traditional building forms and limitations in roof span with wooden beams.

Amazon EX1014 Page 338



	I	II	III	IV
F_n/F_b	0.63	0.55	0.64	0.67
d/d_I	1	0.83	0.78	0.79

Fig. 9.19 Mean stage–listener distance and net efficiency of floor area for four different room shapes. F_n is the net area occupied by the audience, F_b is the total floor area, d is the average source–listener distance; $d_I = d$ for case I (after [9.28])

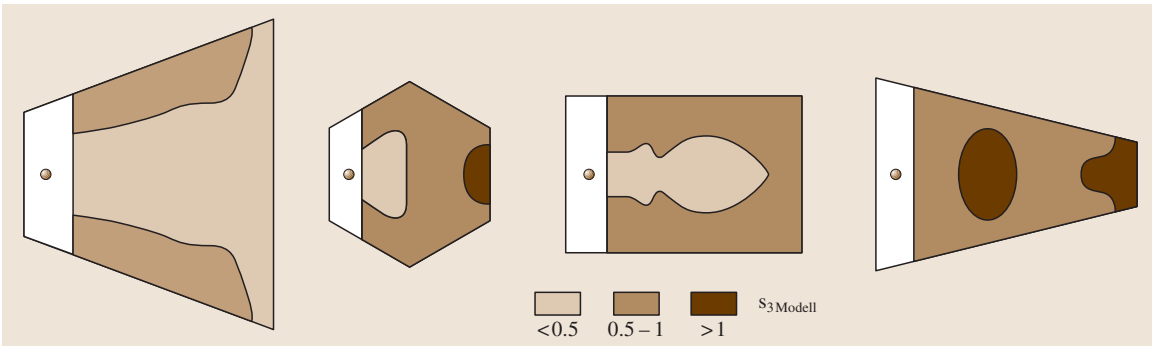


Fig. 9.20 Calculated distributions of early lateral reflection levels in rooms of different plan shape: fan, hexagon, rectangular and reverse fan. *Darker areas* represent higher LEF levels (after [9.5])

When building an auditorium, efficient use of the available floor space and the average proximity of the audience to the performers are very important parameters. Depending on the plan shape of the room different values appear as shown in Fig. 9.19.

Low average distance is the main reason for the frequent use of the fan shape (III and IV in Fig. 9.19), although the directivity of sound sources (like the human voice) as well as the quality of lines of sight sets limits on the opening angle of the fan.

Another limitation of the fan shape is that it does not generate the strong lateral reflections so important in halls for classical music. This was already seen in the empirical equations in Sect. 9.5.6 and confirmed in Fig. 9.20, showing the calculated distributions of early lateral reflection levels in rooms of different plan shapes.

The reason is that sound waves from the source hitting the side walls will produce reflections that will run almost parallel to the direct sound and so hit the listener from an almost frontal direction, therefore not contribut-

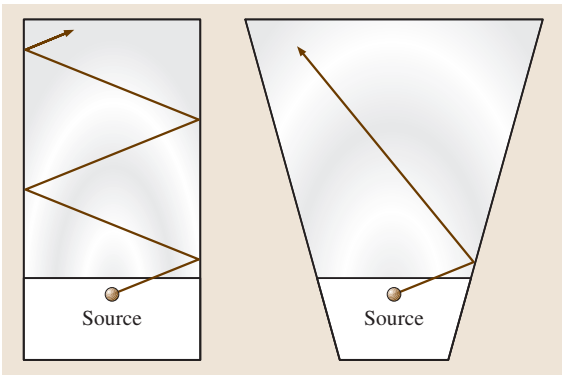


Fig. 9.21 Directions of side-wall reflections depending on room shape (after [9.4])

ing to the LEF or causing dissimilarity between the two ear signals. This is also illustrated in Fig. 9.21. In addition, many of the side-wall reflections will not reach the center of the hall at all.

9.6.2 Seating Arrangement in Section

When several rows of seated audience are placed on a flat floor, the direct sound from a source on the stage and the reflection off the audience surface will hit a receiver in this audience area after having traveled almost identical distances, and the reflection will hit the audience surface at a very small angle relative to the surface plane. At this grazing angle of incidence, all the energy will be reflected regardless of the diffuse absorption value of the surface, but the phase of the pressure in the reflected wave will also be shifted 180°. Hereby the direct sound and the grazing incidence reflection will almost cancel each other over a wide frequency range. Typical attenuation values found at seats beyond the tenth row can amount to 10–20 dB in the range 100–800 Hz relative to an unobstructed direct sound. Moreover, the higher frequencies above 1000 Hz will also be attenuated by 10 dB or more due to scattering of the sound by the heads of people sitting in the rows in front.

The same values of attenuation mentioned above for the direct sound are likely also to be valid for first-order reflections off vertical side walls. The result will be weaker direct sound and early reflections in the horizontal stalls area and subjective impressions of reduced *clarity*, *intimacy* and *warmth* in the stalls seats.

To avoid this grazing incidence attenuation, the seating needs to be sloped relative to the direction towards the source. The vertical angle of the audience floor can be designed to be constant or to vary gradually with

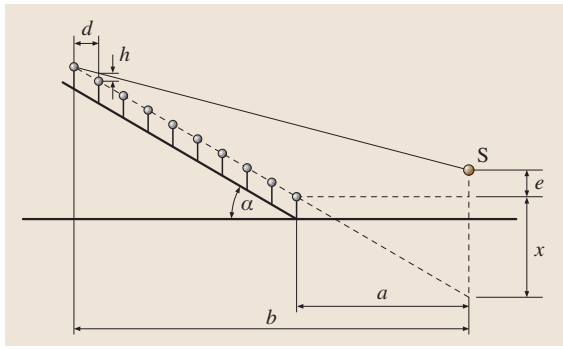


Fig. 9.22 Relevant parameters for the calculation of the necessary angle α for a constantly sloping floor section. a: distance from the source to the first seating row, b: distance from the source to the last seating row, d: distance between the seat rows, e: source height above the head of listeners in the first row (determined by the height of the stage floor over the stalls floor), h: line-of-sight clearance at the last row. Rows closer to the source will have clearance values larger than h (after [9.28])

distance. If a constant slope is planned, the angle α necessary to obtain a certain clearance h at the last row (equal to the vertical distance between sight lines from this and the preceding row) can be calculated from the variables indicated in Fig. 9.22:

$$\tan\alpha = \frac{hb}{da} - \frac{e}{a} \,. \tag{9.30}$$

In particular, it is of interest to use this formula to see how far away from the source it is possible to maintain a certain slope (one obvious example being $\alpha = 0$, corresponding to a horizontal floor):

$$b = \frac{d}{h}\,(e + a\tan\alpha) \,. \tag{9.31}$$

Normally, a clearance value of minimum 8 cm will be sufficient to avoid grazing-incidence attenuation; but sight lines will still be unsatisfactory unless staggered seating is applied so that each listener is looking towards the stage between two other heads and not over a person sitting directly in front. However, if the clearance is increased to about 12 cm, then both the acoustic and visual conditions will be satisfactory, but this implies a steeper floor slope.

In rows closer to the sound source a linear slope will cause the clearance to be higher than the design goal – and so higher than necessary. Often this is not desirable as it may cause the last row to be elevated high above the source and the steeply sloped floor will reduce

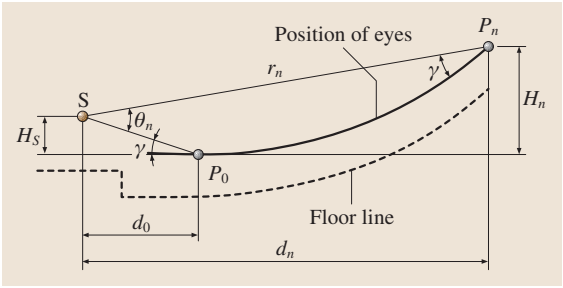


Fig. 9.23 Relevant parameters for the calculation of a floor slope with constant clearance. The variables are explained in the text (after [9.29])

the effective volume of the room. Therefore, it can be relevant to consider instead a curved floor slope with constant clearance, as illustrated in Fig. 9.23.

In Fig. 9.23, the height of a person's head in the n-th row H_n relative to the height of the head in the first row is calculated according to the formula (derived from a logarithmic spiral)

$$H_n = d_0\gamma + d\,(\theta - \gamma) = \gamma\left[d_n \log_e\left(\frac{d_n}{d_0}\right) - (d_n - d_0)\right] . \qquad (9.32)$$

In this expression d_0 is the distance from the source to the first row, d_n the distance to the n-th row and γ the desired angle between the tangent to the seating plane and the line of sight to the source. For a one meter row-to-row distance, an angle γ of 8° will correspond to a clearance of about 12 cm.

A curved slope can also be obtained by simply applying the linear-slope formula to each of smaller sections of seat rows successively (e.g. for every five rows). In this case the variables a and e in (9.30) should refer to the first row in the relevant section.

As mentioned earlier, a steep slope will reduce the volume and so the reverberation time T. Besides, geometric factors may reduce T beyond what a Sabine calculation may predict. The reason is that the elevated seating – and its mirror image in the ceiling – will cover a larger solid angle as seen from the source. Hereby a larger part of the emitted sound energy will quickly reach the absorbing seating and leave less for fueling a reverberant sound field in the room. The result is higher *clarity* and less *reverberance* in halls with steeply raked seating. This is in line with the empirical equation for clarity as a function of T (C_{exp}) and floor slope listed in Table 9.4, Sect. 9.5.6. Consequently, it is a good idea to limit the clearance/angle in halls for classical music to about 8 cm or 6°, whereas for speech auditoriums and drama theaters, in which both intelligibility and good visual sight lines have priority, values closer to 12–15 cm or 8°–10° may be preferable.

9.6.3 Balcony Design

The introduction of balconies allows a hall with a larger seat count to be built where a limited footprint area is available; but often balconies are introduced simply in order to avoid longer distances between stage and listeners. The average stage–listener distance can be reduced significantly by elevating the rearmost seat rows and moving them forward on rear or side balconies above parts of the audience in the stalls. In this way not only is the direct sound increased by shortening the direct sound path, but the elevated seats also become closer to a possibly reflecting ceiling. The result is higher strength as well as higher clarity in the balcony seats compared to the alternative placement of seats at the back of an even deeper hall.

Since the line-of-sight and clearance criteria still have to be fulfilled for balcony seats, the slope often becomes quite steep on balconies, as illustrated in Fig. 9.24 showing a long section in a hall with constant clearance on main floor and balcony seats. For safety reasons the slope must be limited to no more than about 35°.

When balconies are introduced, the acoustics in the seats below the balconies need special attention. It is important to ensure sufficient opening height below balcony fronts relative to the depth of the seating under the balcony (Fig. 9.25). Otherwise, particularly the level of the reverberant energy in the overhung seats will be reduced, causing the overall sound level to be weak and lacking in fullness. Rules of thumb in design are $H \geq 2D$ for theaters (in which reverberant sound is less important) and $H \geq D$ for concert halls (in which fullness is

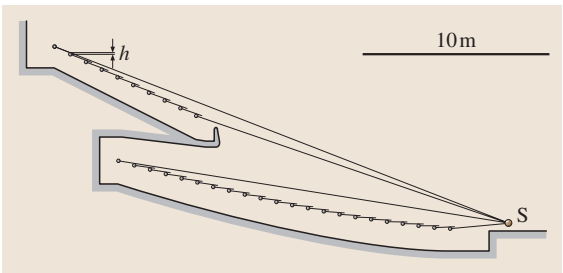


Fig. 9.24 Section of a hall with a rear balcony and constant clearance. The result is an increased slope for the elevated balcony (after [9.30])

Amazon EX1014 Page 341

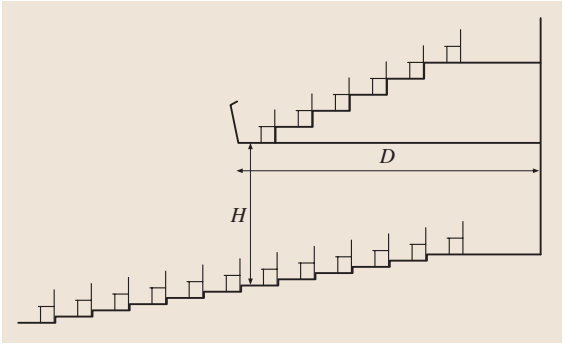


Fig. 9.25 Important parameters for maintaining proper sound in seats overhung by balconies (after [9.9])

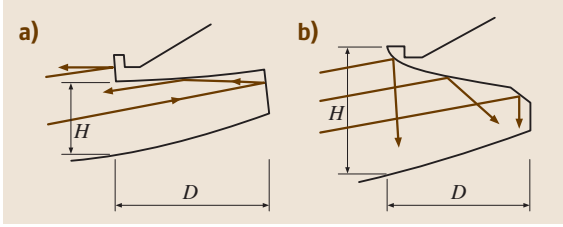


Fig. 9.26a,b Poor (**a**) and good (**b**) design of balcony profile (after [9.29])

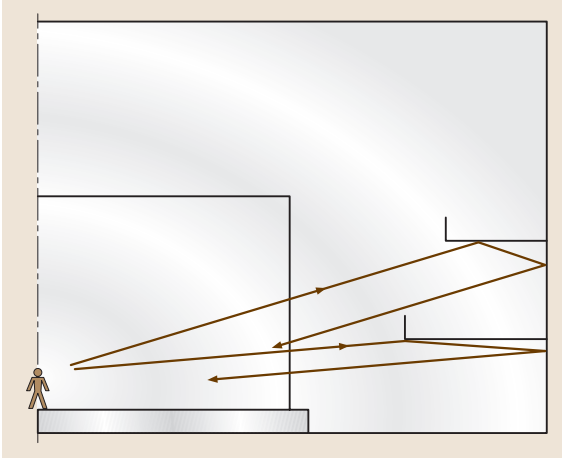


Fig. 9.27 A half cross section of a hall with side balconies that, in combination with the side wall, return early reflection energy towards the stalls area from a lateral angle (after [9.4])

more important) [9.9]. The criteria may also be stated in terms of the vertical viewing angle from the seat to the main hall volume. Recently, *Beranek* [9.3] has suggested this be a minimum of 45°.

Sometimes, the requirement for limited depth of balcony overhangs is also specified as a percentage of the ceiling area needed to be visible from the last row of seating or as a vertical opening angle β between the balcony soffit and the head of listener below the balcony front as seen from the last row. In the latter case, Barron has suggested 25° as suitable for drama halls and 35° for concert halls.

As illustrated in Fig. 9.26, it is often advantageous to let the soffit be sloped, so that it will help distribute reflected sound to the overhung seats. If a diffusing profile is added to the balcony soffit, the reflection off this surface may even gain a lateral component, so that *spaciousness* will not be reduced. The drawing also illustrates how the otherwise vertical balcony front and the right angle between the rear wall and balcony soffit have been modified to avoid sound being reflected back to the stage where it could generate an echo. However, in a rectangular hall with balconies along the side walls, vertical soffits in combination with the side-wall surface can increase the early and lateral reflection energy in the stalls area as shown in the half cross section in Fig. 9.27.

In general, as large-scale modulations in the hall geometry, balconies also provide low-frequency diffusion, which is considered an advantage.

9.6.4 Volume and Ceiling Height

As the people present are often the main source of sound absorption in an auditorium, the ratio between the volume and area covered by the audience and performers is an important factor for the reverberation time achievable. The general relationship valid for concert halls is shown in Fig. 9.28.

Because the variation in area per person does not vary significantly between halls, the volume per seat is also used as a rule of thumb for the calculation of the required volume. For speech theaters, a volume per seat of 5–7 m^3 will result in a reverberation time around 1 s, whereas at least 10 m^3 per seat is needed to obtain a T value of 2 s.

Since the audience normally occupies most of the floor area, the volume criterion can also be translated to a ceiling height criterion. In such cases, the abscissa in Fig. 9.28 can be interpreted as roughly equal to the required room height. A ceiling height of about 15 m is required if a T value of 2 s is the target, whereas a height of just 5–6 m will result in an auditorium with a reverberation time of about 1 s. It should be emphasized that, in the discussion above, we have assumed the other room surfaces to be reflective.

Amazon EX1014 Page 342

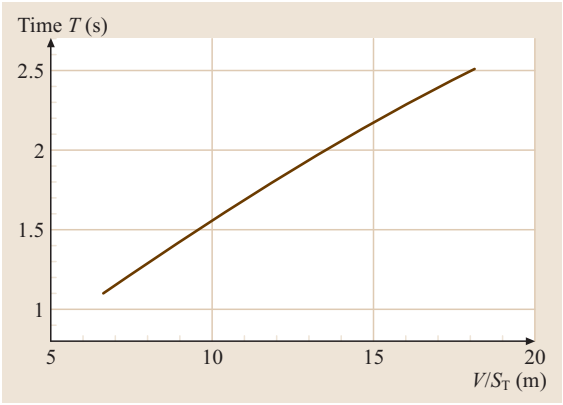


Fig. 9.28 Reverberation time T in occupied concert halls versus the ratio between the volume V and area occupied by the audience and orchestra S_T (including aisle space up to 1 m width). Derived from empirical data (after [9.9])

These guidelines on volume should be regarded as rough rules of thumb only, as room shape, floor slope, the presence of balconies and the distribution of absorption on other surfaces can cause significant variations in the objective values obtained for T and other parameters. In smaller halls (say below 1500 seats) for symphonic concerts, it is better to start with a volume slightly larger than required rather than the opposite, as one can always add a little absorption – but not more volume – to a hall near completion. On the other hand, in larger halls, where maintaining a suitably high G value is also of concern, a better strategy is to limit the volume and minimize the absorption instead. One way of minimizing absorption is to place some of the seats under balcony overhangs, where they are less exposed to the reverberant sound field. However, the other side of the coin is the poor conditions in these seats, as explained in Sect. 9.6.3.

9.6.5 Main Dimensions and Risks of Echoes

In order for early reflections to contribute to the *intelligibility* of speech or to the *clarity* of music, they must not be delayed more than about 30 ms and 80 ms, respectively, relative to the direct sound. If the first reflection arrives later than 50 ms, it is likely to be perceived as a disturbing echo when impulsive sounds are emitted (Sect. 9.2).

In large rooms it is a challenge to shape the surfaces so that reflections from the main reflecting areas arrive within 50 ms at all seats. It is possible to identify which surfaces are able to generate echoes at the receiver point

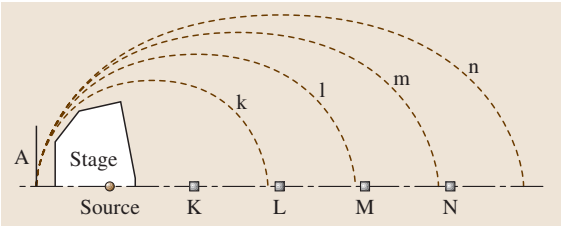


Fig. 9.29 Echo ellipses drawn over the plan of an auditorium (after [9.4])

in question by means of simple geometrical studies of plan and sectional drawings as seen in Fig. 9.29. Ellipses are drawn so that the source and relevant receiver positions are placed at the focal points and so that the sum of distances from the focal points to any point on the ellipse equals the distance between the focal points plus 17 m (times the scale of the drawing). Then, if a surface outside the ellipse marked "m" in the figure directs sound towards seats near point M, this surface must be made absorbing, diffusing or be reoriented so that the reflection is directed towards areas farther away from the source than M.

In particular, it is important to check the risk of echoes being generated by the rear wall behind the audience and from a high ceiling.

9.6.6 Room Shape Details Causing Risks of Focusing and Flutter

Concave surfaces can cause problems as they may focus the sound in certain areas while leaving others with too little sound. Thus, vaulted ceilings as seen in Fig. 9.30 are only acceptable if the radius of curvature is less than half the height of the room (or rather half the distance from peoples' heads to the ceiling) so that the focus center is placed high above the listeners. Situations with

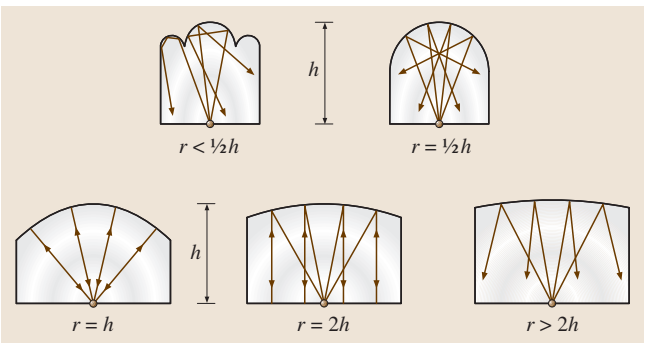


Fig. 9.30 Focusing by concave ceilings (after [9.31])

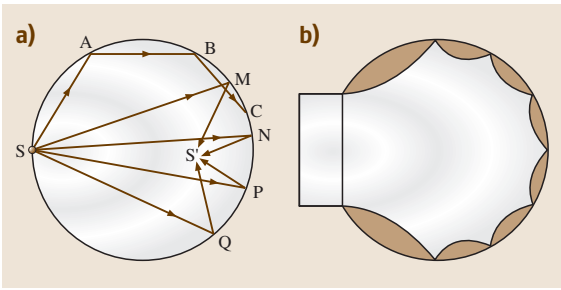


Fig. 9.31a,b Circular room with focused and creeping waves **(a)**, and with the surface shape modified **(b)** so that these phenomena are avoided

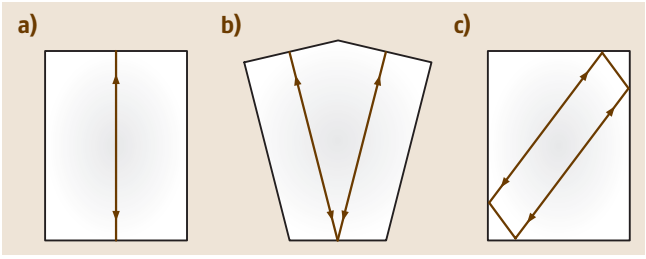


Fig. 9.32 Room shapes causing risk of flutter echoes

a slight curvature, $r \geq 2h$, may also be acceptable if the floor is covered with absorptive seating so that multiple reflections between floor and ceiling will not arise.

Circular rooms are also unfavorable both due to the risk of focusing and the risk of a whispering-gallery effect (boundary waves creeping along the wall). Modifications of the concave wall within the overall circular shape as shown in Fig. 9.31 can solve the problem, at least at mid and high frequencies.

Another frequent problem is regular, repeated reflections, so-called flutter echoes, which arise between parallel, smooth reflecting walls or along other simple reflection paths that can repeat themselves as shown in Fig. 9.32. Particularly in small rooms such reflections may cause strong coloration of the sound (through comb filtering as explained in Sect. 9.2.3) while in larger rooms they may be perceived as a long series of distinct echoes evenly spaced in time.

In any case, the flutter can be avoided by simple absorptive or diffusive treatment of at least one of the opposite reflecting surfaces, as shown in Fig. 9.33. Besides, if it is possible to change the angle between the opposite walls just by a few degrees (3–5°), the problem will also disappear.

9.6.7 Cultivating Early Reflections

In most rooms accommodating more than say 25 listeners, it is of relevance to consider how early reflections can help distribute the sound evenly from the source(s) to the audience. This will increase the early reflection energy and so improve *clarity/intelligibility* and perhaps even reduce the reverberation time by directing more sound energy towards the absorbing audience.

In rooms for speech, the ceiling height should be moderate (Sect. 9.6.4) so that this surface becomes the main surface for the distribution of early reflection energy to the audience. Figure 9.34 shows examples of how this can be accomplished in rooms with both low and high ceilings.

In this figure the two situations at the top need improvement. In the low-ceiling room to the left the echo from the rear wall/ceiling combination can be removed by introducing a slanted reflector covering the corner. In this way the sound is redirected to benefit the last rows, which may need this to compensate for the weak direct sound at this large distance. Alternatively, the energy from this rear corner may simply be attenuated by one or both surfaces near the corner being supplied with absorption.

If the ceiling is so high that echoes can be generated in the front part of the room, as shown to the right, the front part may likewise be treated with slanted reflectors that redirect the sound to the remote seat rows, or this ceiling area may simply be made sound absorbing. The lower-right section in the figure also illustrates how the

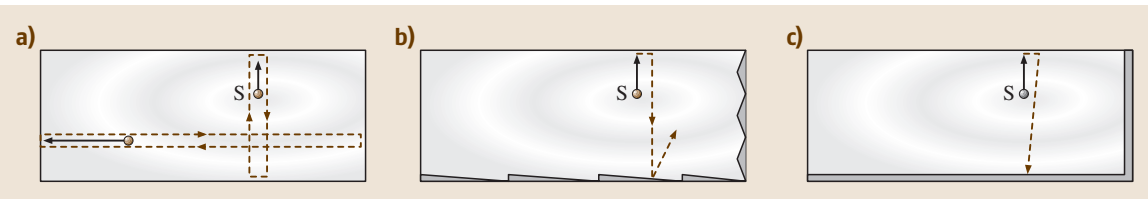


Fig. 9.33a–c Room with flutter echoes **(a)**, and measures against this by means of diffusion **(b)** or absorption **(c)** (after [9.28])

floor can be tilted to reduce volume and allow the rear seats benefit from being closer to the reflecting ceiling.

If for architectural reasons the ceiling is higher than necessary in a room for speech or rhythmic music, substantial areas of the room surfaces must be treated with sound-absorbing material to control the reverberation time. However, more sound absorption will reduce the overall sound level in the room so this is only recommended if the sound can be amplified.

If the ceiling is to be partly absorbing, the ceiling area that should be kept reflecting in order to distribute useful first-order reflections to all listeners can be generated geometrically by drawing the image of the source in the ceiling surface and connecting lines from this image point to the boundaries of the seating area. Where these lines cross the actual ceiling we find the boundaries for the useful reflecting area, as shown in Fig. 9.35.

In larger auditoria, a more-detailed shaping of the ceiling profile may be needed to ensure even distribution of the reflected sound to the listeners. Notice in Fig. 9.36 that the concave overall shape of the ceiling can be maintained if just the size of the ceiling panel segments is large enough to redirect the reflections at suitably low frequencies.

Local reshaping of surfaces while maintaining the overall form is also illustrated in Fig. 9.37, in which the weak lateral reflections in a fan-shaped hall are improved by giving the walls a zigzag shape with almost parallel sets of smaller areas along the splayed side walls. In the example shown in the photo, the panels are even separated from the wall surface and tilted downward.

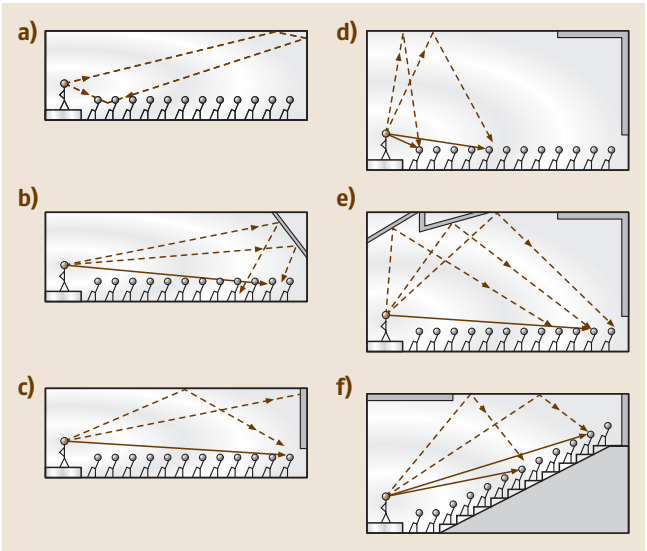


Fig. 9.34 Means of controlling early reflections in auditoria (after [9.28])

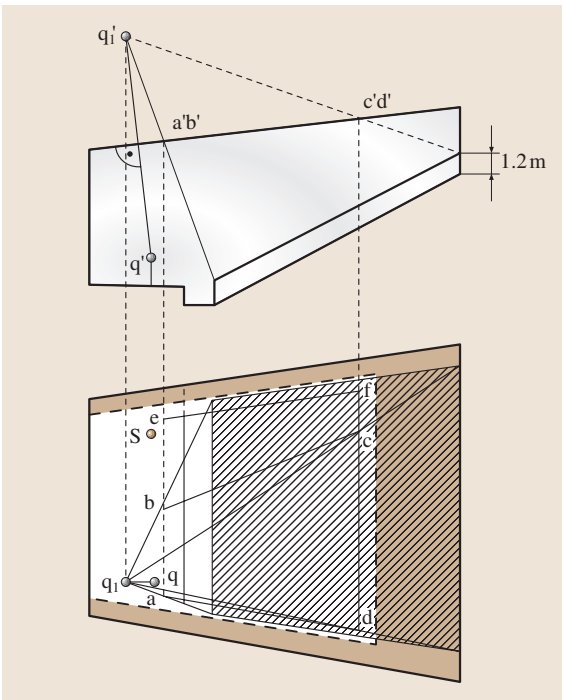


Fig. 9.35 Sketch identifying the area responsible for covering the audience area with a first-order ceiling reflection (after [9.32])

9.6.8 Suspended Reflectors

In many cases it is advantageous to reduce the delay and increase the level of a reflection without changing the room volume. In this case it is obvious to suggest individual reflecting surfaces suspended within the room boundaries, as shown to the right in Fig. 9.38. If the reflector is placed over an orchestra pit, it can improve mutual hearing among the musicians in the pit as well as increase the early energy from the singing on the stage to the seating primarily in the stalls area.

Suspended reflectors are also suitable for acoustic renovation of existing, listed buildings, because the main building structure is not seriously affected (as in the case of the Royal Albert Hall in London [9.3]).

For a given source position, a small reflector will cover only a limited area of receiver positions. There-

Amazon EX1014 Page 345

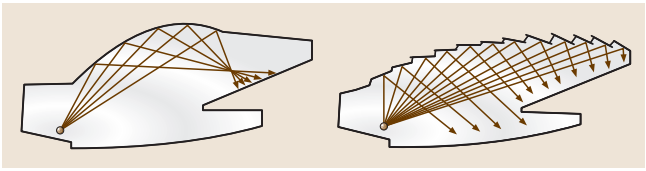


Fig. 9.36 Reshaping of ceiling profile to avoid focusing and provide more-even sound distribution

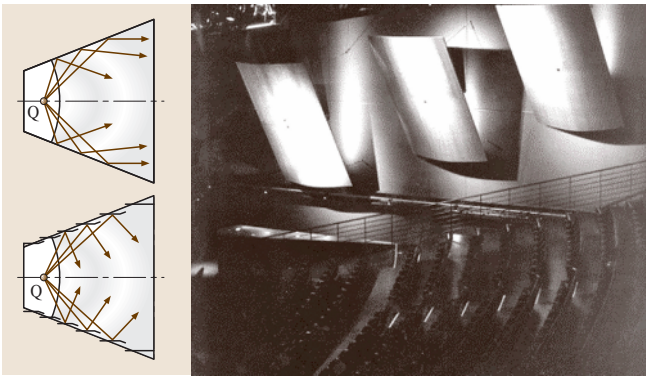


Fig. 9.37 Improvement of side-wall reflections in a fan-shaped hall by local reshaping or adding of panels parallel to the long axis of the room (drawing after [9.32])

fore, how the reflector(s) influence the balance heard between different sources such as the different sections in a symphony orchestra must be thoroughly considered. With small reflectors, the balance between sections may be perceived as very different in different parts of the audience. Often it will be advantageous to give the reflector(s) a slightly convex shape to extend the area covered by the reflection and to soften its level.

The nature of the reflection off a panel of limited size is strongly frequency dependent and may be influenced by several factors: distance to source and receiver, absorption by the panel material, diffraction governed by the ratio between panel size and the wavelength, and focusing/diffusion due to panel curvature.

The attenuation with distance is as for a spherical wave in free field: 6 dB per doubling of the total distance from the source to the reflector a_1 plus the distance from the reflector to the receiver a_2. Normally, the attenuation due to absorption is negligible, as reflectors should obviously be made from a reflective material. The only exception could be when a thin, lightweight material is used, such as plastic foil or coated canvas. In this case the attenuation ΔL_{abs} can be calculated from the mass law

$$\Delta L_{\text{abs}} = -10 \log_{10} \left[1 + \left(\frac{\rho c}{\pi f m \cos\theta} \right)^2 \right] \qquad (9.33)$$

where ρc equals the specific impedance of air ($\approx 414\,\text{kg}\,\text{m}^{-2}\,\text{s}^{-1}$), f is the frequency in Hz, m is the mass per square meter of the material and θ is the angle of incidence relative to the normal direction.

As a simple design guide, diffraction can be said to cause the specular reflection to be attenuated by 6 dB per octave below a limiting frequency f_{g} given by:

$$f_{\text{g}} = \frac{ca^*}{2S\cos\theta} \qquad (9.34)$$

where a^* is the weighted *characteristic distance* of the reflector from the source and receiver:

$$a^* = \frac{2a_1 a_2}{a_1 + a_2}\,, \qquad (9.35)$$

S is the area of the reflector, and θ is the angle of incidence as before. Above the frequency specified in (9.34), the influence of diffraction can be neglected.

If the reflector is curved, the attenuation ΔL_{curv} can be calculated as:

$$\Delta L_{\text{curv}} = -10 \log_{10} \left| 1 + \frac{a^*}{R\cos\theta} \right|\,. \qquad (9.36)$$

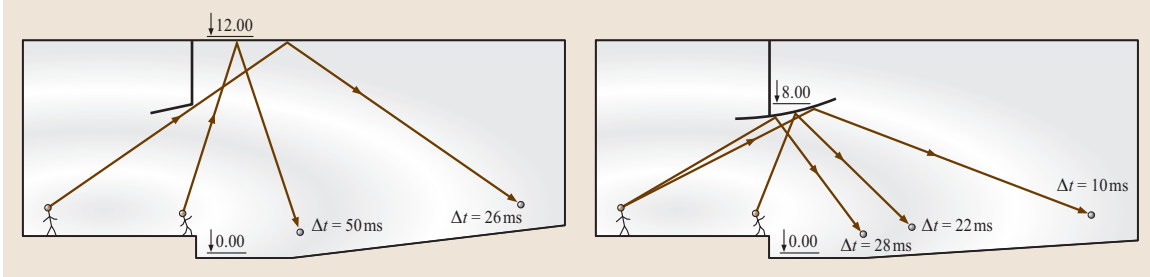


Fig. 9.38 The influence on reflection paths of a reflecting panel suspended in front of a stage (after [9.33])

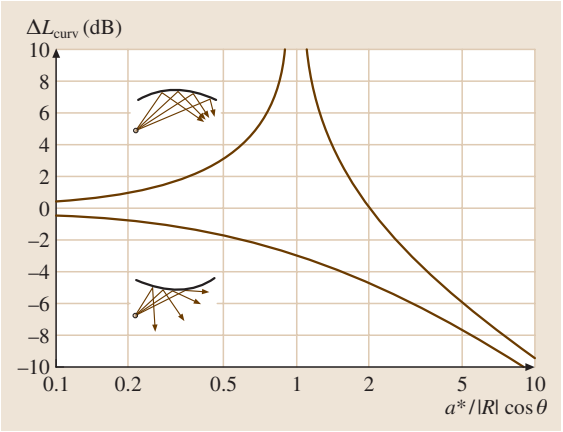


Fig. 9.39 Attenuation of a reflection caused by concave and convex reflectors (after [9.34])

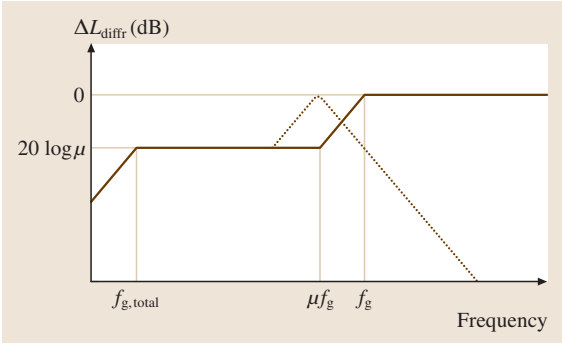


Fig. 9.40 Approximate attenuation due to diffraction from a reflector array. All non-horizontal lines have a slope of 6 dB per octave. See text for further explanation (after [9.34])

Here, the only new variable R is the radius of curvature. R should be entered positive for convex surfaces and negative for concave surfaces. When the reflector is concave, strong focussing occurs when $R = -a^*/\cos\theta$. In Fig. 9.39, graphs of ΔL_{curv} are shown for both the concave and convex cases. It is noted that the values can be both positive and negative. In the case where the surface is curved in both the x- and y-directions of the reflecting plane, the ΔL_{curv} correction should be applied twice using the appropriate radii of curvature in the two directions.

In certain cases an array of (identical) reflectors are suspended as an alternative to having one large reflector. Such an array can be characterized by the area of each element, S, the total area covered by the array S_{total} and the degree of coverage, μ, equal to the ratio between the total area of the n reflector elements nS and S_{total}. In this case, the attenuation due to diffraction depends on frequency as shown in Fig. 9.40.

Here, f_{g} is the limit frequency of the individual elements calculated according to (9.34), while $f_{\text{g,total}}$ is the lower-limit frequency related to the entire array area S_{total}

$$f_{\text{g,total}} = \frac{ca^*}{2S_{\text{total}}\cos\theta} \,. \tag{9.37}$$

First we look at the situation above the frequency f_{g} in Fig. 9.40. As in the case of the single reflector, the attenuation is zero as long as the specular reflection point falls on one of the reflector elements. If the specular point falls between two elements, the attenuation can be approximated by the dotted line, according to which we have an attenuation of 6 dB per octave as frequency increases. In the frequency range between $f_{\text{g,total}}$ and μf_{g}, the attenuation is roughly frequency independent and governed by the array density μ, and below $f_{\text{g,total}}$ the level again rolls off by 6 dB per octave. It is seen that, in order to obtain a wide range of frequency-independent reflection, one should aim to use many small reflectors rather than fewer, larger elements. Fortunately, the strong position-dependent variation in reflection level at high frequencies can be reduced by making the individual elements slightly convex.

It should be remembered that whenever surfaces are shaped with the purpose of efficiently directing sound energy from the source towards the absorbing audience, this energy is no longer available for the creation of the reverberant sound field. In other words, the more we try to improve the *clarity* and *intelligibility*, the more we reduce reverberation in terms of T and EDT (regardless of the physical absorption area), and the more the sound field will deviate from the diffuse field predicted by any simple reverberation formula. If substantial reverberation is needed as well, the solution may be to increase the volume per seat beyond the $10\,\text{m}^3$ as already suggested for large halls in Table 9.1. As seen in Sect. 9.7.3 this is in line with the current trend in concert hall design.

9.6.9 Sound-Diffusing Surfaces

It is often stated that diffusion of sound from irregular surfaces adds a pleasant, smooth character to the sound in a concert hall. It is a fact that reflections from large smooth surfaces can produce an unpleasant harsh sound. It is also a fact that most of the surfaces in famous old and highly cherished halls such as the Musikvereinsaal

break up and scatter the reflected sound in many directions due to rich ornamentation, many window and door niches and the coffered ceiling (Fig. 9.45). Even the large chandeliers in this hall break up the sound waves.

Many present-day architects are influenced by modernism and are less enthusiastic about filling their designs with rich ornamentation. Therefore, close cooperation is often needed between architect and acoustician in order to reach a surface structure that will scatter the sound sufficiently, although this is difficult to define exactly as good guidelines do not exist for what percentage of the wall/ceiling area one should make diffusive.

Schroeder [9.35] has done pioneering work on creating algorithms for the design of surface profiles that will ensure uniform distribution of scattered sound within a well-defined frequency range. These algorithms are based on number theory using the maximum length, quadratic residue or primitive root number sequences. However, in most cases of practical auditorium design, perfectly uniform scattering is not needed, and good results can be obtained even after a relaxed architectural adaptation of the ideas, as shown in Fig. 9.41.

The primary requirement to achieve diffusion is that the profile depth of the relief is comparable with a quarter of the wavelength and the linear extension of the diffusing element (before it is eventually repeated) is comparable with the wavelength at the lowest frequency of interest to be diffused. Given this, many different solutions can be imagined such as soft, curved forms cast in plaster, broken stone surfaces, wooden sticks, irregular brick walls etc.

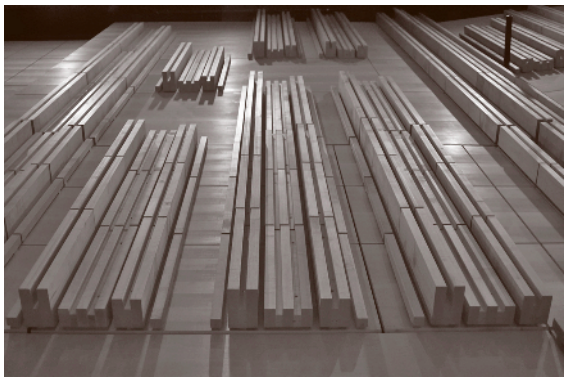


Fig. 9.41 Diffusion treatment by means of wooden sticks screwed on an otherwise plane, wooden wall surface

Finally, it should be mentioned that strict periodic structures with sharp edges equally spaced along the surface can cause problems of harsh sound at frequencies above the design frequency range. Actually, it is safer to apply more-organic, chaotic structures, or to ask the workers not to be too careful with the ruler.

9.7 Room Acoustic Design of Auditoria for Specific Purposes

This section will present some examples showing how the principles outlined in Sect. 9.5 are implemented in the current design of different types of auditoria. The main focus will be on concert halls for classical music as well as drama and opera theaters. It will also be shown that design choices are not only compromises between acoustic and other concerns, but sometimes even between different room acoustic aspects. This is often the case with multipurpose auditoria. In any case, the compromises are more serious the larger the seating capacity.

9.7.1 Speech Auditoria, Drama Theaters and Lecture Halls

In the Western world, rooms for spoken drama as well as for opera have their roots in the Greek and Roman theaters in which the audience is arranged in sloping concentric rows, the *cavea*. In Roman times, this audience area covered a 180° angle around the orchestra or apron stage. The advantage of such an arrangement is of course the possibility to accommodate a large audience within a limited maximum distance from the stage. Many modern theaters and speech auditoria have adopted this layout – perhaps extended with one or more sets of balconies.

An example of such a modern speech auditorium (the lecture hall at Herlev Hospital, Denmark) is seen in Fig. 9.42 along with a Roman theater for comparison.

In a drama theater, a moderate reverberation time around 1 s will provide a good compromise between high intelligibility and a high strength value, as both are needed for unamplified speech. The T value may be reduced to about 0.5 s in small rooms for 50–100 people and also increased slightly for audiences above 1000. To obtain a high strength value in a large theater, it is important to choose a modest volume per seat (about 5 m^3) rather than to control the reverberation in an unneces-

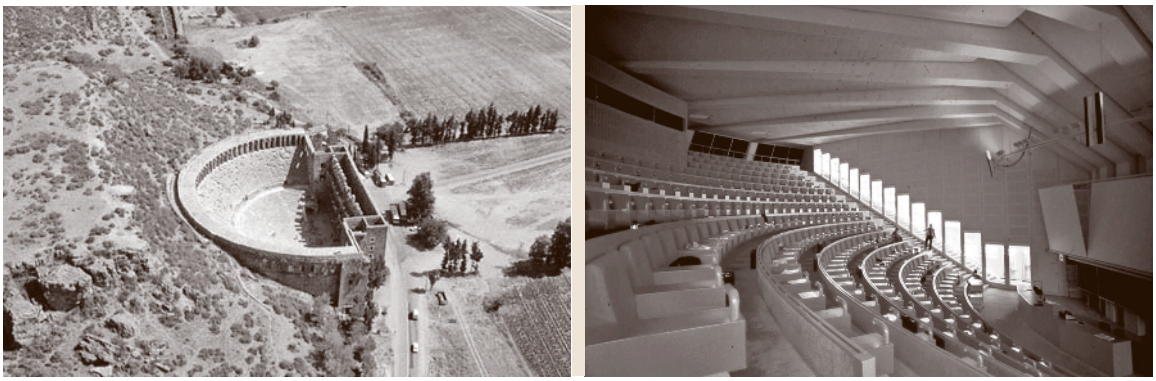


Fig. 9.42 Photos of Roman theater (in Aspendos, Turkey) and of a modern speech auditorium (Herlev Hospital, Denmark)

sarily large volume by applying additional absorption beyond that provided by the audience chairs. Thus, the height of the ceiling should be chosen fairly low – which is also advantageous for its ability to distribute early reflections to the audience. According to diffuse field theory, the steady-state energy density equivalent to G_{exp} is inversely proportional to the total absorption area A in the room.

High *clarity* and sound *level* are also promoted by careful shaping of the reflecting ceiling and a substantial sloping of the audience floor. Due to the directivity of the human voice, a reflecting rear wall (or stage scenery in a theatre) is of less importance, unless an actor should turn his back to the audience. Also the wall surfaces close to the stage opening are important, because they will send sound across the room so that sound also reaches people behind the speaker when he turns the head towards one side. The rear wall/balcony soffits behind/over the rear part of the audience should diffuse or redirect the sound to benefit nearby listeners.

As *spaciousness* is not a quality of particular importance in a speech auditorium, reflections from the side walls farther from the stage are often given lower priority than a low average stage-to-audience distance, which is promoted by a fan-shaped floor plan. However, as indicated by the problems encountered in the Olivier Theater in London [9.4], there should be limits to the opening angle of large fan-shaped theaters.

In a modern theater, the floor slope will often be more modest than seen in Fig. 9.42. In contrast to the lecture hall, a modern drama theater also needs to be highly flexible regarding the layout of stage and audience areas.

The classical proscenium frame inherited from the Italian Baroque theater is often considered to limit intimacy as well as the creativity of the stage-set designers and directors. If the classical proscenium stage is not needed, intimacy and audience involvement can be maximized by an arena arrangement, with audience all around the stage. In other cases, the stage can be extended into the audience chamber, for instance by removing the first rows of seats and raising the floor below to stage level. If such a flexible forestage is considered, the line-of-sight design should be adjusted accordingly, particularly if the hall features balconies.

An example of a flexible modern theater (without balcony) is seen in Fig. 9.43. In this room, the circular side/rear wall was made highly sound absorbing, leaving most of the distribution of early reflection energy to the ceiling.

Dedicated drama theaters are seldom built with a seat count higher than 500–1000, because visual as well as acoustic intimacy, including close view of facial expressions and freedom from artificial amplification, are given high priority. Thus, the maximum distance from the stage front to any seat should not exceed about 20 m for drama performances.

From this discussion it can be concluded that a reflecting ceiling of modest height is the main source of early reflection energy in theaters. However, lighting staff will often want to penetrate this surface with slots for stage lights and rigging or place light bridges across the room, which will act as possible barriers for the sound paths going via the ceiling. This challenge must always be considered in the acoustic design of drama theaters.

9.7.2 Opera Halls

There is a 350 year tradition of performing opera in halls of the Italian Baroque theater style. These are halls with a horse-shoe plan shape with several balcony levels running continuously from one side of the proscenium arch

Amazon EX1014 Page 349

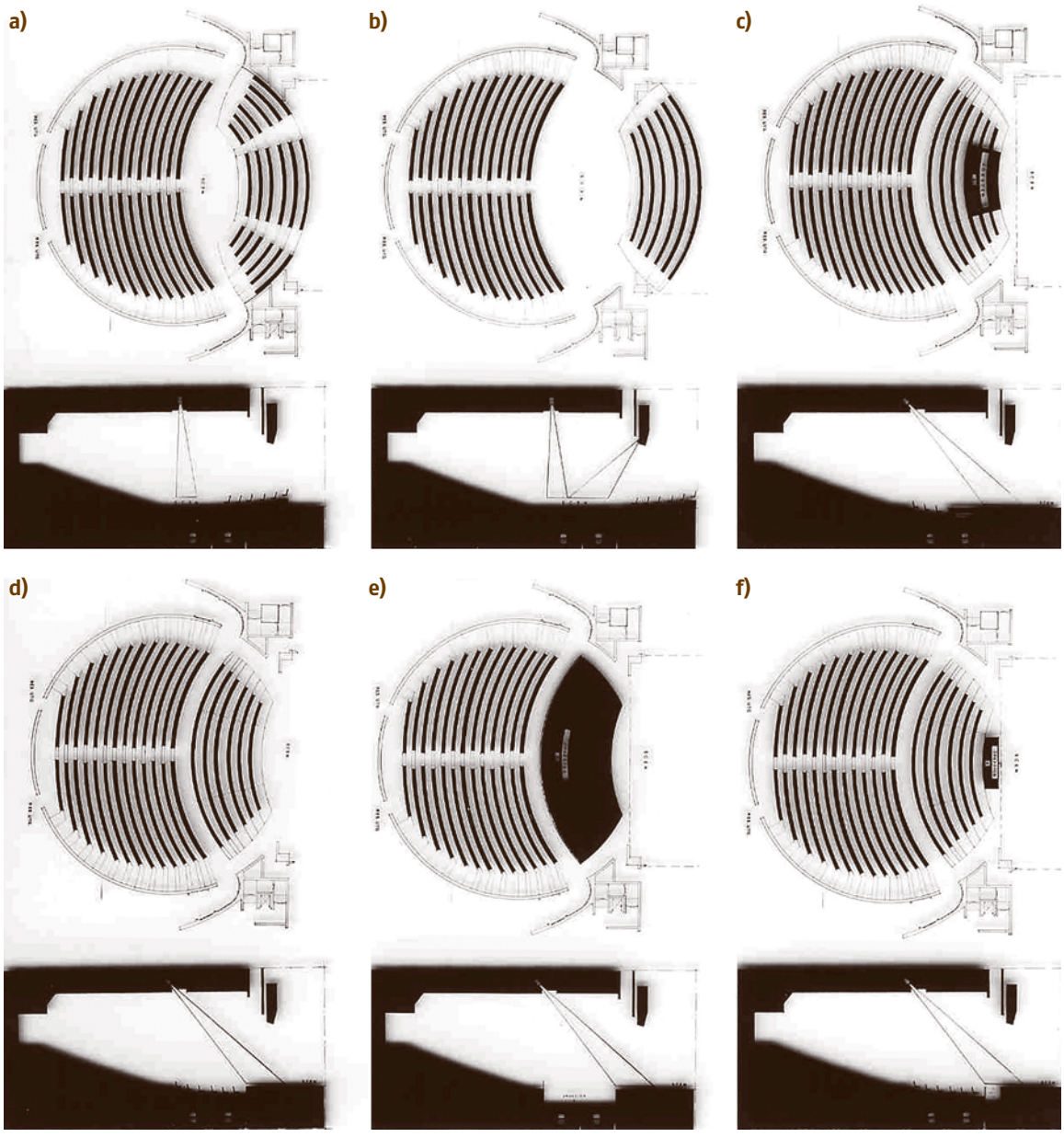


Fig. 9.43a–f Plans showing a flexible theater in Hëlsingborg, Sweden, with different layouts of the stage, orchestra pit and audience seating. **(a)** and **(b)**: arena with small and large stage respectively, **(c)**, **(e)**, and **(f)**: proscenium theater with three different sizes of orchestra pits. **(d)**: proscenium theater without pit

to the other along the curved side and rear walls. The balconies are either divided into boxes separated by thin walls or are open with long seat rows. The orchestra is placed in front of the proscenium opening that connects the auditorium with the stage house.

This tradition has undergone few changes during the past 350 years. Since the 19th century, the orchestra has been placed in a lowered pit separated from the stalls by a closed barrier. In this arrangement, seats in the stalls and lower rear bal-

Amazon EX1014 Page 350

cony experience an attenuation of the direct sound from the orchestra, because the pit rail blocks the free line of sight to the orchestra. The result may be an improved singer-to-orchestra balance in these seating areas. The pit rail will also reflect sound back and forth between the stage and pit, giving improved contact between the singers and the lowered orchestra.

Another development is that the separate boxes have been exchanged for open balconies, which improves the possibility for the now exposed side walls to help distribute the sound to seating farther backwards in the auditorium. This development is the result of advances in building technology as well as of changes in social attitudes.

The multiple levels of horseshoe-shaped balconies facilitates a short average distance between the audience and the singers, promoting both the visual and the acoustic intimacy, which is important for both parties. However, balcony seats close to the proscenium and on the upper balcony levels often have a very restricted view of the stage.

The short audience–singer distance obtained by the horseshoe form gives high levels of both strength and clarity. On the other hand, the horseshoe form and the proscenium arch also cause some acoustic problems:

- with all wall surfaces covered by balconies, only small free wall areas are left to generate the early reflections that assist an even distribution of the sound, clarity and the generation of reverberation in the room;
- often focusing effects occur from the curved wall surfaces and a cupola-shaped ceiling which, along with the placement of the orchestra out of sight in the pit, give a high risk of false localization of certain instruments;
- only part of the energy generated by the singer will reach the auditorium, because the singers are placed in a large stage house coupled to the auditorium via a proscenium opening of limited size, and this energy even depends on the acoustic properties of the ever-changing stage sets.

Recent developments in the acoustic design of opera halls have mainly aimed at avoiding these problems. Only a few attempts have been tried to radically change the overall shape (e.g. the Deutsche Oper in Berlin, and the Opéra Bastille in Paris) but without much success. Therefore, we are left with tradition and acoustic properties of horseshoe halls to shape the current acoustic ideals for opera, which are

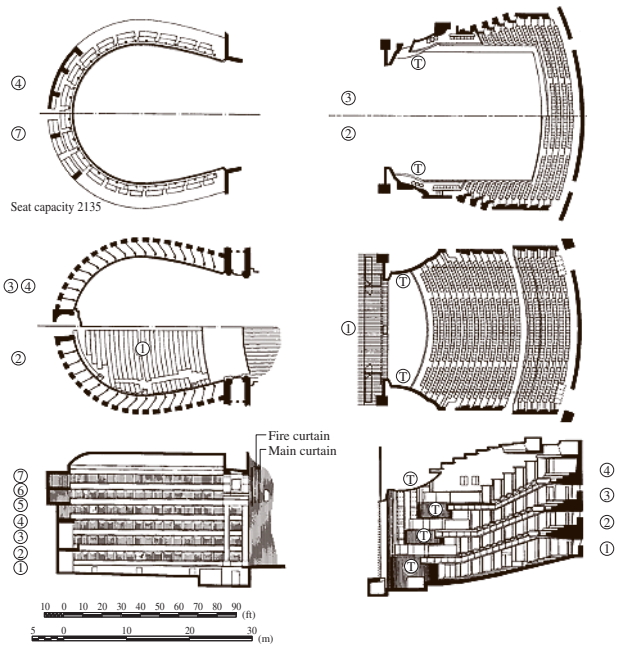


Fig. 9.44 Plan and section drawings of the Teatro alla Scala, Milan, Italy, from 1778 (*left*) and the New National Theatre Opera, Tokyo, Japan, from 1997 (*right*) (after [9.3])

- sufficient level of the singers' voices relative to the orchestra,
- some emphasis on *clarity* and *intelligibility* (although not as much as in speech auditoria),
- a certain *fullness of tone* and *reverberance*; but less than in a classical concert hall.

In recent years the trend has moved towards higher T values than found in the old halls, like Opéra Garnier, Paris (2131 seats, 1.1 s), Teatro Alla Scala, Milano (2289 seats, 1.2 s). This might well be due to the tendency to perform operas in the original language and display the translated text on a text board above the proscenium. Thus, for many recent opera halls we find T values of 1.4–1.8 s. Examples are: Göteborg (1390 seats, $T = 1.7$ s), Glyndebourne (1200 seats, $T = 1.4$ s), Tokyo (1810 seats, $T = 1.5$ s), and Semper Oper, Dresden (reconstruction, 1290 seats, $T = 1.6$ s).

As the width of the traditional horseshoe shape is about equal to its length, the lateral sound is not very pronounced and so *spaciousness* is seldom regarded a factor of particular importance in opera hall acoustics.

In order to ensure a sufficient level and *clarity* of the singers' sound, efficient sound reflections off the proscenium frame, ceiling, balcony fronts and from the

corners between balcony soffits and rear/side walls are important. Of course the scenery may also be designed to help in this respect, but generally this is out of the hands of the acoustician.

If a seat count larger than, say, 1500 is demanded it is better to extend the top balcony backwards, as seen in Covent Garden in London and the Metropolitan in New York rather than increase the overall dimensions of the horseshoe, which would give poorer acoustics for all seats. As this extended balcony is close to the ceiling surface, the acoustics here can often be very intimate, clear and loud, in spite of the physical and visual distance to the stage being very large. In the Metropolitan Opera in New York the farthest seat is about 60 m from the stage.

The culmination of the old Baroque theater tradition is represented by the Teatro alla Scala in Milan, Italy, which dates from 1778. Drawings of this theater are shown in Fig. 9.44 along with a modern opera with almost the same number of seats, the New National Opera in Tokyo from 1997.

The horseshoe-shaped Milan opera contains six levels of balconies subdivided into boxes with the openings covered with curtains so that the audience could decide whether to concentrate on the play or on socializing with their guests or family in the box. The audience in the boxes receive a rather weak sound unless they are very close to the opening.

This aspect has been improved in the modern opera with open balconies. Steeper slopes on the stalls floor and even on the side balconies have also improved the direct sound propagation to all seats. Fewer balcony levels have created higher openings at the balcony fronts, allowing the reverberant sound to reach the seats in the last rows below the balconies, providing more early reflection energy from the walls to all seats.

In the Tokyo opera, much emphasis has been put on retracting the balconies away from the proscenium so that extensive side-wall areas close to the proscenium are available for projecting the sound from the stage into the auditorium. Along with the ceiling above the pit, these wall areas almost form a trumpet flare. In the section, it is also seen that the side balconies are tilted downwards towards the stage in order to improve lines of sight. Finally, the side-wall areas and the side-balcony fronts have been made straight to avoid focusing and improve lateral reflection energy.

9.7.3 Concert Halls for Classical Music

Public concerts with large orchestras started in Europe about 250 years ago in halls that resembled those in which the composers normally found audience for their nonreligious music such as the salons and ballrooms of the nobility and courts. It is likely that this tradition as well as the limited span of wooden beam ceilings are the main reasons why many of the old concert rooms were built in what we now call the classical shoe-box shape: a narrow rectangle with a flat floor and ceiling and with the orchestra placed on a platform at one end.

The most famous of the shoe-box-shaped halls is the Musikvereinsaal in Vienna, Austria, shown in Fig. 9.45. This hall, dating from 1870, has a volume of 15 000 m^3 and seats 1670 people. Today, we understand many of the factors responsible for its highly praised acoustics:

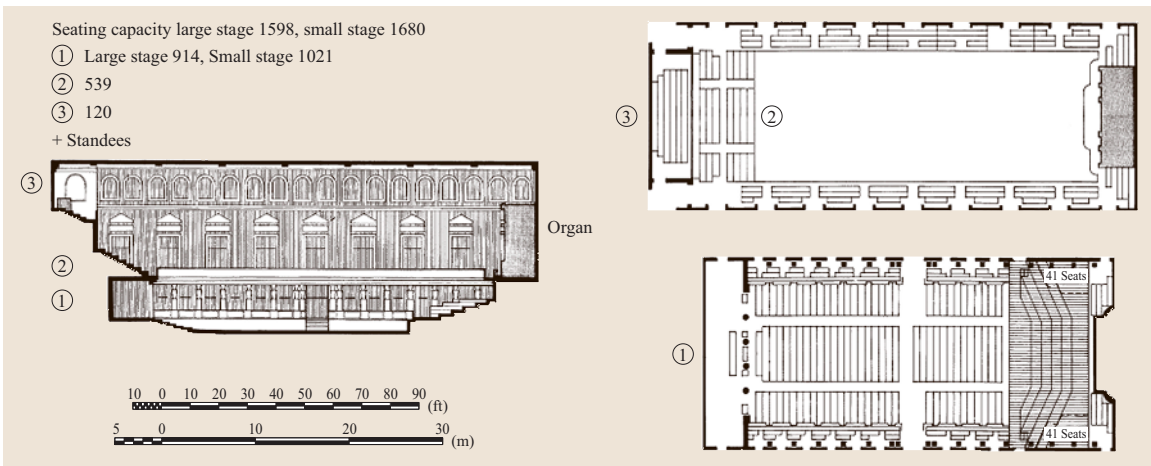


Fig. 9.45 The Musikvereinsaal, Vienna. Balcony and floor plans and long section (after [9.3])

- The volume per seat is only 9 m^3; but T is still 2.0 s in the fully occupied hall because the seating area is relatively small (the seats are narrow and the row-to-row distance is small) and because of the flat main floor. Thus, the sound becomes both loud and reverberant.
- The reverberation also benefits from the fact that the absorbing audience (apart from a few on the top rear balcony) is placed in the lower half of the volume, whereby a *reservoir* of reverberant energy can exist in the upper part of the room.
- The close side walls create strong lateral reflections giving high *clarity* and *envelopment*, and all surfaces are modulated with rich ornamentation, which prevents echoes and causes the room reflections to sound soft and pleasant.

There are however elements of the acoustics, and lack of comfort, that we would not accept in a modern concert hall. The sound is somewhat unclear and remote in the rear part of the flat main floor due to grazing-incidence attenuation. Also, in the side balconies only people sitting close to the balcony front can see well. Other old and famous shoe-box halls (such as the Boston Symphony Hall) have too-shallow balcony overhangs, which results in weak reverberant energy below the balconies.

Because of the general and proven success of the shoe-box design, many new halls are built with this shape. However, in most of the modern shoe-box halls, the drawbacks mentioned above are avoided by a slightly sloping stalls floor and a different layout of the side balconies.

One of the modern shoe-box halls in which the above modifications have been introduced is the Kyoto Concert Hall, Japan, shown in Fig. 9.46. Comparing Fig. 9.45 and Fig. 9.46, the increased floor slope and the subdivided side balconies turned towards the stage are clearly seen.

Recalling the influence of floor slope on T (and other parameters, Sect. 9.5.6) the increased slope will result in a reduced T and higher C. In modern designs T will also tend to be reduced by more-spacious seating and less-shallow balcony overhangs, which result in a larger absorptive audience area being exposed to the sound field. All these factors call for a larger volume per seat in modern shoe-box halls than found in the classical ones, if a high T value is to be maintained. On the other hand, this might well be the reason why the newer halls seldom provide the same intensity and warmth as experienced when listening from one of the sparsely upholstered wooden chairs in Vienna.

Early in the 20th century the development of concrete building technology allowed architects to shape the halls more freely. At the same time it became popular to treat sound as light, by designing the surfaces by means of geometric acoustics and by trying to transmit the sound as efficiently as possible from the source on stage to the (absorbing) audience. Another desire was to put the audience as close as possible to the stage.

The result was the fan-shaped concert hall, which dominated modernist architecture between about 1920 and 1980. An example is seen in Fig. 9.47: Kleinhans Music Hall in Buffalo, USA. Comparison with Fig. 9.45 reveals a very wide hall with a relatively low ceiling and the stage surrounded by concave, almost parabolic, walls and with the ceiling projecting the sound into the absorbing audience area after just a single reflection. This phenomenon, as well as the low volume per seat (about 6.4 m^3), are the reasons for the resulting low reverberation time and the weak sense of *reverberance*. The audience receives most of the sound from frontal directions or from the ceiling, so the sound is clear but lacks envelopment. Other examples of modernist fan-shaped halls are, the Salle Pleyel in Paris (1927) and the old Danish Broadcasting Concert Hall (Copenhagen, 1946). In conclusion, few concert halls from the modernist era have been acoustically successful.

Fortunately, mainly as a result of advances in room acoustic research, two alternative and more-promising designs appeared in the 1960s and 1970s: the *vineyard*

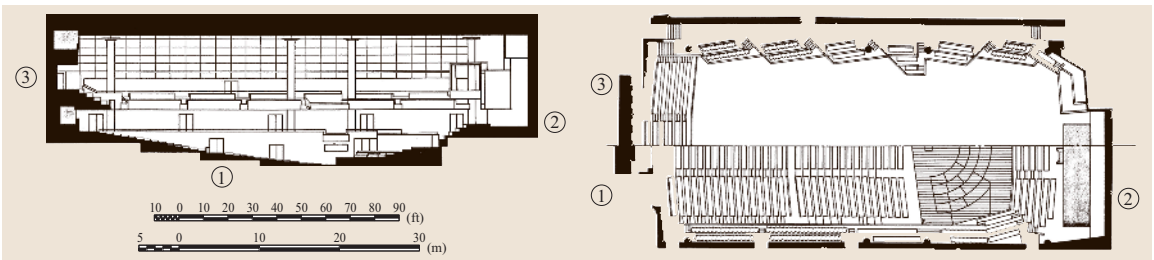


Fig. 9.46 Kyoto Concert Hall, Japan. Built 1995, volume 20 000 m^3, 1840 seats, RT = 2.0 s (after [9.3])

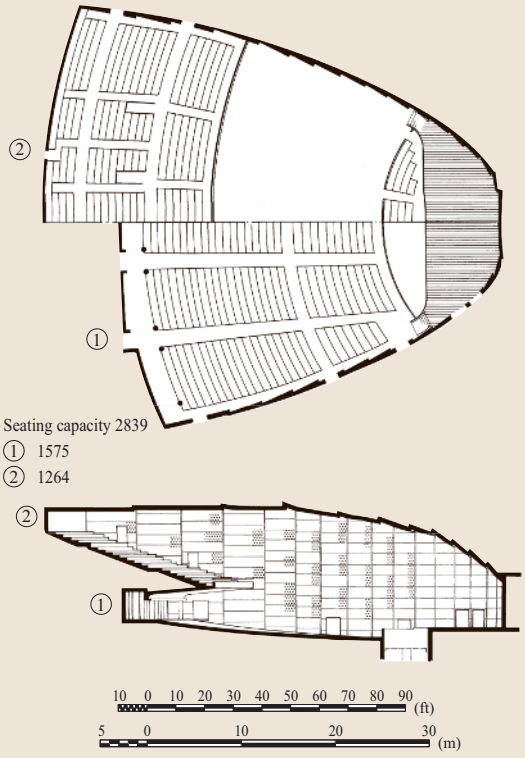


Fig. 9.47 Kleinhans Music Hall, Buffalo, USA. Built 1940, volume 18 240 m^3, 2839 seats, RT = 1.5 s. Plans and long section (after [9.3])

and the *directed reflection sequence* (DRS). Both are arena-shaped with the possibility of the audience being placed all around the orchestra platform. Even more than the fan shape, this causes the audience to come closer to the orchestra for the sake of intimacy. However, both designs are primarily relevant for larger halls, say more than 1500 seats. For halls seating more than about 2000, these designs may even be more successful than the classical rectangular shape, which, with so many seats, possess a risk of some listeners being placed too far from the stage.

The first vineyard concert hall was the Philharmonie Berlin, shown in Fig. 9.48, which opened in 1963. This hall has no balconies. Instead, the seating area is subdivided into terraces elevated relative to each other, whereby the terrace fronts and sides can act as *local* reflectors for specific seating areas. By careful design of these terraces, it is possible to provide plenty of early, and even lateral, reflections to most of the seats in an arena-shaped hall. With no seats being overhung by balconies, a large absorptive area is exposed to the sound, whereby a generous volume per seat is recommended. Thus, in the Sapporo Concert Hall in Japan (1997) as well as in the new Danish Radio Concert Hall in Copenhagen, the volume per seat is about 14 m^3. This is a quite high value for concert halls, as a 2 s T value should normally be obtained with just 10 m^3 per seat.

In *directed reflection sequence* halls such as the Christchurch Town Hall shown in Fig. 9.49, most of the early reflections are provided by huge, suspended and tilted reflectors. These are distinctly separate from

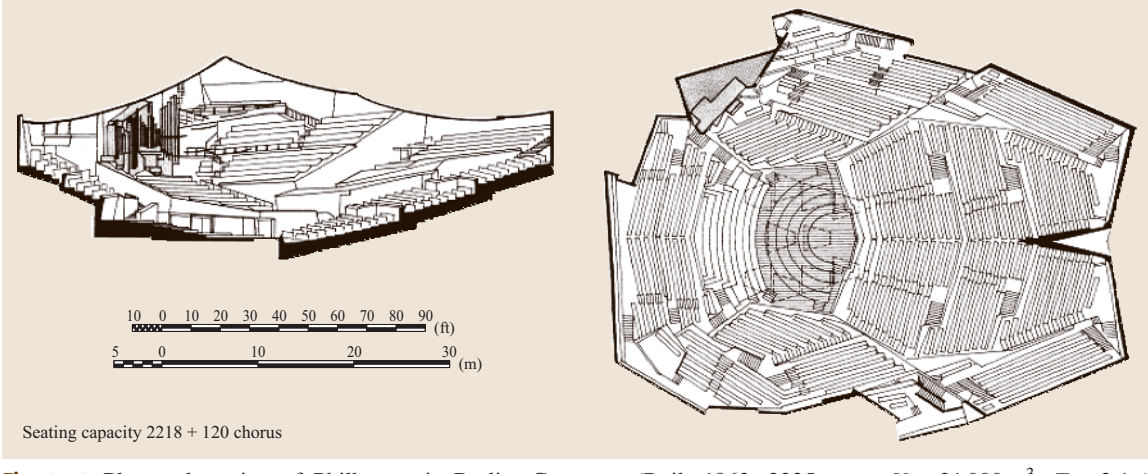


Fig. 9.48 Plan and section of Philharmonie Berlin, Germany. (Built 1963, 2335 seats, $V = 21\,000\,m^3$, $T = 2.1$ s) Acoustician: Lothar Cremer (after [9.3])

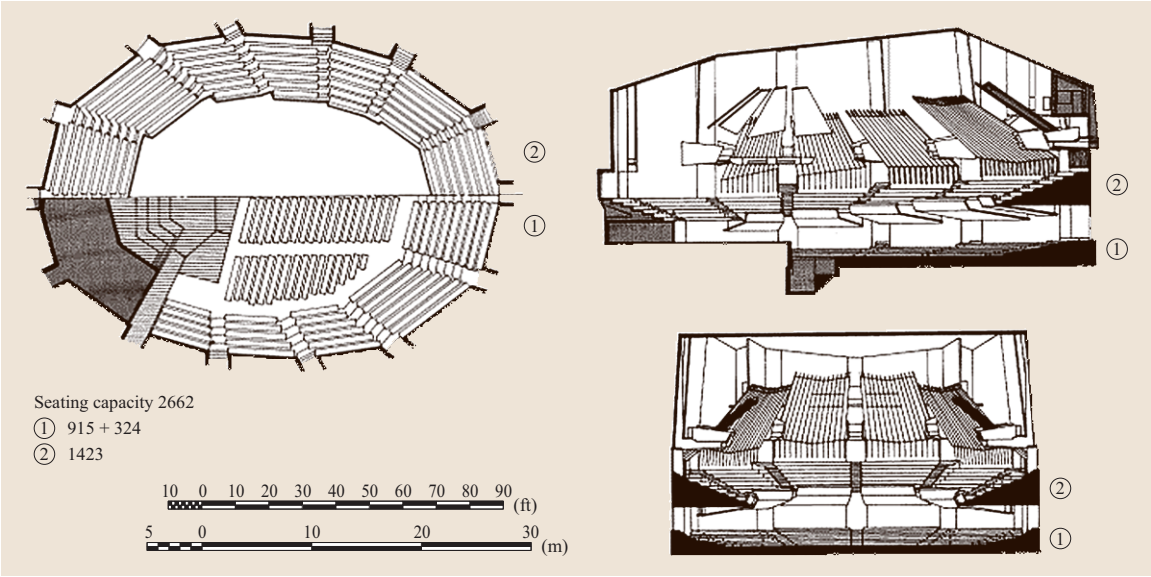


Fig. 9.49 Christchurch Town Hall, New Zealand. (Built 1972, 2662 seats, $V = 20\,500\,\mathrm{m}^3$, $T = 2.4\,\mathrm{s}$) Acoustician: Harold Marshall (after [9.3])

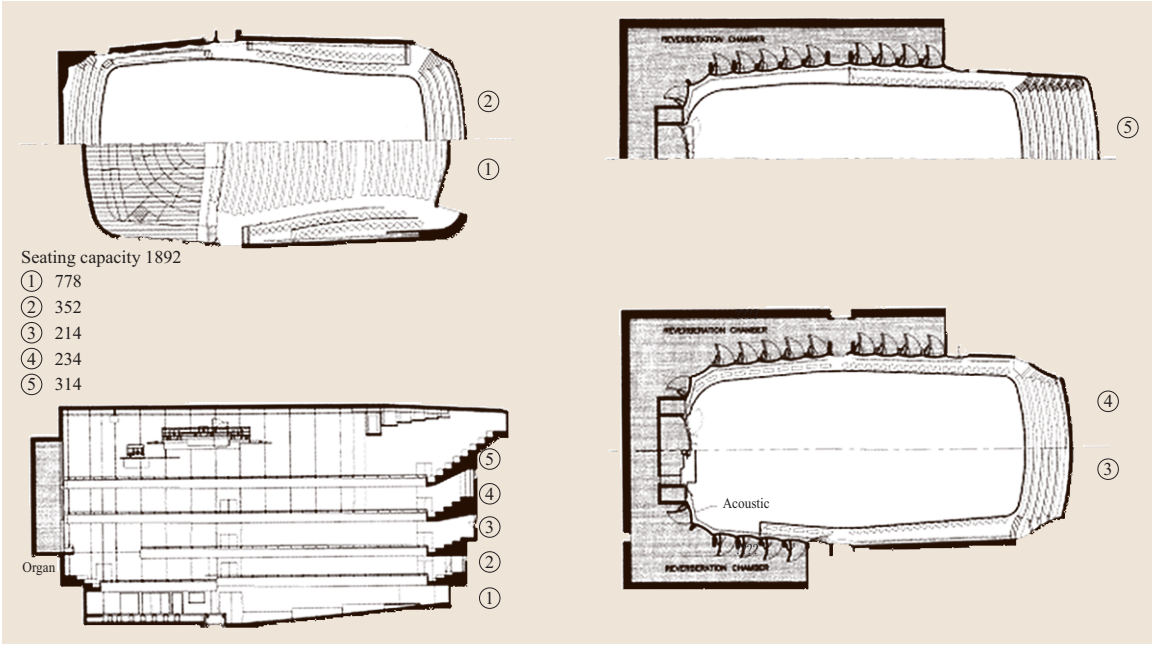


Fig. 9.50 The KKL Concert hall in Luzern, Switzerland. (Built 1999, 1892 seats, $V = 17800 + 6200\,\mathrm{m}^3$, $T = 1.8–2.2\,\mathrm{s}$) Acoustician: Russell Johnson (after [9.3])

the boundaries, which define the reverberant volume. The fronts and soffits of the sectioned balconies surrounding the main floor and stage in the elliptical plan provide additional early reflections. Because of the arena layout combined with the extensive use of balconies (containing more than half of the seats) the hall is very

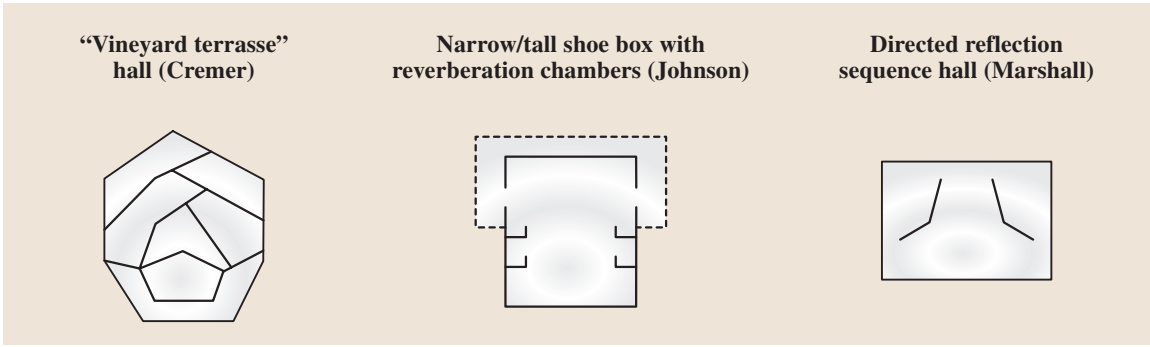


Fig. 9.51 Three concert hall concepts with the potential of combining high clarity with long reverberation. *From left to right*: vineyard (plan), shoe box with coupled reverberation chambers (cross section), and directed reflection sequence (cross section)

intimate, and the large reflecting surfaces provide a stunningly clear, dynamic and almost too loud sound. Still, the reverberation is long and rich because of the significant volume without absorption above/behind the reflectors.

This aim towards a combination of high *clarity* and high *reverberance*, probably developed through our extensive listening to recorded music, requires close reflecting surfaces as well as a large volume. The natural consequence of this demand is to separate the early reflecting surfaces from the room boundaries. Both the terrace fronts in the vineyard halls and the suspended reflectors in the DRS halls offer this possibility. A third way to achieve high *clarity* as well as high *reverberance* is a narrow shoe box with added surrounding volume, as found in a number of halls since about 1990. In these halls the volume of the narrow tall shoe-box hall providing the early reflections is quite moderate; but an extra volume is coupled to the main hall through openings that can be closed by heavy doors, so that the total volume can be varied. However, in such halls one should be careful to make the coupling area large enough for the added volume to have any significant effect (unless one sits close to one of the open doors), otherwise it is hard to justify the enormous costs of the extra volume and door system. With weak coupling, a double-sloped decay curve with a knee point perhaps 20 dB down is created, whereby the added, longer reverberation becomes barely audible except during breaks in the music.

A recent design of such a rectangular hall with a coupled volume is shown in Fig. 9.50.

The sketches in Fig. 9.51 summarize the basic design of the three types of halls mentioned. It is seen that they all possess the possibility of separating the surfaces that generate the early reflections from the volume boundaries that generate the reverberation.

9.7.4 Multipurpose Halls

Many modern halls built for cultural purposes often have to accommodate a variety of different types of events, from classical to pop/rock concerts, drama and musicals, opera, conferences, banquets, exhibitions, cinema and perhaps even sports events. From an acoustical point of view, the first concern in these cases is whether a variable reverberation time will be necessary. The answer is yes in most cases where some functions primarily require intelligibility of speech while others require a substantial reverberation, such as for classical music.

A hall in which these demands have been met by means of variable absorption is Dronningesalen, at the Royal Library in Copenhagen (Fig. 9.52). For this hall both chamber-music concerts and conferences with amplified speech were given high priority. The variable absorption is provided by means of moveable panels on the side walls as well as by folded curtains and roller blinds on the end walls. Combining these measures in different ways, T values in the range from 1.1 to above 1.8 s can be obtained (in the empty hall) as seen in Fig. 9.52.

In many cases, stage performances with extensive scenery are also required. The most common way to accomplish this is to design a stage house, mount an orchestra shell on the stage and raise the pit to stage level for concerts. If a hall is also to be used for banquets and exhibitions requiring a flat floor, it is common to place the stalls seats on a telescopic riser system on a flat floor instead of having a fixed, sloped seating. When not in use, this riser system is stored along the rear wall or un-

Amazon EX1014 Page 356

der a balcony. If the reverberation time suits classical music concerts with the sloped seating in place, a substantial area of variable absorption is needed to reduce the T value when the chairs are absent. Another problem may be that the telescopic systems offer only a rather steep and linear slope, which is not optimal for classical concerts. If the chairs are fixed on the riser steps, the minimum step height is about 25 cm, corresponding to a rake of about 25%.

If a change in T is accomplished by means of variable absorption, G will be reduced along with lowering T. However, if instead T is lowered by reducing the volume, for instance by moving a wall or lowering the ceiling, G will remain approximately constant or perhaps even increase. This can be advantageous when the low-T setting is to be used, for instance for unamplified drama or for chamber music in a larger symphony concert hall.

An example of a hall with variable volume is found in Umeå, Sweden (Fig. 9.53). In this hall the auditorium volume can be adjusted by moving the proscenium wall between three positions. When the proscenium is stored against the rear wall in the *concert* format, the entire volume is available for concerts with large symphonic orchestras on an open stage. In this situation, the stage tower can be closed off by horizontal panels at ceiling level. But when the proscenium is moved forward to the *opera* setting, the auditorium becomes smaller and a proscenium stage area is created, while the ceiling panels are removed for access to the fly tower above. In the third position, *drama*, the auditorium volume is further reduced to create an intimate theater. The variable elements also include a hinged side-wall section to improve the shape of the room for theater as well as moveable reflectors and variable absorption curtains above the grid ceiling level.

It should be mentioned that many of the problems related to the successful design of speech and music auditoria become more severe with increased size of the room. In other words, it is much easier to design a hall for less than 1000 people, like the two examples presented above, than for 2000 plus. The problems become even more complicated if multipurpose function is requested.

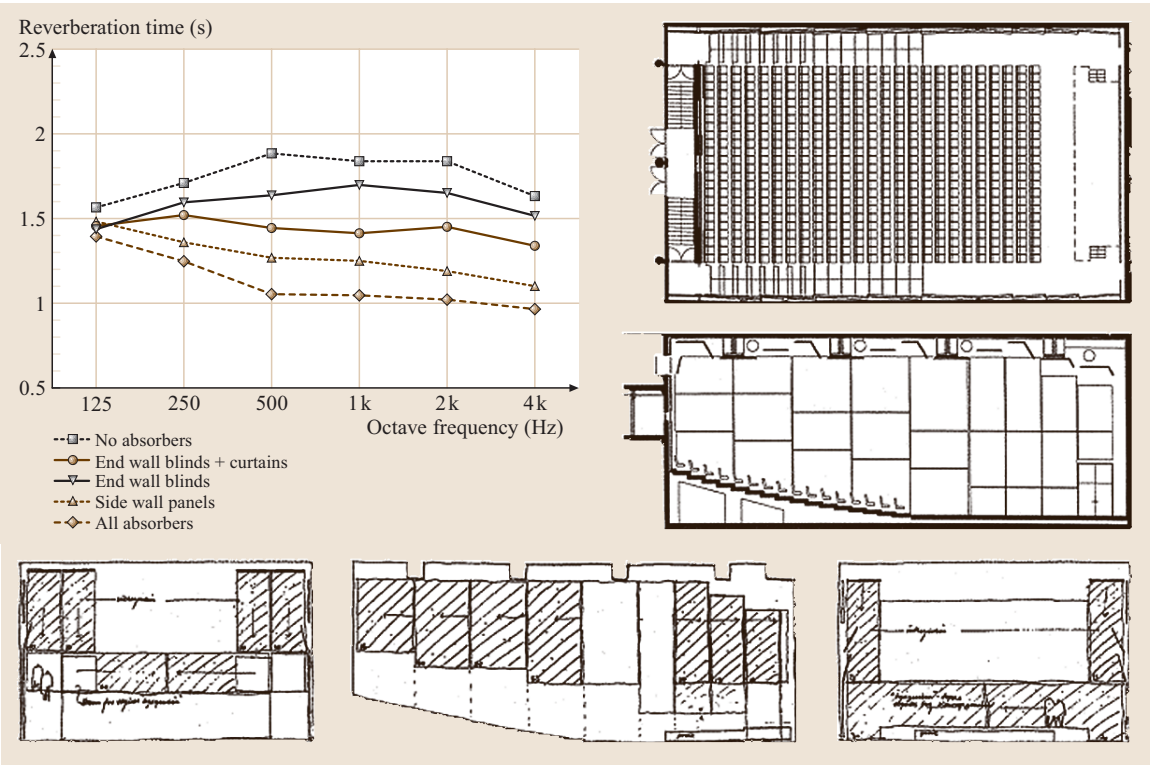


Fig. 9.52 Dronningesalen in the Royal Library in Copenhagen, Denmark; 1999, 400–600 seats. Reverberation time curves, plan, section and wall elevations with *hatched areas* indicating variable absorption

Amazon EX1014 Page 357

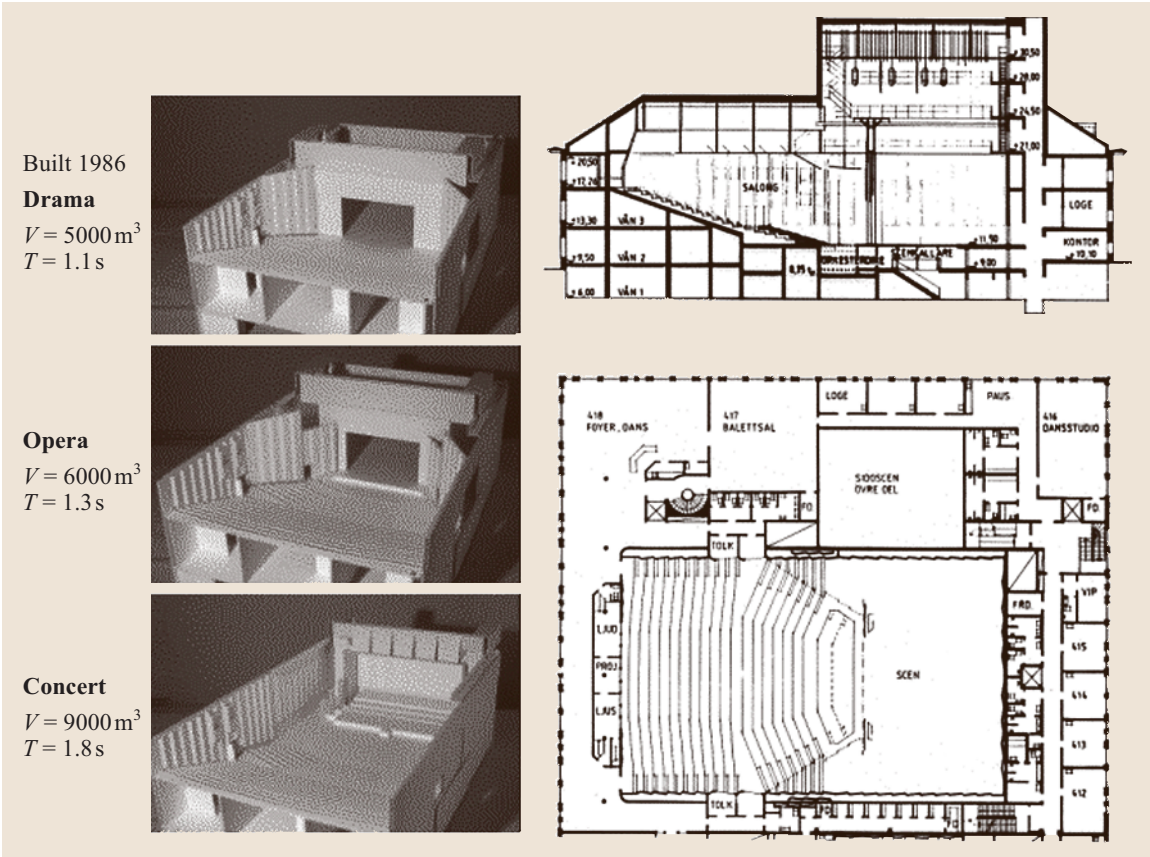


Fig. 9.53 The Idun multipurpose hall in Umeå, Sweden, with variable volume and stage configurations. The hall seats up to 860 people. Brief data (*left*), photos from cardboard model, long section and plan (*right*)

Still, a few examples of such successful designs exist. The most innovative is probably the Segerstrom Hall in Orange County, California, USA, which is shown in Fig. 9.54. This hall features a special layout of the almost 3000 audience seats, which are distributed on four levels. Each of these forms an almost rectangular, or even reverse fan, shape. However, the orientation of each level is shifted so that the overall plan becomes a wide fan. In this way the virtues of the fan shape for audience proximity to the stage are combined with the advantages of the rectangular shape for strong, early reflections from lateral directions. Thus, although the total width is about 50 m, the lateral energy level in this hall resembles that found in rectangular, 25 m-wide concert halls with typical seating capacity up to, say, 2000 people [9.36].

9.7.5 Halls for Rhythmic Music

Halls intended for rhythmic, amplified music concerts range in size from less than 100 to perhaps more than 50 000 listeners (in roofed sports arenas). The reason for this broad range of audience capacity is that the size and number of the sound amplification loudspeakers as well as of event video screens can be chosen at will to ensure good vision and adequate (often too high) sound levels for any size of audience. In any case, the reverberation time should be close to the optimal for speech, i. e., not more than 1 s if possible.

In huge arenas, of course, the reverberation time will be higher; but if efficient absorption is applied to the critical surfaces, a higher reverberation time need not compromise clarity. This is because the reverber-

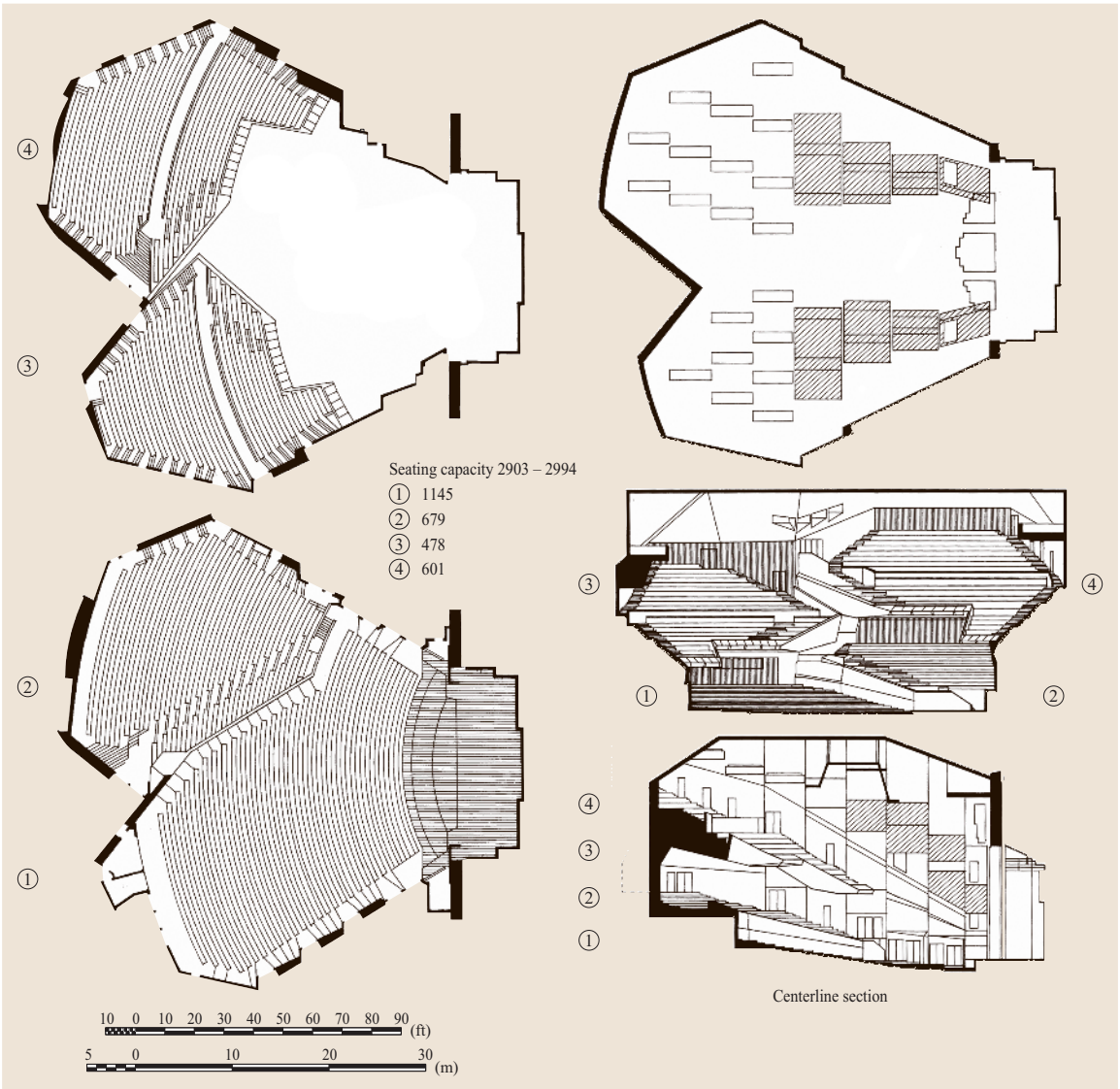


Fig. 9.54 Segerstrom Hall, Orange County, California, seating 2900 people. Acoustic design: Marshall, Hyde, Paoletti (after [9.3])

ation level will be low compared to the direct sound from the loudspeakers. In large spaces, designed to have a moderate T value, one must be careful to avoid echoes from distant, hard surfaces facing the often highly directional loudspeakers. Even small surfaces such as doors can cause echo problems if most other surfaces in the room are efficiently absorbing.

It is also very important to maintain modest low-frequency T values. A bass ratio larger than unity can create serious problems with control of the low-frequency levels. This is largely because most loudspeakers are almost omnidirectional at low frequencies, at which they willingly excite the room *reverberance*. In contrast, with high loudspeaker directivity in mid- and high-frequency range, it is easy to control the level in the audience area at these frequencies without spilling additional sound energy into the room.

Unfortunately, low-frequency reverberation control can be difficult to achieve, because low-frequency absorbers are normally less efficient (have lower α values) than mid-/high-frequency absorbers, and in a room in which these already occupy most of the available surfaces, it can be difficult to also find spaces for dedicated low-frequency absorbers. In particular, this poses a problem in buildings made from heavy concrete or masonry (which on the other hand is favorable for insulation of low-frequency sound towards the neighbors). When the low-frequency reverberation is too high the string bass and pedal drum will sound muddy. The audience area layout should be shaped to allow an even coverage by a sensible loudspeaker set up, and in large halls, a sloping floor is not suitable if a standing audience is expected. Fortunately, a sloping floor is not always mandatory, since in larger venues, the loudspeakers can be elevated to avoid grazing-incidence attenuation and elevated video projection screens improve the visual experience. It is advantageous to design the hall as well as the loudspeaker system so that zones with lower levels are also created, whereby people can have a chance to rest their ears, for instance while picking up refreshments in the bar. This may be accomplished by placing the bar in a smaller, partly decoupled, space with a shallower room height and an efficiently absorbing ceiling.

9.7.6 Worship Spaces/Churches

The liturgies in old Christian and Muslim worship were developed in highly reverberant cathedrals and mosques, both of which have their roots in Byzantine architecture. When new churches are built for the traditional Catholic or Protestant Christian denominations, it is customary to aim at a rich reverberation, which suit the pipe organ literature and congregational singing, while the intelligibility of the words by the priests is taken care of by a distributed public address (PA) sound system.

The same is true for new large mosques, in which the speech is amplified while the long reverberation supports traditional singing by the imams. Apart form these songs music is not performed in mosques.

In contrast to this, several Christian denominations, particularly in the US, have built new churches or temples (often accommodating several thousand people) and brought in newer art forms making use of large rock-music sound systems. In these spaces, therefore, the solution is often for the acoustician to create fairly dry, natural acoustics and to install both a heavy PA sound system as well as an artificial, electronic reverberation-enhancement system. Both types of sound systems are briefly described in Sect. 9.8.

9.8 Sound Systems for Auditoria

It is important to realize that sound systems will always be installed in both new and existing halls. As these systems are often used to amplify natural sound sources appearing on the stage in the hall, they may be regarded as a means for creating variable acoustics of the space. In this regard, the following will focus on how these systems interact with the natural acoustics, rather than talking about specific details or trends in the design of the loudspeakers themselves.

In principle, one may consider two different types of loudspeaker to be installed in an auditorium: traditional PA systems intended for the modification of the early part of the impulse response of the hall, and reverberation-enhancement systems intended for modification of the later part. Below, important aspects of both types will be briefly explored.

9.8.1 PA Systems

Normally, PA sound systems are installed with the purpose of increasing the sound level and/or the intelligibility of the performance. Use of the sound system for the reproduction of prerecorded sounds is of less interest in this context, as in this case it just acts like any other sound source working under the acoustic conditions provided by the natural acoustics of the room. In Sect. 9.2 we learned that, in order to improve *clarity/intelligibility* as well as the *level*, we need to increase the clarity C by adding sound components to the early part of the impulse response. (The alternative, to increase C by reducing the late part, can so far only be accomplished by adding physical absorption; but some day in the future, active systems with multiple microphones and loudspeakers integrated into *acoustic wallpaper* could be imagined as well.) Doing this requires that the main part of the sound energy leaving the loudspeakers hits the absorbing audience area. Alternatively, if substantial parts of the energy hit other, reflecting room surfaces, it may be reflected randomly and feed the reverberant field, which will reduce the clarity. For this reason, it is very important to apply loudspeakers with well-controlled directivity for this purpose, espe-

Amazon EX1014 Page 360

cially in auditoria with T values larger than optimal for speech.

For the acoustician, the most important technical specification for PA loudspeakers relates to their directivity, either in terms of vertical and horizontal radiation angles or in terms of the Q factor. Sometimes, the term *directivity index* $\mathrm{DI} = 10\log(Q)$ is used instead. Q is defined as the ratio between

1. the intensity emitted in the axis direction of the directive loudspeaker, and
2. the averaged intensity emitted by the same loudspeaker integrated over all directions.

The higher the Q value, the more the sound emitted will be concentrated in the forward direction (or rather, within the radiation angles specified).

When the loudspeaker is placed in a room, the sound field close to the loudspeaker will be dominated by the direct sound, whereas at longer distances, this component is negligible compared to the (diffuse) reverberant sound, as illustrated in Fig. 11.6. The (squared) sound pressure as a function of distance from the source in a room can be written as follows:

$$p^2(r) = \rho c\, P_{\mathrm{SP}} \left(\frac{Q_{\mathrm{SP}}}{4\pi r^2} + \frac{4}{A} \right) . \qquad (9.38)$$

In (9.38) the first term in the parenthesis is the direct sound component and the second one, $4/A$, the diffuse, reverberant component, P_{SP} is the emitted sound power, Q_{SP} is the Q factor of the loudspeaker, and A is the total sound absorption area of the room. Notice that in (9.38), for simplification, we assume a diffuse field intensity that is constant throughout the room, although in Sect. 9.5.1 and Fig. 9.13 we demonstrated that this is seldom the case.

Equation (9.38) shows that the direct component will dominate as long as the direct term is larger than the diffuse term. This will be the case for distances below the *critical distance* or the *reverberation distance*, $r_{\mathrm{cr,SP}}$. At this distance, the two components have equal magnitude, whereby

$$r_{\mathrm{cr,SP}} = \sqrt{\frac{Q_{\mathrm{SP}} A}{16\pi}} = \sqrt{\frac{Q_{\mathrm{SP}} V}{100\pi T}} . \qquad (9.39)$$

As seen from (9.39), we maximize the range within which we are able to provide the listeners with a clear sound from the speakers by increasing Q (as well as A).

One should be careful not to apply a larger number of loudspeakers than needed for coverage of the audience area, as each new loudspeaker will contribute to the generation of the reverberant diffuse field, while only one at a time will send direct sound towards a given audience position. Likewise, the number of open microphones should be minimized (which is often done automatically by an intelligent microphone mixer), as all open microphones will pick up and amplify the unclear reverberant field in the room.

The number of active loudspeakers and microphones should also be minimized in order to reduce the risk of feedback, which is a consequence of the microphone(s) picking up the sound from the loudspeaker(s) and the amplifier gain being set to high. If the amplification through the entire closed loop (microphone–amplifier–loudspeaker–room and back to the microphone) exceeds unity, the system will start to oscillate at a random frequency, which is emitted at maximum power level into the room, a most annoying experience.

The risk of feedback can be reduced by minimizing the loudspeaker–listener distance and the source–microphone distance, whereby a suitable listening level can be obtained with a moderate amplifier gain setting. Use of head-borne microphones represents the ultimate reduction of the source–microphone distance. Consequently, with such microphones the risk of feedback is substantially reduced. Also, the microphone–loudspeaker distance should be maximized and directional microphones and loudspeakers should be used and positioned so that the transmission from the loudspeaker back into the microphone is minimized. Thus, the loudspeakers and the microphones should be placed pointing away from each other with the microphone pointing towards the stage and the loudspeaker pointing toward the audience.

Considering the factors mentioned above, we still need to choose a configuration of the loudspeakers for a given room. Often the choice is between

1. a central system, consisting of a single or a limited number of highly directional units arranged close together in a cluster over the stage, or
2. a distributed system of several smaller units placed closer to the listeners.

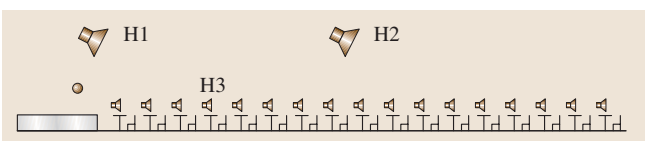


Fig. 9.55 Alternative solutions for loudspeaker coverage in an auditorium: a central (cluster) system H1, the same with a delayed secondary unit H2, or a highly distributed (pew back) system with small loudspeakers placed in each row of seats

These options are illustrated schematically in Fig. 9.55 as H1 and H3, respectively. H2 illustrates a compromise with a secondary unit or cluster assisting the main cluster, H1.

In meeting rooms or churches, highly distributed systems with small loudspeakers built into fixed meeting tables or pews can be very successful. With each loudspeaker being within a distance of say 1 m from the listener, each unit can be set to a very moderate level. Therefore, it is often possible to serve the listeners with a clear direct sound from such speakers without the total number of speakers exciting too much of the room reverberation. If the system is subdivided into independent sections, it is even possible to install automatic switches that only activate loudspeakers in zones where people are sitting. In this way, the reverberant field can be even further reduced in cases of fewer people attending the meeting or service. This is actually very fortunate as, especially in churches, fewer people also means less absorption and louder reverberation. The performance regarding intelligibility of pew-back systems with perhaps one hundred, moderately directive speakers is often equal to or better than when a highly directive central-cluster system is used. This is simply due to the moderate sound level needed from each of the closely placed loudspeakers compared to the much higher level needed for the cluster to reach the distant listeners. As can be imagined, the sound level distribution produced by the pew-back system will also be more uniform.

Of course, the highly distributed systems are not suitable for amplification of stage performances, where a high sound level of music is needed. For this purpose, a main cluster over the stage or a stereo system with loudspeakers on each side of the stage is preferable. For such systems, it is also more important to maintain a correct localization of the sound as coming from the performers on the stage.

In cases where some additional, distributed loudspeakers are needed to assist coverage in certain parts of the room, for instance on and under deep balconies, these are equipped with an electronic delay. If this delay is set so that the sound from the distributed assisting loudspeakers arrives 5–25 ms after the sound from the main loudspeakers, the listener will still perceive the sound as coming from the direction of the source. This is even true when the assisting loudspeakers are up to 5 dB louder than the main cluster at the listener's position. This very convenient phenomenon, called the precedence effect or Hass effect, is generally used as a guideline for the choice of level and delay settings in distributed loudspeaker systems.

9.8.2 Reverberation-Enhancement Systems

For auditoria where a long reverberation time is only needed occasionally or when the available room volume is simply inadequate, one may consider creating more reverberation by means of an electronic reverberation-enhancement system. Like the systems described in the previous section, these also consist of a number of loudspeakers and microphones connected by amplifiers; but there are many important differences.

The basic difference is that enhancement systems primarily attempt to add energy to the late part of the impulse response. Therefore, delays must be built into the signal processing between microphones and loudspeakers. In the first systems of this kind, *assisted*

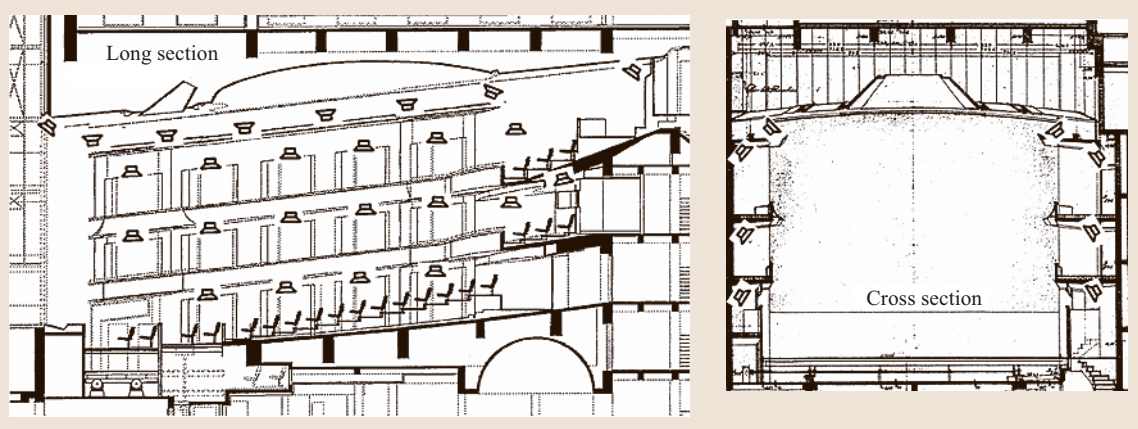


Fig. 9.56 Example of distribution of loudspeakers for reverberation enhancement in a theatre

resonance (AR) and *multichannel reverberation* (MCR), which appeared in the 1960s, this delay was provided by letting the loop gain of a large number of microphone–loudspeaker channels be tuned to just below the feedback limit. By spacing the microphones and loudspeakers far apart in the room (normally close to the ceiling), the sound would spend sufficient time traveling several times between loudspeakers and microphones for an audible prolongation of the reverberation to be perceived. In more-modern systems, electronic delays and reverberators are normally used.

In order to create the illusion of diffuse reverberation coming from the room itself, a large number of loudspeakers with low directivity must be distributed over the main surfaces in the room, as illustrated in Fig. 9.56. Since many speakers are needed, each can be fairly small and with limited acoustic power capability. The density and placement of these speakers as well as the balancing of their levels is essential to prevent localization of individual loudspeakers from any seat.

Regarding the microphones, these may be highly directive, but they must be placed quite far from the sound sources (and preferably out of sight) in order to cover the stage without the reverberant level being strongly dependent on the position of the sound source.

In view of what was explained in Sect. 9.8.1 about the measures needed to reduce the risk of feedback in loudspeaker systems, it is no wonder that most reverberation-enhancement manufacturers have to implement a solid strategy against uncontrolled feedback. One very efficient approach is to introduce time-varying delays in the sound-processing equipment, as this seems to destroy sharp room resonances when they start to build up to form a hauling frequency. Still, most enhancement systems work with a loop gain close to the feedback limit (which is actually participating actively in increasing the reverberation time), which is acceptable as long as it is under control and does not cause audible coloration of the sound.

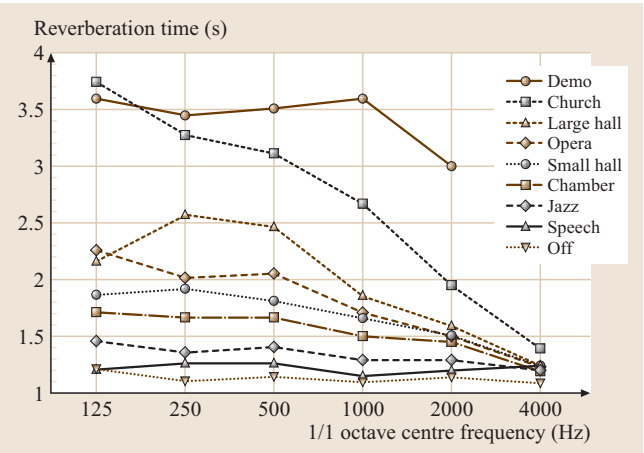


Fig. 9.57 Reverberation time values measured in a hall for different settings of the reverberation enhancement system (LARES)

It is essential for the acceptance of enhancement systems that neither audience nor performers experience the sound as coming from a loudspeaker system. Likewise, artefacts such as coloration due to feedback or the poor frequency characteristics of speakers are unacceptable.

The maximum increase in reverberation time possible with feedback-based systems like AR and MCR was typically about 50%. With modern digital reverberators, a much larger range can be achieved, as shown in Fig. 9.57. Increasing the reverberation time by a factor of three or more is not impossible. However, the illusion, when you still see the physical, much-dryer room around you, tends to break down when the value increases much above the 50%, which ironically is what the old systems could produce.

From a purist standpoint, reverberation enhancement can be regarded as an acoustic prosthesis, but in many cases a pragmatic approach to the use of these systems can increase the range of programs that a cultural venue can host.

References

9.1 A.C. Gade: *Subjective Room Acoustic Experiments with Musicians. PhD thesis, Rep. No. 32* (Technical University of Denmark, Copenhagen 1982)

9.2 W.C. Sabine: *Collected Papers on Acoustics* (Peninsula, Los Altos 1994), Reissued

9.3 L.L. Beranek: *Concert Halls and Opera Houses, Music Acoustics and Architecture*, 2nd edn. (Springer, Berlin, New York 2004)

9.4 M.F.E. Barron: *Auditorium Acoustics and Architectural Design* (Spun, London 1993)

9.5 H. Kuttruff: *Room Acoustics* (Applied Science, London 1976)

9.6 ISO: *ISO 3382*, Acoustics – Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters (ISO, Geneva 1997)

Amazon EX1014 Page 363

9.7 M. Vorländer: *International Round Robin on Room Acoustical Computer Simulations, Proc. 15th ICA* (Trondheim, Norway 1995) pp. 689–692

9.8 J.S. Bradley, G.A. Soloudre: The influence of late arriving energy on spetial impression, J. Acoust. Soc. Am. **97**(4), 2263–2271 (1995)

9.9 L.L. Beranek: *Music, Acoustics and Architecture* (Wiley, New York 1962)

9.10 Personal communication (summer 2006)

9.11 A.C. Gade: Investigations of musicians' room acoustic conditions in concert halls I, methods and laboratory experiments, Acustica **69**, 193–203 (1989)

9.12 A.C. Gade: Investigations of musicians' room acoustic conditions in concert halls II, Field experiments and synthesis of results, Acustica **69**, 249–262 (1989)

9.13 T. Houtgast, H.J.M. Steeneken: A tool for evaluating auditoria, Brüel Kjær Tech. Rev. **3**, 3–12 (1985)

9.14 J.S. Bradley: Predictors of speech intelligibility in rooms, J. Acoust. Soc. Am. **80**, 837–845 (1986)

9.15 M.R. Schroeder: New method of measuring reverberation time, J. Acoust. Soc. Am. **37**, 409 (1965)

9.16 J.S. Bradley: Accuracy and reproducibility of auditorium acoustics measures, Proc. IOA **10**, 399–406 (1988), Part 2

9.17 K. Sekigouchi, S. Kimura, T. Hanyuu: Analysisi of sound field on spatial information using a four-channel microphone system based on regular tetrahedron peak point method, Appl. Acoust. **37**, 305–323 (1992)

9.18 D. de Vries, A.J. Berkhout, J.J. Sonke: *Array Technology for Measurement and Analysis of Sound Fields in Enclosures, AES 100th Convention 1996 Copenhagen, Preprint 4266 (R-4)* (AES, San Francisco 1996)

9.19 C.M. Harris: Absorption of air versus humidity and temperature, J. Acoust. Soc. Am. **40**, 148–159 (1966)

9.20 C.W. Kosten: The mean free path in room acoustics, Acustica **10**, 245 (1960)

9.21 L.L. Beranek: Analysis of Sabine and Eyring equations and their applications to concert hall audience and chair absorption, J. Acoust. Soc. Am. **120** (2006)

9.22 L.L. Beranek, T. Hidaka: Sound absorption in concert halls by seats, occupied and unoccupied, and by the hall's interior surfaces, J. Acoust. Soc. Am. **104**, 3169–3177 (1998)

9.23 S. Sakamoto, T. Yokota, H. Tachibana: Numerical sound field analysis in halls using the finite difference time domain method. In: *Proc. Int. Symp. Room Acoustics: Design and Science (RADS)*, Awaji Island (ICA, Kyoto 2004)

9.24 F. Spandöck: Raumakustische Modellversuche, Ann. Phys. **20**, 345 (1934)

9.25 A.C. Gade: Room acoustic properties of concert halls: quantifying the influence of size, shape and absorption area, J. Acoust. Soc. Am. **100**, 2802 (1996)

9.26 M. Barron, L.-J. Lee: Energy relations in Concert Auditoria, J. Acoust. Soc. Am. **84**, 618–628 (1988)

9.27 W.-H. Chiang: *Effects of Various Architectural Parameters on Six Room Acoustical Measures in Auditoria, PhD thesis* (University of Florida, Gainesville 1994)

9.28 W. Fasold, H. Winkler: *Bauphysikalische Entwurfslehre, Band 5; Raumakustik* (Verlag Bauwesen, Berlin 1976)

9.29 Z. Maekawa, P. Lord: *Environmental and Architectural Acoustics* (Spon, London 1994)

9.30 W. Furrer, A. Lauber: *Raum- und Bauakustik Lärmabwehr* (Birkhaüser, Basel 1976)

9.31 J.H. Rindel: *Anvendt Rumakustik, Note 4201, Oersted DTU* (Technical University of Denmark, Copenhagen 1984), (In Danish)

9.32 L.I. Makrinenko: *Acoustics of Auditoriums in Public Buildings* (ASA, Melville 1994)

9.33 J. Sadowski: *Akustyka architektoniczna* (PWN, Warszawa 1976)

9.34 J.H. Rindel: *Acoustic design of reflectors in Auditoria, Note 4212 Oersted DTU* (Technical University of Denmark, Copenhagen 1999)

9.35 M.R. Schroeder: Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion, J. Acoust. Soc. Am. **65**, 958–963 (1979)

9.36 J.R. Hyde: *Segerstrom Hall in Orange County – Design, Measurements and Results after years of operation*, Proc. IOA, Vol. 10 (IOA, St. Albans 1988), Part 2

Amazon EX1014 Page 364

10. Concert Hall Acoustics Based on Subjective Preference Theory

This chapter describes the theory of subjective preference for the sound field applied to designing concert halls. Special attention is paid to the process of obtaining scientific results, rather than only describing a final design method. Attention has also been given to enhancing satisfaction in the selection of the most preferred seat for each individual in a given hall. We begin with a brief historical review of concert hall acoustics and related fields since 1900.

A neurally grounded theory of subjective preference for the sound field in a concert hall, based on a model of the human auditory–brain system, is described [10.1]. Most generally, subjective preference itself is regarded as a primitive response of a living creature and entails judgments that steer an organism in the direction of maintaining its life. Brain activities relating to the scale value of subjective preference, obtained by paired-comparison tests, have been determined. The model represents *relativity*, relating the autocorrelation function (ACF) mechanism and the interaural cross-correlation function (IACF) mechanism for signals arriving at the two ear entrances. The representations of ACF have a firm neural basis in the temporal patterning signal at each of the two ears, while the IACF describes the correlations between the signals arriving at the two ear entrances. Since Helmholtz, it has been well appreciated that the cochlea carries out a rough spectral analysis of sound signals. However, by the use the of the spectrum of an acoustic signal, it was hard to obtain factors or cues to describe subjective responses directly. The auditory representations from the cochlea to the cortex that have been found to be related to subjective preference in a deep way involve these temporal response patterns, which have a very different character from those related to power spectrum analyses. The scale value of subjective preference of the sound field is well described by four orthogonal factors. Two are temporal factors (the initial delay time between the direct sound and the first reflection, Δt_1, and the reverberation time, T_{sub}) associated with the left cerebral hemisphere, and two are spatial factors [the binaural listening level (LL) and the magnitude of the IACF, the IACC] associated with the right hemisphere. The theory of subjective preference enables us to calculate the acoustical quality at any seat in a proposed concert hall, which leads to a seat selection system.

The temporal treatment enables musicians to choose the music program and/or performing style most suited to a performance in a particular concert hall. Also, for designing the stage enclosure for music performers, a temporal factor is proposed. Acoustical quality at each seating position examined in a real hall is confirmed by both temporal and spatial factors.

10.1 **Theory of Subjective Preference for the Sound Field** 353
10.1.1 Sound Fields with a Single Reflection 353
10.1.2 Optimal Conditions Maximizing Subjective Preference 356
10.1.3 Theory of Subjective Preference for the Sound Field 357
10.1.4 Auditory Temporal Window for ACF and IACF Processing 360
10.1.5 Specialization of Cerebral Hemispheres for Temporal and Spatial Factors of the Sound Field 360
10.2 **Design Studies** 361
10.2.1 Study of a Space-Form Design by Genetic Algorithms (GA) 361
10.2.2 Actual Design Studies 365
10.3 **Individual Preferences of a Listener and a Performer** 370
10.3.1 Individual Subjective Preference of Each Listener 370
10.3.2 Individual Subjective Preference of Each Cellist 374
10.4 **Acoustical Measurements of the Sound Fields in Rooms** 377
10.4.1 Acoustic Test Techniques 377
10.4.2 Subjective Preference Test in an Existing Hall 380
10.4.3 Conclusions 383
References 384

Amazon EX1014 Page 365

In the first half of the 20th century, studies were mostly concentrated on temporal aspects of the sound field. In 1900 *Sabine* [10.2] initiated the science of architectural acoustics, developing a formula to quantify reverberation time. In 1949, *Haas* [10.3] investigated the echo disturbance effect from adjustment of the delay time of the early reflection by moving head positions on a magnetic tape recorder. He showed the disturbance of speech echo to be a function of the delay time, with amplitude as a parameter.

After investigations with a number of existing concert halls throughout the world, *Beranek* in 1962 [10.4] proposed a rating scale with eight factors applied to sound fields from data obtained by questionnaires on existing halls given to experienced listeners. Too much attention, however, has been given to monaural temporal factors of the sound field since Sabine's formulation of reverberation theory. For example, the Philharmonic Hall of Lincoln Center in New York, opened in 1962, was not satisfactory to listeners even after many improvements.

On the spatial aspect of the sound field, *Damaske* in 1967 [10.5] investigated subjective diffuseness by arranging a number of loudspeakers around the listener. In 1968, *Keet* [10.6] reported the variation of apparent source width (ASW) in relation to the short-term cross-correlation coefficient between two signals fed into the stereo loudspeakers as well as the sound pressure level. *Marshall* in 1968 [10.7] stressed the importance of early lateral reflections of just 90 degrees, and *Barron* in 1971 [10.8] reported *spatial impressions* or *envelopment* of sound fields in relation to the short-term interaural cross-correlation coefficient.

As a typical spatial factor of the sound field, *Damaske* and *Ando* in 1972 [10.9] defined the IACC as the maximum absolute value of the interaural cross-correlation function (IACF) within the possible maximum interaural delay range for the human head, such that

$$\mathrm{IACC} = |\phi_{\mathrm{lr}}(\tau)|_{\max} \ , \text{for } |\tau| < 1\,\mathrm{ms} \qquad (10.1)$$

and proposed a method of calculating the IACF for a sound field. In 1974, *Schroeder* et al. in [10.10] reported results of paired-comparison tests asking listeners which of two music sound fields was preferred. In an anechoic chamber, sound fields in existing concert halls were reproduced at the ears of listeners through dummy head recordings and two loudspeaker systems, with filters reproducing spatial information. They found that two significant factors, the reverberation time and the IACC, had a strong influence on subjective preference.

In 1977, *Ando* discussed subjective preference in relation to the temporal and spatial factors of the sound field simulated with a single reflection [10.11]. In 1983, he described a theory of subjective preference in relation to four orthogonal factors consisting of the temporal and spatial factors for the sound field, which enable us to calculate the scale value of the subjective preference at each seat [10.12, 13]. *Cocchi* et al. reconfirmed this theory in an existing hall in 1990 [10.14]. In 1997, *Sato* et al. [10.15] reconfirmed this clearly by use of the paired-comparison judgment in an existing hall, switching identical loudspeakers located at different positions on the stage, instead of changing the seats of each subject. This method may avoid effects of other physical factors than the acoustic factors. In addition to the orthogonal factors, they found the interaural delay in the IACF, τ_{IACC}, as a measure of image shift of the sound source that is to be kept at zero realizes good balance in the sound field. Using this method, the dissimilarity distance has also been described by the temporal and spatial factors of the sound field in 2002 [10.16].

Thus far, the theory has been based on the global subjective attributes for a number of subjects. The theory may be applied for enhancing each individual's satisfaction by adjusting the weighting coefficient of each orthogonal factor [10.1], even though a certain amount of individual differences exist [10.17]. The seat selection system [10.18] was introduced in 1994 after construction of the Kirishima International Concert Hall.

For the purpose of identifying the model of the auditory–brain system, experiments point toward the possibility of developing the correlation between brain activities, measurable with electroencephalography (EEG) [10.19, 20]. Correspondences between subjective preference and brain activity have been found from EEG and magnetoencephalography (MEG) studies in human subjects [10.21, 22], but details of these results are not included in this chapter due to limited space. Results show that orthogonal factors comprise two temporal factors: the initial time delay gap between the direct sound and the first reflection (Δt_1) and the subsequent reverberation time (T_{sub}) are associated with the left hemisphere, and two spatial factors: the listening level (LL) and the magnitude of the IACF (IACC) are associated with the right hemisphere. The information corresponding to subjective preference of the sound field can be found in the effective duration of the autocorrelation function (ACF) of the alpha (α) waves of both EEG and MEG. A repetitive feature in the α-wave, as measured in its ACF at the preferred condition, has

been found. The evidence ensures that the basic theory of subjective preference may also be applied to each individual preference [10.21].

Since individual differences of subjective preference in relation to the IACC in the spatial factor are small enough, at the first stage of acoustic design we can determine the architectural space form of the room. The temporal factors are closely related to the dimensions of a specific concert hall, which can be altered to exhibit specific types of music, such as organ music, chamber music or choral works.

On the neural mechanism in the auditory pathway, a possible mechanism for the interaural time difference and correlation processors in the time domain was proposed by *Jeffress* in 1948 [10.23], and by *Licklider* in 1951 [10.24]. By recording left and right auditory brainstem responses (ABR), Ando et al. in 1991 [10.25] revealed that the maximum neural activity (wave V at the inferior colliculus) corresponds well to the magnitude of the interaural cross-correlation function. Also, the left and right waves $IV_{l,r}$ are close to the sound energies at the right- and left-ear entrances. In fact, the time-domain analysis of the firing rate of the auditory nerve of a cat reveals a pattern of ACF rather than the frequency-domain analysis as reported by *Secker-Walker* and *Searle* in 1990 [10.26]. Cariani and Delgutte in 1996 showed that pooled inter-spike interval distributions resemble the short-time or the running autocorrelation function (ACF) for the low-frequency component. In addition, pooled interval distributions for sound stimuli consisting of the high-frequency component resemble the envelope for running ACF [10.27, 28].

Remarkably, primary sensations such as the pitch of the missing fundamental [10.29], loudness [10.30], and duration sensation [10.31] can be well described by the temporal factors extracted from the ACF [10.32, 33]. *Timbre* investigated by the dissimilarity judgment [10.16] of the sound field has been described by both the temporal and spatial factors. The typical spatial attributes of the sound field, such as subjective diffuseness [10.34] and apparent source width (ASW), as well as subjective preference, are well described by the spatial factor [10.32–36].

Besides the design of concert halls, other acoustical applications such as speech identification [10.36], environmental noise measurement [10.37], and sound localization in the median plane [10.38] should benefit from guidelines derived from this model.

10.1 Theory of Subjective Preference for the Sound Field

Subjective preference judgment is the most primitive response in any subjective attribute and entails judgments that steer an organism in the direction of maintaining and/or animating *life*. Subjective preference, therefore, may relate to an aesthetic issue. It is known that judgment in an absolute manner presents a problem in reliability; rather, data is judged in a relative manner such as by paired-comparison tests. This is the simplest method, in that any person may participate, and the resulting scale value may be utilized in the wide range of applications. From the results of subjective preference studies in relation to the temporal factor and the spatial factor of the sound field, the theory of subjective preference is derived. Examples of calculating subjective preference at each listener's position are demonstrated in Sect. 10.2 for the global listener and Sect. 10.3 for the individual listener. The relationship between the resulting scale value of subjective preference in an existing hall and the physical factors obtained by calculation using architectural plan drawings, has been examined by factor analysis in Sect. 10.4.

10.1.1 Sound Fields with a Single Reflection

Preferred Delay Time of a Single Reflection

First of all, the simplest sound field, which consists of the direct sound with the horizontal angle to a listener: $\xi_0 = 0°$ (the elevation angle $\eta_0 = 0°$), and a single reflection from a fixed direction $\xi_1 = 36°$ ($\eta_1 = 9°$), was investigated. These angles were selected since they are typical in a concert hall. The delay time Δt_1 of the reflection was adjusted in the range of 6–256 ms. The paired-comparison test was performed for all pairs in an anechoic chamber using normal hearing subjects with two different music motifs A and B (Table 10.1). The effective duration of the ACF of the sound source defined by τ_e may be obtained by the delay at which the envelope of the normalized autocorrelation function (ACF) becomes 0.1 (10-percentile delay) as shown in Fig. 10.1. The value of $(\tau_e)_{min}$ indicated in Table 10.1 is obtained from the minimum value of the running ACF, $2T$=2 s, with an interval of 100 ms. The recommended $2T$ is given by (10.14). As far as these source signals are concerned, values of (τ_e) extracted from the long-term ACF

Table 10.1 Music and speech source signals and their minimum effective duration of the running ACF, $(\tau_e)_{min}$

Sound source	Title	Composer or writer	$(\tau_e)_{min}$ (ms)
Music motif A	Royal Pavane	Orlando Gibbons	125
Music motif B	Sinfonietta, opus 48; IV movement	Malcolm Arnold	40
Speech S	Poem read by a female	Doppo Kunikida	10

[1] The value of $(\tau_e)_{min}$ is the minimum value extracted from the running ACF, $2T = 2$ s, with a running interval of 100 ms. The recommended value of $2T$ is given by (10.14)

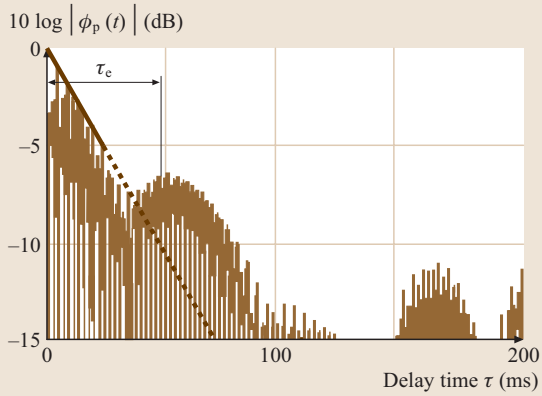


Fig. 10.1 Determination of the effective duration of the running ACF, τ_e. The value of τ_e may be obtained by the delay for which the envelope of the normalized ACF becomes 0.1 or −10 dB (10-percentile delay)

were similar to the minimum values from the running ACF, $(\tau_e)_{min}$.

For simplicity, the score was simply obtained, in this section, by accumulating a score giving +1 and −1 corresponding to positive and negative judgments, respectively, and the total score is divided by $S(F-1)$ to get the normalized score, where S is the number of subjects and F is the number of sound fields tested. The normalized scores for two motifs and the percentage of preference for speech signal as a function of the delay time are shown in Fig. 10.2.

Obviously, the most preferred delay time with the maximum score differs greatly between the two motifs. When the amplitude of reflection $A_1 = 1$, the most preferred delays are around 130 ms for music motif A, 35 ms for music motif B (Fig. 10.2a), and 16 ms for

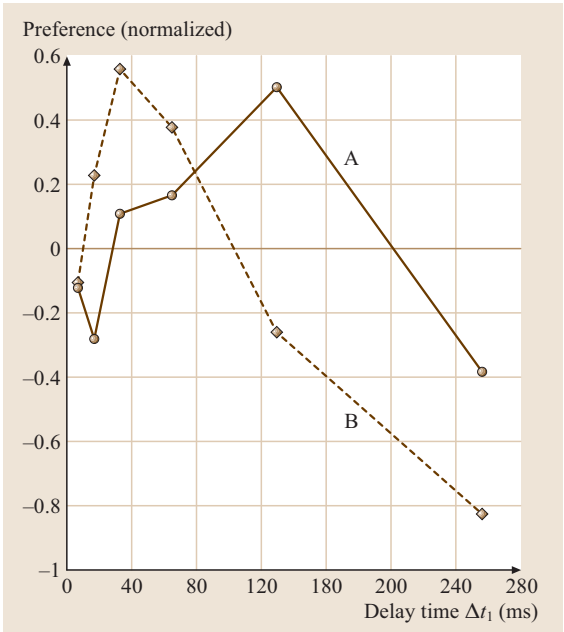


Fig. 10.2 Preference scores of the sound fields as a function of the delay time $A_1 = 1.0$ (six sound fields and 13 subjects [10.11]). Preference scores are directly obtained for each pair of sound fields by assigning the values +1 and −1, corresponding to positive and negative judgments, respectively. The normalized score is obtained by accumulating the scores for all sound fields (F) tested and all subjects (S), and then dividing by the factor $S(F-1)$ for 13 subjects. A: music motif A, Royal Pavane by Gibbons, $(\tau_e)_{min} = 125$ ms; B: music motif B, Sinfonietta, opus 48, III movement by Malcolm Arnold, $(\tau_e)_{min} = 40$ ms

speech [10.36]. Later, it was found that this corresponds well to the minimum values of the effective durations of the running ACF [10.39] of the source signals, so that the most preferred delay time were 125 ms (motif A), 40 ms (motif B) and 10 ms (speech). After inspection, the preferred delay is found roughly at certain durations of the ACF, defined by τ_p, such that the envelope of the ACF becomes $0.1A_1$. Thus, $[\Delta t_1]_p \approx \tau_e$ only when $A_1 = 1$. As shown in Fig. 10.3, changing the amplitude A_1, collected data of $[\Delta t_1]_p$ are expressed approximately by

$$[\Delta t_1]_p = \tau_p \approx (1 - \log_{10} A_1)(\tau_e)_{min} \,. \qquad (10.2)$$

Note that the amplitude of reflection relative to that of the direct sound should be measured by the most accurate method (ex. the square-root value of the ACF at the origin of the delay time).

Amazon EX1014 Page 368

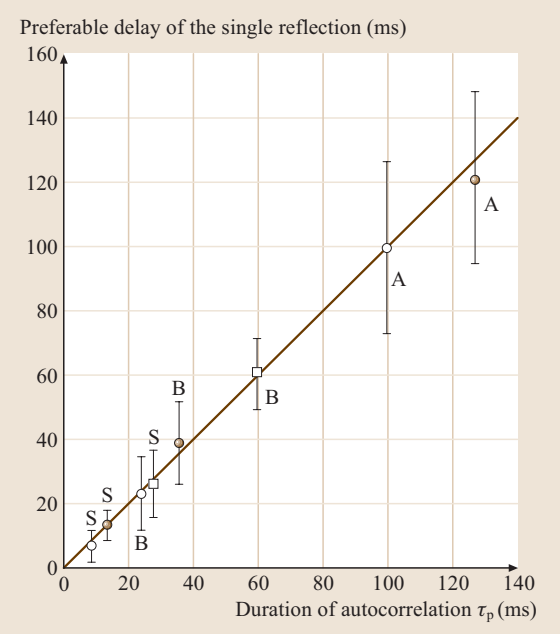


Fig. 10.3 The relationship between the preferred delay time of the single reflection and the duration of the ACF such that its envelope becomes 0.1 A_1 (for $k = 0.1$ and $c = 1.0$). The ranges of the preferred delays are graphically obtained at 0.1 below the maximum score. A, B and S refer to motif A, motif B, and speech, respectively. Different symbols indicate the center values obtained at the reflection amplitudes of +6 dB (○), 0 dB (●), and −6 dB (□), respectively (13–19 subjects). (After [10.11])

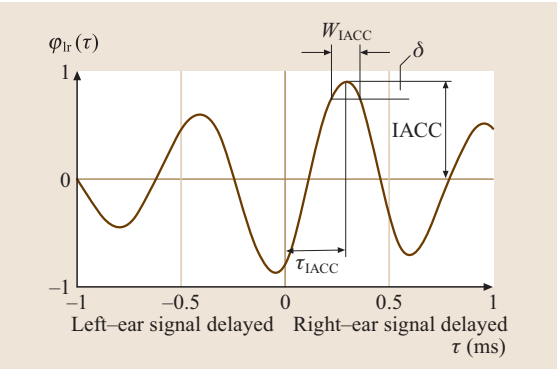


Fig. 10.4 Definition of the three spatial factors IACC, τ_{IACC} and W_{IACC}, extracted from the interaural cross-correlation function (IACF)

Two reasons can be given for why the preference decreases for the short delay range of reflection, $0 < \Delta t_1 < [\Delta t_1]_p$:

1. Tone coloration effects occur because of the interference phenomenon in the coherent time region; and
2. The IACC increases when Δt_1 is near 0. The definition of the IACC, which may be extracted from the IACF, as given by [10.1], is shown in Fig. 10.4.

On the other hand, echo disturbance effects can be observed when Δt_1 is greater than $[\Delta t_1]_p$.

Preferred Horizontal Direction of a Single Reflection to a Listener

The direction was specified by loudspeakers located at $\xi_1 = 0°$ ($\eta_1 = 27°$), and $\xi_1 = 18°, 36°, \ldots, 90°$ ($\eta_1 = 9°$), where the delay time of the reflection was fixed at 32 ms [10.11]. Results of the preference tests for the two motifs are shown in Fig. 10.5. No fundamental differences are observed between the curves of the sound source in spite of the large differences in the value of $(\tau_e)_{min}$. The preferred score increases roughly with decreasing IACC, the typical spatial factor. The correla-

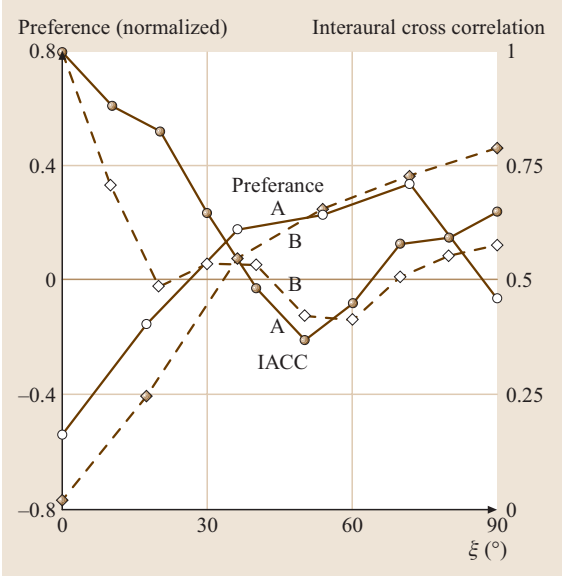


Fig. 10.5 Preference scores and the IACC of the sound fields with extreme music motifs A and B, as a function of the horizontal angle of a single reflection, $A_1 = 0$ dB (six sound fields and 13 subjects)

tion coefficient between the score and the IACC is -0.8 ($p < 0.01$). The score with motif A at $\xi_1 = 90°$ drops to a negative value, indicating that the lateral reflections coming only from around $\xi_1 = 90°$, thus, are not always preferred. The figure shows that there is a preference for angles less than $\xi_1 = 90°$, and on average there may be an optimum range centered on about $\xi_1 = 55°$. Similar results can be seen in the data from speech signals [10.40]. These results are due to the spatial factor independent of the temporal factor, which consists of the source signal.

10.1.2 Optimal Conditions Maximizing Subjective Preference

According to a systematic investigation of simulating the sound field with multiple reflections and reverberation by the aid of a computer and the listening test, the optimum design objectives and the linear scale value of subjective preference may be derived. The optimum design objectives can be described in terms of the subjectively preferred sound qualities, which are related to the four orthogonal factors describing the sound signals arriving at the two ears. They clearly lead to comprehensive criteria for achieving the optimal design of concert halls as summarized below [10.11–13].

Listening Level (LL)

The listening level is, of course, the primary criterion for listening to sound in a concert hall. The preferred listening level depends upon the music and the particular passage being performed. The preferred levels obtained with 16 subjects were similar for two extreme music motifs: 77–79 dBA in peak ranges for music motif A (Royal Pavane by Gibbons) with a slow tempo, and 79–80 dBA for music motif B (Sinfonietta by Arnold) with a fast tempo Fig. 10.7a.

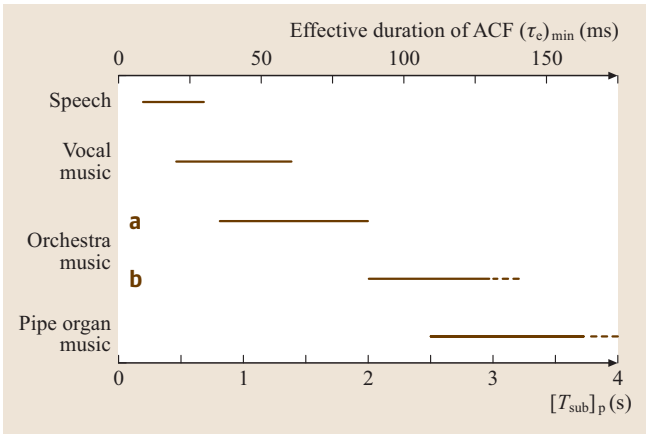


Fig. 10.6 Recommended reverberation time for several sound programs

Early Reflection after the Direct Sound (Δt_1)

An approximate relationship for the most preferred delay time has been discovered in terms of the ACF envelope of source signals and the total amplitude of reflections A. Generally, it is expressed by $[\Delta t_1]_p = \tau_p$

$$\left|\phi_p(\tau)\right|_{envelope} \approx kA^c, \text{ at } \tau = \tau_p\,, \tag{10.3}$$

where k and c are constants that depend on the subjective attributes [10.13, Fig. 42]. If the envelope of ACF is exponential, then

$$[\Delta t_1]_p = \tau_p \approx \left[\log_{10}(1/k) - c\log_{10} A\right](\tau_e)_{min}\,, \tag{10.4}$$

where the total pressure amplitude of reflection is given by

$$A = \left(A_1^2 + A_2^2 + A_3^2 + \ldots\right)^{1/2}\,. \tag{10.5}$$

The relationship given by (10.2) for a single reflection may be obtained by putting $A = A_1$, $k = 0.1$ and $c = 1$.

The value of $(\tau_e)_{min}$ is observed at the most active part of a piece of music containing artistic information such as a vibrato, a quick passage in the music flow, and/or a very sharp sound signal. Echo disturbance, therefore, may be perceived at $(\tau_e)_{min}$. Even for a long musical composition, the minimum part of $(\tau_e)_{min}$ of the running ACF in the whole music, which determines the preferred temporal condition, may be taken into consideration for the choice of music program to be performed in a given concert hall. A method of controlling the minimum value $(\tau_e)_{min}$ in performance, which determines the preferred temporal condition for vocal music has been discussed for blending the sound source and a given concert hall [10.41, 42]. If vibrato is introduced during singing, for example, it decreases $(\tau_e)_{min}$, blending the sound field with a short reverberation time.

Reverberation Time after the Early Reflection (T_{sub})

It has been observed that the most preferred frequency response to the reverberation time is a flat curve [10.39]. The preferred reverberation time, which is equivalent to that defined by *Sabine* [10.2], is expressed approxi-

mately by

$$[T_{sub}]_p \approx 23(\tau_e)_{min} \,. \qquad (10.6)$$

The total amplitudes of reflections A tested were 1.1 and 4.1, which cover the usual conditions of sound fields in a room. Recommended reverberation times for several sound sources are shown in Fig. 10.6. A lecture and conference room must be designed for speech, and an opera house mainly for vocal music but also for orchestra music. For orchestral music, there may be two or three types of concert hall designs according to the effective duration of the ACF. For example, symphony no. 41 by Mozart, *Le Sacre du Printemps* by Stravinsky, and Arnold's Sinfonietta have short ACFs and fit orchestra music of type (A). On the other hand, symphony no. 4 by Brahms and symphony no. 7 by Bruckner are typical of orchestra music (B). Much longer ACFs are typical for pipe-organ music, for example, by Bach.

The most preferred reverberation times for each sound source given by (10.6) might play important roles for the selection of music motifs to be performed. Of interest is that the most preferred reverberation time expressed by (10.6) implies about four times the *reverberation time* containing the source signal itself.

Magnitude of the Interaural Cross-Correlation Function (IACC)

All individual data indicated a negative correlation between the magnitude of the IACC and subjective preference, i. e., dissimilarity of signals arriving at the two ears is preferred. This holds only under the condition that the maximum value of the IACF is maintained at the origin of the time delay, keeping a balance of the sound field at the two ears. If not, then an image shift of the source may occur (Sect. 10.4.2). To obtain a small magnitude of the IACC in the most effective manner, the directions from which the early reflections arrive at the listener should be kept within a certain range of angles from the median plane centered on ±55°. It is obvious that the sound arriving from the median plane ±0° makes the IACC greater. Sound arriving from ±90° in the horizontal plane is not always advantageous, because the similar detour paths around the head to both ears cannot decrease the IACC effectively, particularly for frequency ranges higher than 500 Hz. For example, the most effective angles for the frequency ranges of 1 kHz and 2 kHz are centered on ±55° and ±36°, respectively. To realize this condition simultaneously, a geometrical uneven surface has been proposed [10.43].

10.1.3 Theory of Subjective Preference for the Sound Field

Theory

Since the number of orthogonal acoustic factors of the sound field, which are included in the sound signals

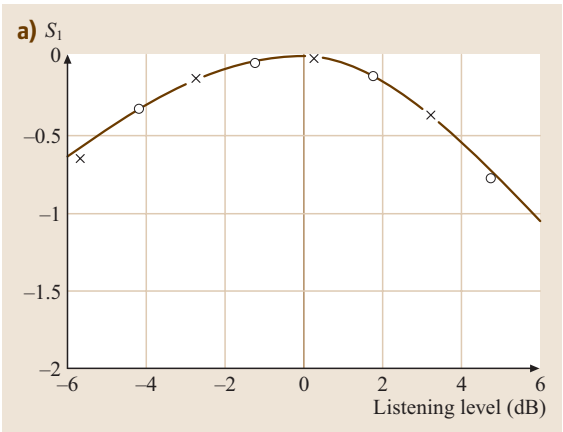


Fig. 10.7a–d Scale values of subjective preference obtained by the paired-comparison test for simulated sound fields in an anechoic chamber. Different symbols indicate scale values obtained from different source signals [10.12]. Even if different signals are used, a consistency of scale values as a function of each factor is observed, fitting a single curve. **(a)** As a function of the listening level, LL. The most preferred listening level, $[LL]_p = 0$ dB;

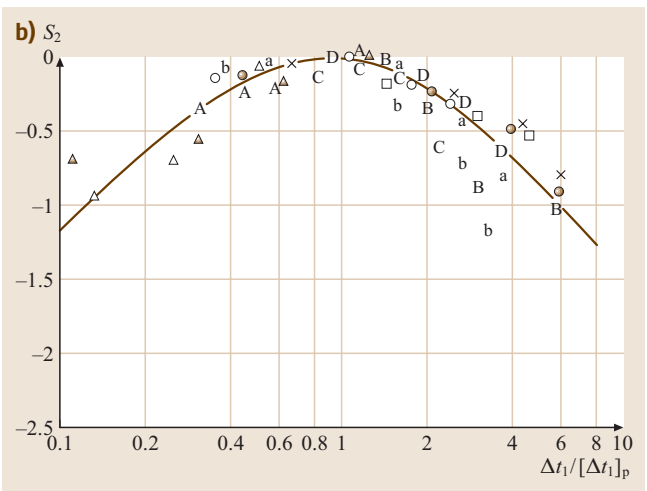


Fig. 10.7 (b) As a function of the normalized initial delay time of the first reflection by the most preferred delay time calculated by (10.4), $\Delta t_1/[\Delta t_1]_p$;

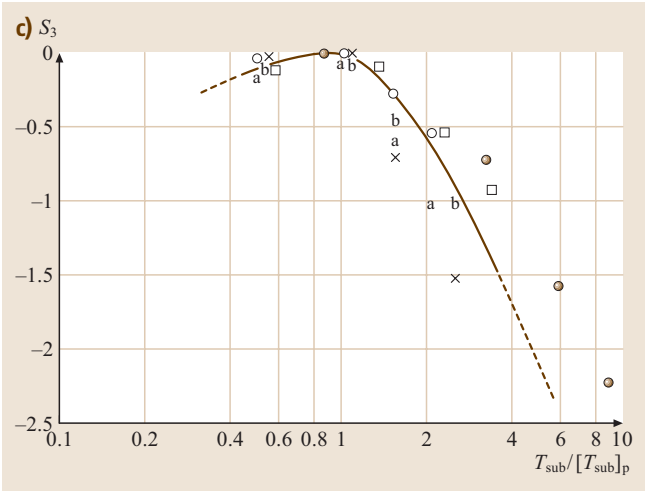


Fig. 10.7 (c) As a function of the normalized reverberation time for the most preferred calculated by (10.6), $T_{sub}/[T_{sub}]_p$;

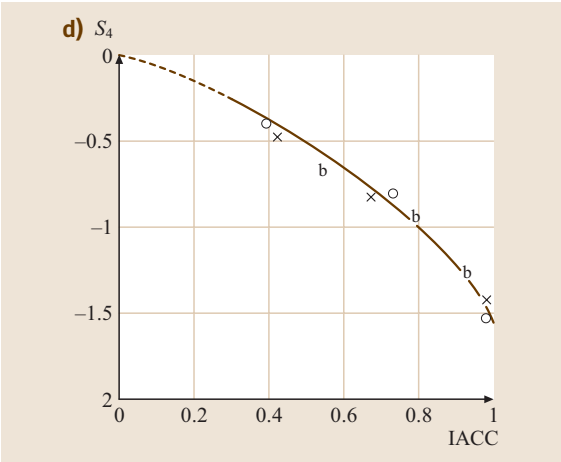


Fig. 10.7 (d) As a function of the IACC (After [10.12])

at both ears, is limited [10.12]; the scale value of any one-dimensional subjective response may be expressed by:

$$S = g(x_1, x_2, \ldots x_i) \,. \tag{10.7}$$

It has been verified by a series of experiments that four objective factors act independently on the scale value when changing two of the four factors simultaneously [10.13]. Results indicate that the units of the scale value of subjective preference derived by a series of experiments with different sound sources and different subjects have appeared to be constant [10.12], so that we may add scale values to obtain the total scale value such as

$$\begin{aligned} S &= g(x_1) + g(x_2) + g(x_3) + g(x_4) \\ &= S_1 + S_2 + S_3 + S_4 \,, \end{aligned} \tag{10.8}$$

where S_i $(i = 1, 2, 3, 4)$ is the scale value obtained relative to each objective parameter. Equation (10.8) indicates a four-dimensional continuity.

A Common Formula for the Four Normalized Orthogonal Factors

The dependence of the scale value on each objective parameter is shown graphically in Fig. 10.7. From the nature of the scale value, it is convenient to put a zero value at the most preferred conditions, as shown in this figure. These results of the scale value of subjective preference obtained from the different test series, using different music programs, yield the following common formula

$$S_i \approx -\alpha_i \, |x_i|^{3/2} \,, \quad i = 1, 2, 3, 4 \tag{10.9}$$

where values of α_i are weighting coefficients as listed in Table 10.2, which were obtained with a number of subjects. These coefficients depend on the individual. If α_i is close to zero, then a lesser contribution of the factor x_i on subjective preference is signified.

The factor x_1 is given by the sound pressure level (SPL) difference, measured by the A-weighted network, so that

$$x_1 = 20 \log P - 20 \log [P]_p \,, \tag{10.10}$$

P and $[P]_p$ being, respectively, the sound pressure at a specific seat and the most preferred sound pressure that may be assumed at a particular seat position in the

Table 10.2 Four orthogonal factors of the sound field and their weighting coefficients α_i obtained by a paired-comparison test of subjective preference with a number of subjects in conditions without any image shift of the source sound ($\tau_{IACC} = 0$)

i	x_i	α_i	
		$x_i > 0$	$x_i < 0$
1	$20 \log P - 20 \log [P]_p$ (dB)	0.07	0.04
2	$\log(\Delta t_1/[\Delta t_1]_p)$	1.42	1.11
3	$\log(T_{sub}/[T_{sub}]_p)$	$0.45 + 0.75A$	2.36–0.42 A
4	IACC	1.45	–

Amazon EX1014 Page 372

room under investigation:

$$x_2 = \log\left(\Delta t_1/[\Delta t_1]_p\right), \quad (10.11)$$

$$x_3 = \log\left(T_{sub}/[T_{sub}]_p\right), \quad (10.12)$$

$$x_4 = \text{IACC}. \quad (10.13)$$

Scale values of preference have been formulated approximately in terms of the 3/2 powers of the normalized objective parameters, expressed in the logarithm for the parameters, x_1, x_2 and x_3. Thus, scale values are not greatly changed in the neighborhood of the most preferred conditions, but decrease rapidly outside of this range. The remarkable fact is that the spatial binaural parameter x_4 is expressed in terms of the 3/2 powers of its real values, indicating a greater contribution than those of the temporal parameters.

Limitation of Theory

Since experiments were conducted to find the optimal conditions, this theory holds in the range of preferred conditions obtained by the test. In order to demonstrate the independence of the four orthogonal factors, under the conditions of fixed Δt_1 and T_{sub} around the preferred conditions, scale values of subjective preference calculated by (10.8) for the LL with (10.10) and the IACC with (10.13) with constants listed in Table 10.2 are shown in Fig. 10.8. Agreement between the calculated and observed values are satisfactory [10.13]. Even though both LL and the IACC are spatial factors, which are associated with the right cerebral hemisphere (Table 10.3), these are quite independent of each other. The same is true for the temporal factors of Δt_1 and T_{sub} associated with the left hemisphere. Of course, the spatial factor and the temporal factor are highly independent.

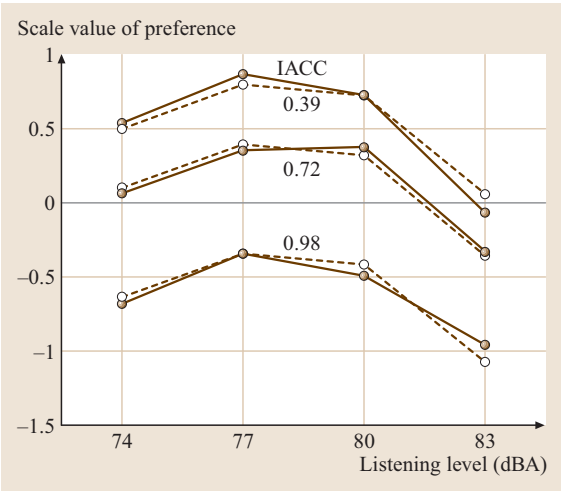


Fig. 10.8 Scale values of subjective preference for the sound field with music motif A as a function of the listening level and as a parameter of the IACC [10.13]. *Solid line*: calculated values based on (10.8) together with (10.9) taking the two factors (10.10) and (10.13) into consideration; *Dashed line*: measured values

Table 10.3 Hemispheric specializations determined by analyses of AEP (SVR), EEG and MEG[1]

Factors changed	AEP (SVR) $A(P_1 - N_1)$	EEG, ratio of ACF τ_e values of α-waves	MEG, ACF τ_e value of α-wave
Temporal			
Δt_1	L>R (speech)[2]	L>R (music)	L>R (speech)
T_{sub}	–	L>R (music)	–
Spatial			
LL	R>L (speech)	–	–
IACC	R>L (vowel /a/) R>L (band noise)	R>L (music)[3]	–

[1] See also [10.44] for a review of these investigations
[2] The sound source used in experiments is indicated in the bracket
[3] Flow of α-wave (EEG) from the right hemisphere to the left hemisphere for music stimuli when changing the IACC was observed by $|\phi(\tau)|_{max}$ between α-waves recorded at different electrodes [10.45]

Example of Calculating the Sound Quality at Each Seat

As a typical example, we shall discuss the quality of the sound field at each seating position in a concert hall with a shape similar to that of Symphony Hall in Boston. Suppose that a single source is located at the center, 1.2 m above the stage floor. Receiving points at a height of 1.1 m above the floor level correspond to the ear positions. Reflections with their amplitudes, delay times, and directions of arrival at the listeners are taken into account using the image method. Contour lines of the total scale value of preference calculated for music motif B are shown in Fig. 10.9. Results shown in Fig. 10.9b demonstrate the effects of the reflection from the sidewalls adjusted to the stage, which produce decreasing values of the IACC for the audience area. Thus the preference value at each seat is increased compared with that

Amazon EX1014 Page 373

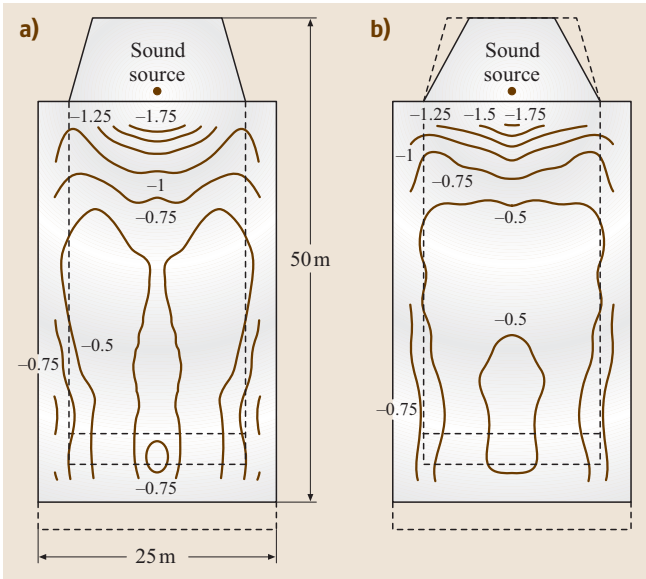


Fig. 10.9a,b An example of calculating the scale value with the four orthogonal factors using (10.8) through (10.13) with weighting coefficients (Table 10.2). **(a)** Contour lines of the total scale value for the Boston Symphony Hall, with original side reflectors on the stage. **(b)** Contour lines of the total scale values for the side reflectors optimized on the stage

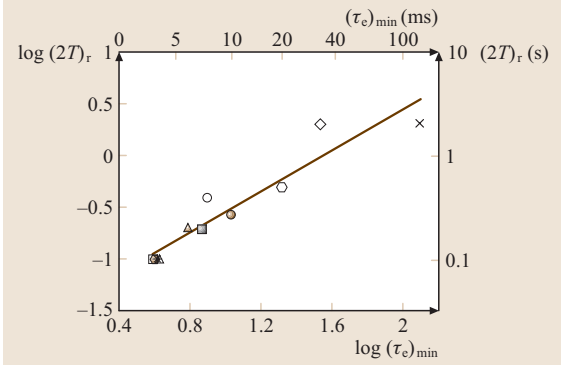


Fig. 10.10 Recommended temporal window $(2T)_r$ for the ACF proceeding as a function of the minimum value of effective duration of the ACF $(\tau_e)_{min}$. Different symbols represent experimental results using different sound sources

in Fig. 10.9a. In this calculation, the reverberation time is assumed to be 1.8 s throughout the hall and the most preferred listening level, $[LL]_p = 20\log[P]_p$ in (10.10), is set for a point on the center line 20 m from the source position.

10.1.4 Auditory Temporal Window for ACF and IACF Processing

Auditory Temporal Window for ACF Processing

In analyzing the running ACF, the so-called *auditory–temporal window* $2T$ must be carefully determined. The initial part of the ACF within the effective duration of the ACF contains important information about the source signal. In order to determine the auditory–temporal window, successive loudness judgments in pursuit of the running SPL have been conducted. Results shown in Fig. 10.10 indicate that the recommended signal duration $(2T)_r$ to be analyzed is approximately expressed by [10.46]

$$(2T)_r \approx 30(\tau_e)_{min}\,, \qquad (10.14)$$

where $(\tau_e)_{min}$ is the minimum effective duration, extracted from the running ACF [10.47]. This signifies an adaptive temporal window depending on the temporal characteristics of the sound signal in the auditory system. For example, the temporal window may differ according to music pieces $[(2T)_r = 0.5–5\,s]$ and to the vowels $[(2T)_r = 50–100\,ms]$ and consonants $[(2T)_r = 5–10\,ms]$ in continuous speech signals. It is worth noticing that the time constant represented by "fast" or "slow" of the sound level meter should be replaced by the temporal window, which depends on the effective duration of the ACF of the source signal. The running step (R_s), which signifies a degree of overlap of signal to be analyzed, is not critical. It may be selected as $K_2(2T)_r$; K_2 is chosen, say, in the range 0.25–0.5.

Auditory Temporal Window for IACF Processing

For the sound source fixed on the stage in a concert hall as usual, the value of $2T$ can be selected longer than 1.0 s for the measurement of the spatial factor at a fixed audience seat. But, when a sound signal is moving in the horizontal direction on the stage, we must know a suitable temporal window for $2T$ in analyzing the running IACF, which describes the moving image of sound localization. For a sound source moving sinusoidally in the horizontal plane with less than 0.2 Hz, $2T$ may be selected in a wide range from 30 to 1000 ms. If a sound source is moving below and/or at 4.0 Hz, $2T = 30–100$ ms is acceptable. In order to obtain a reliable result, it is recommended that such a temporal window for the IACF covering a wide range of movement velocity in the horizontal localization be fixed at about 30 ms.

Amazon EX1014 Page 374

10.1.5 Specialization of Cerebral Hemispheres for Temporal and Spatial Factors of the Sound Field

The independent influence of the aforementioned temporal and spatial factors on subjective preference judgments has been achieved by the specialization of the human cerebral hemispheres [10.44]. Recording over the left and right hemispheres of the slow vertex response (SVR) with latency of less than 500 ms, electroencephalograms (EEG) and magnetoencephalograms (MEG) have revealed various pieces of evidence (Table 10.3), with the most significant results being:

1. The left and right amplitudes of the evoked SVR, $A(P_1 - N_1)$ indicate that the left and right hemispheric dominance are due to temporal factors (Δt_1) and spatial factors (LL and IACC), respectively [10.48, 49].
2. Both left and right latencies of N_2 of SVR correspond well to the IACC [10.49].
3. Results of EEGs for the cerebral hemispheric specialization of the temporal factors, i. e., Δt_1 and T_{sub} indicated left-hemisphere dominance [10.50, 51], while the IACC indicated right-hemisphere dominance [10.45]. Thus, a high degree of independence between temporal and spatial factors was indicated.
4. The scale value of subjective preference is well described in relation to the value of τ_e extracted from the ACF of α-wave signals over the left hemisphere and the right hemisphere according to changes in temporal and spatial factors of the sound field, respectively [10.45, 50, 51].
5. Amplitudes of MEGs recorded when Δt_1 was changed reconfirm the left-hemisphere specialization [10.52].
6. The scale values of individual subjective preference relate directly to the value of τ_e extracted from the ACF of the α-wave of the MEG [10.52]. It is worth noting that the amplitudes of the α-wave in both the EEG and the MEG do not correspond well to the scale value of subjective preference.

In addition to the aforementioned activities in the time domain, in both the left and right hemispheres, spatial activity waves were analyzed by the cross-correlation function of alpha waves from the EEGs and MEGs. The results showed that a larger area of the brain is activated when the preferred sound field is presented [10.53] than when a less preferred one. This implies that the brain repeats a similar temporal rhythm in the α-wave range over a wider area of the scalp under the preferred conditions.

It has been reported that the left hemisphere is mainly associated with speech and time-sequential identifications, and the right is concerned with nonverbal and spatial identification [10.54, 55]. However, when the IACC was changed using speech and music signals, right-hemisphere dominance was observed, as indicated in Table 10.3. Therefore, hemispheric dominance is a relative response depending on which factor is changed in the comparison pair, and no absolute behavior could be observed.

To date, it has been discovered that the LL and the IACC are dominantly associated with the right cerebral hemisphere, and the temporal factors, Δt_1 and T_{sub} are associated with the left. This implies that such specialization of the human cerebral hemisphere may relate to the highly independent influence of spatial and temporal criteria on any subjective attribute. It is remarkable, for example, that *cocktail party effects* might well be explained by such specialization of the human brain, because speech is processed in the left hemisphere, and spatial information is processed in the right hemisphere independently.

10.2 Design Studies

Using the scale values in the four orthogonal factors of the sound field obtained by a number of listeners, the *principle of superposition* expressed by (10.8) together with (10.9) through (10.13) can be applied to calculate the scale value of preference for each seat. Comparison of the total preference values for different configurations of concert halls allows a designer to choose the best scheme. Temporal factors relating to its dimensions and the absorbing material on its walls are carefully determined according to the purpose of the hall in terms of a range of specific music programs (Fig. 10.6). In this section, we discuss mainly the spatial form of the hall, the magnitude of the interaural correlation function and the binaural listening level.

Amazon EX1014 Page 375

10.2.1 Study of a Space-Form Design by Genetic Algorithms (GA)

A large number of concert halls have been built since the time of the ancient Greeks, but only the halls with good sound quality are well liked. In order to increase the measure of success, a genetic algorithm (GA) system [10.56], a form of evolutionary computing, can be applied to the acoustic design of concert halls [10.57,58]. The GA system is applied here to generate the alternative scheme on the left-hand side of Fig. 10.11 for listeners.

Procedure

In this calculation, linear scale values of subjective preference S_1 and S_4 given by (10.9) are employed as fitting functions due to the LL and IACC, because the geometrical shape of a hall is directly affected by these spatial factors. The spatial factor for a source on the stage was calculated at a number of seating positions. For the sake of convenience, the single omnidirectional source was assumed to be at the center of the stage, 1.5 m above the stage floor. The receiving points that correspond to the ear positions were 1.1 m above the floor of the hall. The image method was used to determine the amplitudes, delay times, and directions of arrival of reflections at these receiving points. Reflections were calculated up to the second order. In fact, there was no change in the relative relationship among the factors obtained from calculations performed up to the second, third, and fourth order of reflection. The averaged values of the IACC for five music motifs (motifs A through E [10.13]) were used for the calculation.

Those hall shapes that produced greater scale values are selected as parent chromosomes. An example of the encoding of the chromosome is given in Fig. 10.12. The first bit indicated the direction of motion for the vertex.

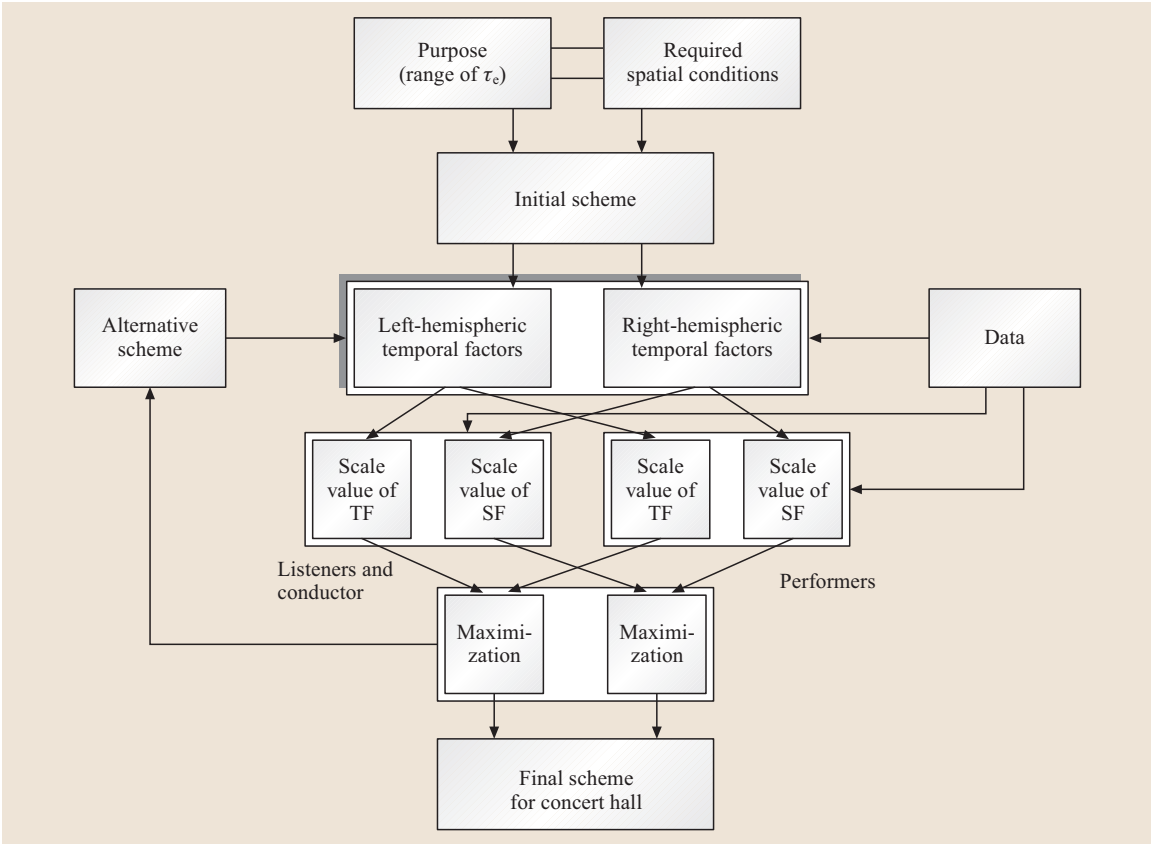


Fig. 10.11 The procedure for designing the sound field in a concert hall maximizing the scale values of subjective preference for a number of listeners (including a conductor) and performers. Data for global values of subjective preference may be utilized when designing a public hall

Amazon EX1014 Page 376

The other $(n-1)$ bits indicated the range over which the vertex moved. To create a new generation, the room shapes are modified and the corresponding movement of the vertices of the walls is encoded in chromosomes, i. e., binary strings. After GA operations that include crossover and mutation, new offspring are created. The fitness of the offspring is then evaluated in terms of the scale value of subjective preference. This process is repeated until the end condition of about 2000 generations is satisfied.

Shoe-Box Optimized

First of all, the proportions of the shoe-box hall were optimized (model 1). The initial geometry is shown in Fig. 10.13. In this form, the hall was 20 m wide, the stage was 12 m deep, the room was 30 m long, and the ceiling was 15 m above the floor. The point source was located at the center of the stage and 4.0 m from the front of the stage; 72 listening positions were selected. The range of motion for each sidewall and the ceiling was ±5 m from the respective initial positions, and the distance through which each was moved was coded on the chromosome of the GA. Scale values at the listening positions other than those within 1 m of the sidewalls were included in the averages ($\overline{S_1}$ and $\overline{S_4}$). These values were employed as the measure of fitness. In this calculation, the most preferred listening level $[LL]_p$ was chosen at the frontal seat near the stage. Results of optimization of the hall for $\overline{S_1}$ and $\overline{S_4}$ are shown in Fig. 10.14a and Fig. 10.14b, respectively. The width and length were almost the same in the two results, but the indicated heights were quite different. The height of the ceiling that maximizes $\overline{S_1}$ was as low as possible within the allowed range of motion to obtain a constant LL (Fig. 10.14a). The height that maximizes $\overline{S_4}$, on the other hand, was at the upper limit of the allowed range of motion to obtain small values of the IACC (Fig. 10.14b). Table 10.4 shows the comparison of the proportions obtained here and those of the Grosser Musikvereinssaal, which is a typical example of an excellent concert hall. The length/width ratios are almost the same. The height/width ratio of

Table 10.4 Comparison of proportions for the optimized spatial form of *shoe-box* type and the *Grosser Musikvereinssaal*

	Length/width	Height/width
Optimized for $\overline{S_1}$ for the listening level	2.50	0.71
Optimized for $\overline{S_4}$ for the IACC	2.57	1.43
Grosser Musikvereinssaal	2.55	0.93

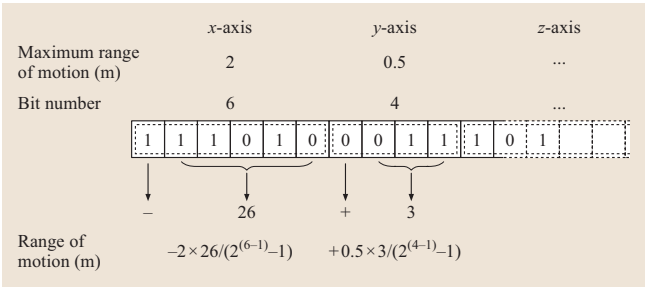


Fig. 10.12 An example of the binary strings used in encoding of the chromosome to represent modifications to the room shape

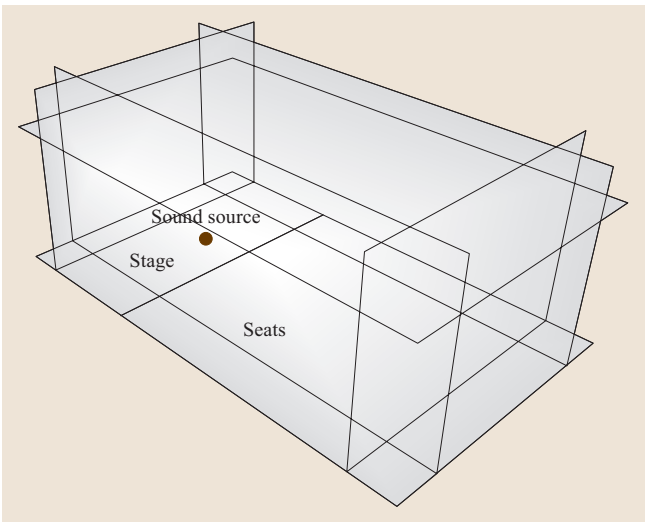


Fig. 10.13 The initial scheme of a concert hall (model 1). The range of sidewall and ceiling variation was ±5 m from the initial scheme

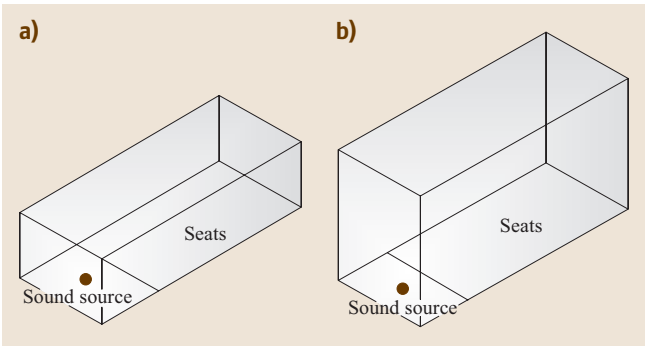


Fig. 10.14a,b Results for the model 1. **(a)** Geometry optimized for $\overline{S_1}$. **(b)** Geometry optimized for $\overline{S_4}$. (After [10.57])

Amazon EX1014 Page 377

the Grosser Musikvereinsaal is intermediate between our results for the two factors. For the ceiling of the hall, the height that maximized $\overline{S_1}$ was the lowest within the allowed range of motion (Fig. 10.14a). This is due to the fact that more energy can be provided from the ceiling to the listening position throughout the seats. To maximize $\overline{S_4}$, on the other hand, the ceiling took on the maximum height in the possible range of motion (Fig. 10.14b). Reflections from the flat ceiling did not decrease the IACC, but those from the sidewalls did.

Modification from the *Shoe-Box*

Next, to obtain even better sound fields, a little more complicated form (model 2), as shown in Fig. 10.15, was examined. The floor plan optimized according to the above results was applied as a starting point. The hall in its initial form was 14 m wide, the stage was 9 m deep, the room was 27 m long, and the ceiling was 15 m above the stage floor. The sound source was again 4.0 m from the front of the stage, but was 0.5 m to one side of the centerline and 1.5 m above the stage floor. The front and rear walls were vertically bisected to obtain two faces, and each stretch of wall along the side of the seating area was divided into four faces. The walls were kept vertical (i. e., tilting was not allowed) to examine only the plan of the hall in terms of maximizing $\overline{S_1}$ and $\overline{S_4}$. Each wall was moved independently of the other walls. The openings between the walls, in the acoustical simulation using the image method, were assumed not to reflect the sound. Forty-nine listening positions distributing throughout the seating area on a 2 m×4 m grid were selected. In the GA operation, the sidewalls were moved, so that none of these 49 listening positions were excluded. The moving range of each vertex was ±2 m in the direction of the line normal to the surface. The coordinates of the two bottom vertices of each surface were encoded on the chromosomes for the GA. In this calculation, the most preferred listening level was set for a point on the hall's long axis (central line), 10 m from the source position.

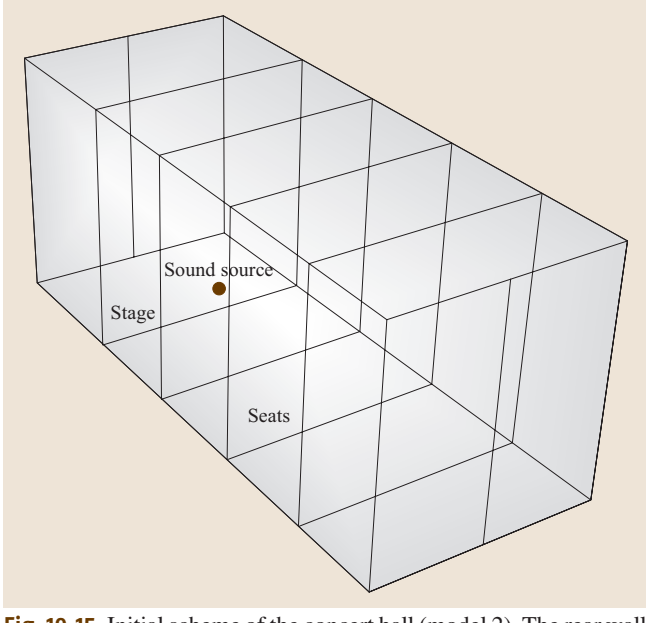


Fig. 10.15 Initial scheme of the concert hall (model 2). The rear wall of the stage and the rear wall of the audience area were divided into two. Sidewalls were divided into four

Leaf-Shape Concert Hall

The result of optimizing the hall for $\overline{S_1}$ is shown in Fig. 10.16 and the contour lines of equal $\overline{S_1}$ values are shown in Fig. 10.17. To maximize $\overline{S_1}$, the rear wall of the stage and the rear wall of the audience area took on concave shapes. The result of optimizing for $\overline{S_4}$ is shown in Fig. 10.18 and contour lines of equal $\overline{S_4}$ values are shown in Fig. 10.19. To maximize $\overline{S_4}$, on the other hand, the rear wall of the stage and the rear wall of the audience area took on convex shapes. With regard to the sidewalls, both $\overline{S_1}$ and $\overline{S_4}$ were maximized by the *leaf-shaped* plan, which is discussed in the following section.

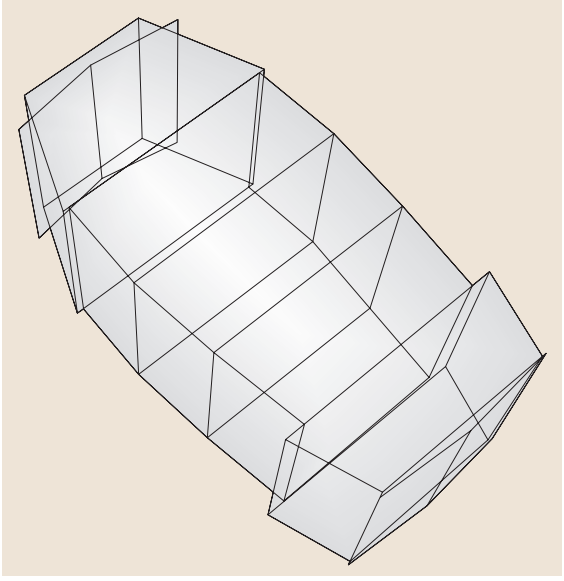


Fig. 10.16 A result for the model 2 optimized for $\overline{S_1}$

As for the conflicting requirements for S_1 and S_4, the maximization of S_4 may take priority over that of S_1, because the preference increases with decreasing IACC without exception [10.1], while there is a large individual difference in the preferred LL [10.59]. It is worth noting that listeners themselves can usually choose the best seat with respect to the preferred LL in a real concert hall.

A conductor and/or music director must be aware of the sound field characteristics of a concert hall. One is then able to select a program of music so that the sound is best in that hall in terms of the temporal factors (Fig. 10.6).

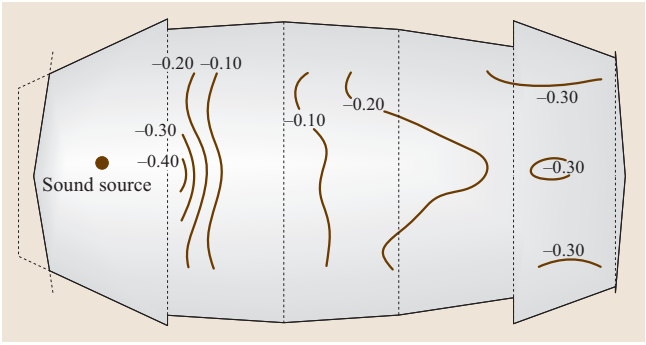


Fig. 10.17 Contour lines of equal $\overline{S_1}$ values calculated for the geometry shown in Fig. 10.16

10.2.2 Actual Design Studies

After testing more than 200 listeners, a small value of the IACC, which corresponds to different sound signals arriving at two ears, was demonstrated to be the preferred condition for individuals without exception. A practical application of this design theory was done in the Kirishima International Concert Hall (Miyama Conceru), which was characterized by a *leaf shape* (Fig. 10.20).

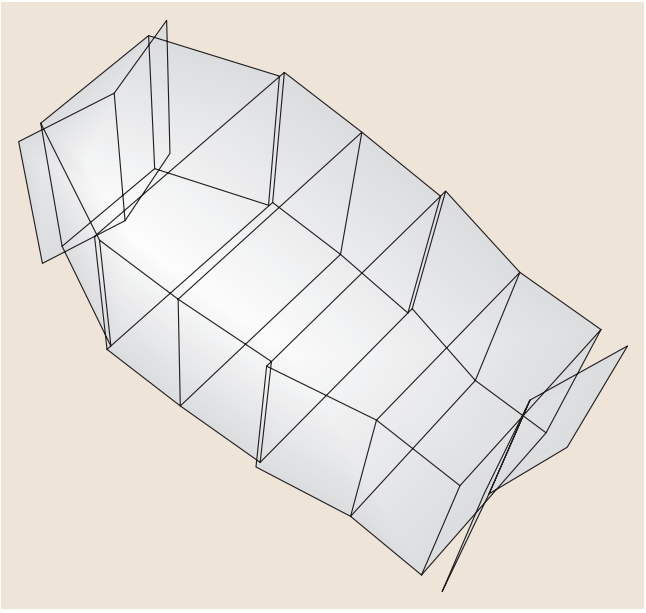


Fig. 10.18 A result for model 2 optimized for $\overline{S_4}$

Temporal Factors of the Sound Field for Listeners

When the space is designed for pipe-organ performance, the range of $(\tau_e)_{min}$, which may be selected to be centered on 200 ms, determines the typical temporal factor of the hall: $[T_{sub}]_p \approx 4.6$ s (10.6). When designed for the performance of chamber music, the range is selected to be near the value of 65 ms ($[T_{sub}]_p \approx 1.5$ s). The conductor and/or the sound coordinator select suitable musical motifs with a satisfactory range of effective duration of the ACF to achieve a music performance that blends the music and the sound field in a hall (Fig. 10.6). To adjust the preferred condition of Δt_1, on the other hand, since the value of $(\tau_e)_{min}$ for violins is usually shorter than that of contrabasses in the low-frequency range, the position of the violins is shifted closer to the left wall on the stage, and the position of the contrabasses is shifted closer to the center, as viewed from the audience.

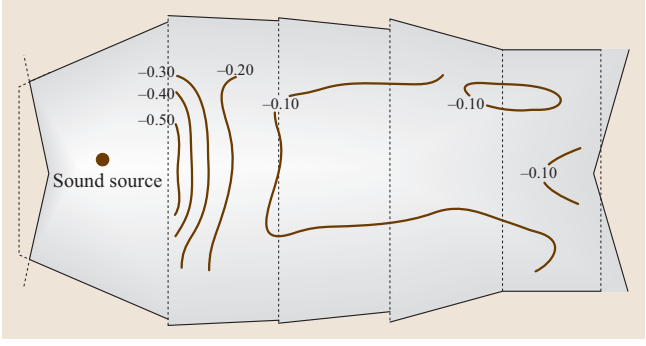


Fig. 10.19 Contour lines of equal $\overline{S_4}$ values calculated for the geometry shown in Fig. 10.18

Spatial Factors of the Sound Field for Listeners

The IACC should be kept as small as possible, maintaining $\tau_{IACC} = 0$. This is realized by suppressing the strong reflection from the ceiling, and by appropriate reflections from the sidewall at certain angles. When the source signal mainly contains frequency compo-

Amazon EX1014 Page 379

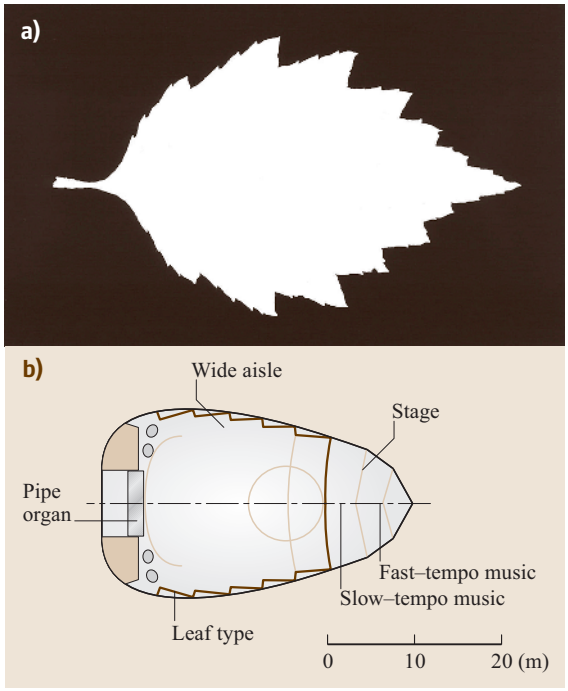


Fig. 10.20a,b A leaf shape for the plan proposed for the Kirishima International Concert Hall. **(a)** Original leaf shape. **(b)** Proposed shape for the plan. As usual, the sound field in *circled seating area* close to the stage must be carefully designed to obtain reflections from the walls on the stage and tilted sidewalls

nents around 1 kHz, the reflection from the side walls is adjusted to be centered roughly 55° to each listener, measured from the median plane. Under actual hearing conditions, the perceived IACC depends on whether or not the amplitudes of reflection exceed the hearing threshold level. This may be one of the reasons why a more diffuse sound field can be perceived with increasing power of the sound source. When the source is weak enough, that only the direct sound is heard, the actual IACC being processed in the auditory–brain system approaches unity, resulting in no diffuse sound impression. Thus, the IACC should be small enough with only strong early reflections.

Sound Field for Musicians

For music performers, the temporal factor is considered to be much more critical than the spatial factor (Sect. 10.3.2). Since musicians perform over a sequence of time, reflections with a suitable delay in terms of the value of $(\tau_e)_{min}$ of the source signals are of particular importance. Without any spatial subjective diffuseness, the preferred directions of reflections are in the median plane of music performers, resulting in IACC $\approx$ 1.0 [10.62, 63]. In order to satisfy these acoustic conditions, some design iterations are required, maximizing scale values for both musicians and listeners and leading to the final scheme of the concert hall as shown in Fig. 10.11.

Sound Field for the Conductor

It is recommended that the sound field for the conductor on the stage should be designed as that of a listener with appropriate reflections from the sidewalls on the stage [10.64].

Acoustic Design with Architects

From the historical viewpoint, architects have been more concerned with spatial criteria from the visual standpoint, and less so from the point of view of temporal criteria for blending human experience and the environment with design. On the other hand, acousticians

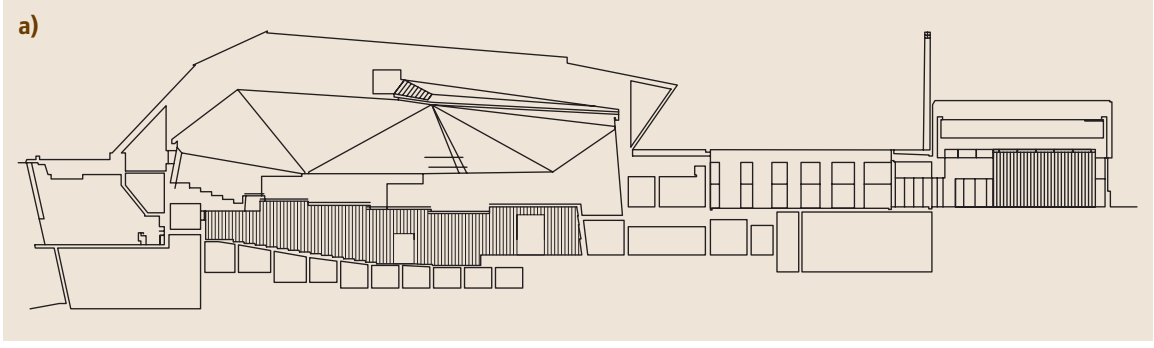


Fig. 10.21a–f Scheme of the Kirishima International Concert Hall, Kagoshima, Japan designed by the architect Maki and associates (1997) [10.60, 61].**(a)** Longitudinal section;

have mainly been concerned with temporal criteria, represented primarily by the reverberation time, from the time of *Sabine* [10.2] onward. No comprehensive theory of design including the spatial criterion as represented by the IACC existed before 1977, so that discussions between acousticians and architects were rarely on the same subject. As a matter of fact, both temporal and spatial factors are deeply interrelated with both acoustic design and architectural design [10.65, 66].

As an initial design sketch of the Kirishima International Concert Hall, a plan shape like a leaf (Fig. 10.20a) was presented at the first meeting for discussion with the architect Fumihiko Maki and associates with the explanation of the temporal and spatial factors of sound field.

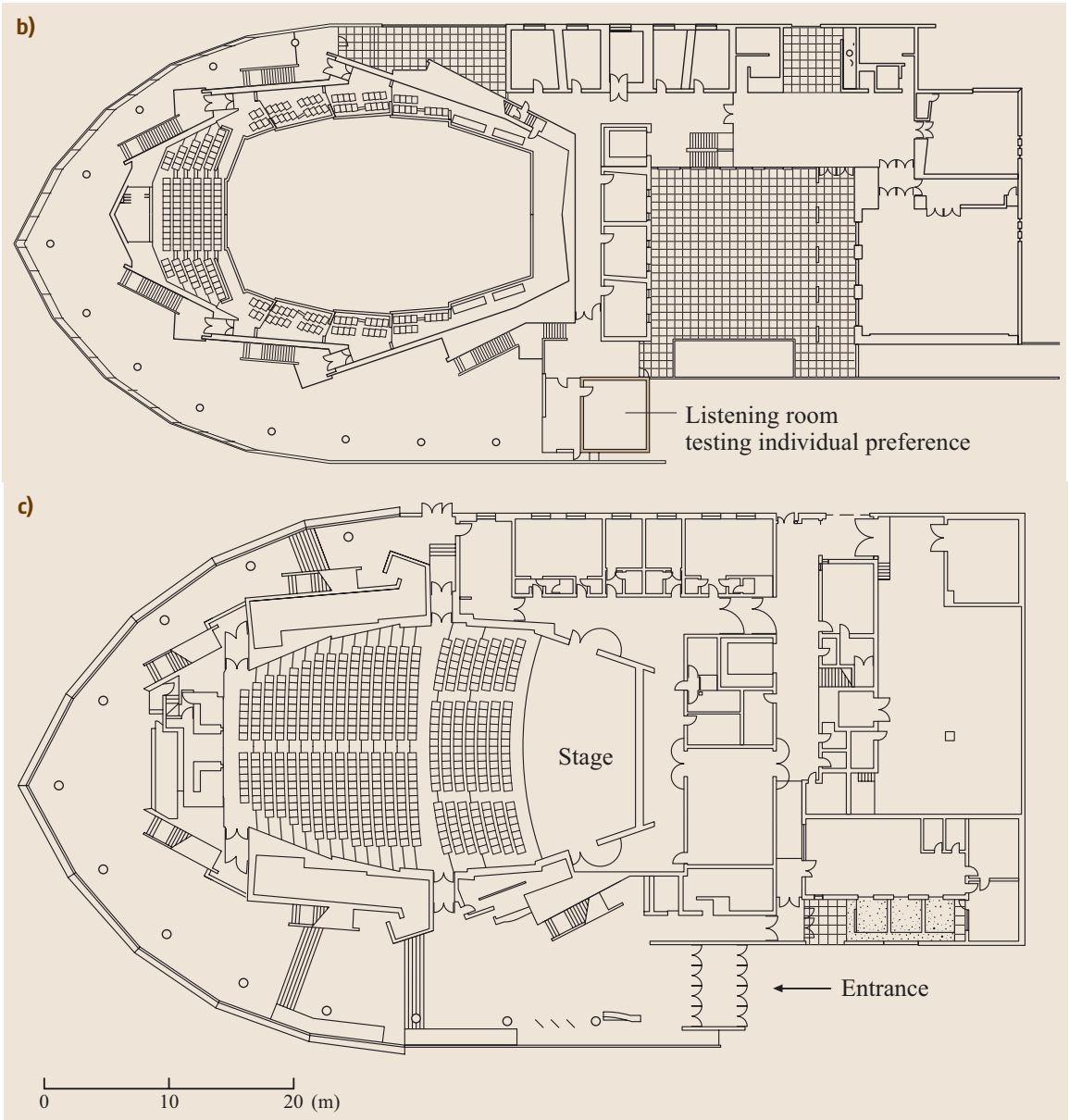


Fig. 10.21 (b) Plan of balcony level; (c) Plan of audience level

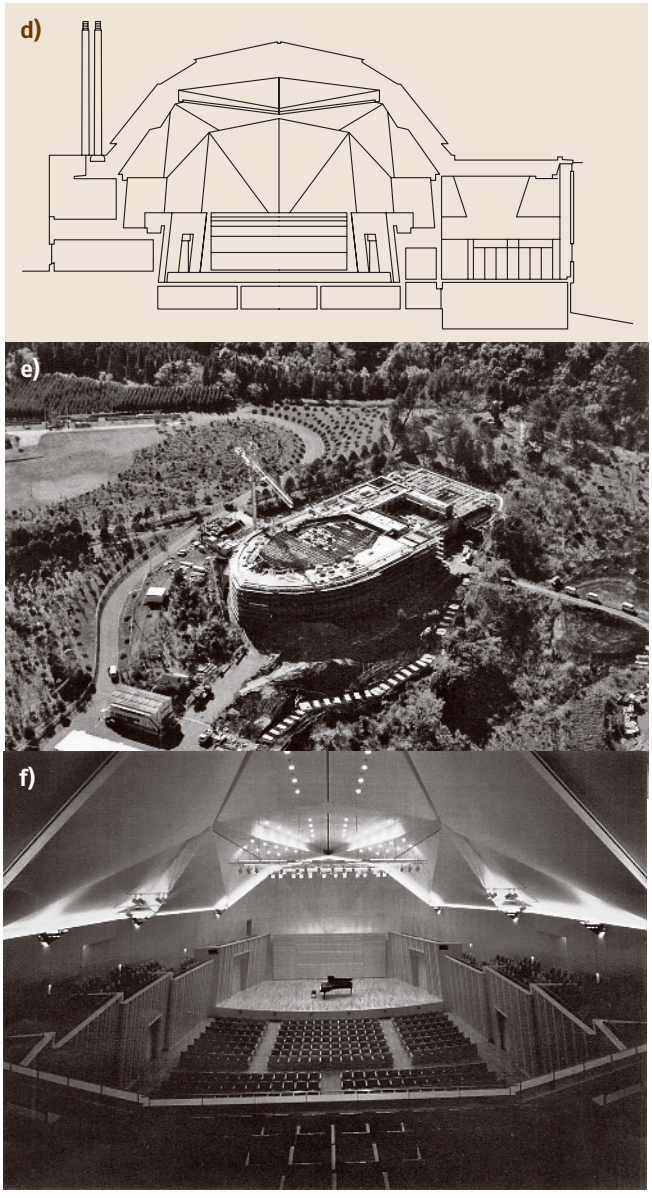


Fig. 10.21 (**d**) Cross section; (**e**) The Kirishima International Concert Hall, Kagoshima under construction. The leaf shape can be seen; (**f**) Tilt sidewalls and triangular ceilings after construction of the hall

After some weeks, Maki and Ikeda indicated a scheme of the concert hall as shown in Fig. 10.21 [10.60,61]. Without any change of plan and cross sections, the calculated results indicated the excellent sound field as shown in Fig. 10.22 [10.67,68].

The final architectural schemes, together with the special listening room for testing individual preference of sound field and selecting the appropriate seats for maximizing individual preference of the sound field, are shown in Fig. 10.21b. In these figures, the concert courtyard, the small concert hall, several rehearsal rooms and dressing rooms are also shown. The concert hall under construction is shown in Fig. 10.21e, in which the leaf shape may be seen; it was opened in 1994 (Fig. 10.21f).

Details of Acoustic Design

For listeners on the main floor. In order to obtain a small value of the IACC for most of the listeners, ceilings were designed using a number of triangular plates with adjusted angles, and the side walls were given a 10% tilt with respect to the main audience floor, as are shown in Fig. 10.21d and Fig. 10.23. In addition, diffusing elements were designed on the sidewalls to avoid the image shift of sound sources on the stage caused by the strong reflection in the high-frequency range above 2 kHz. These diffusers on the sidewalls were designed as a modification of the Schroeder diffuser [10.69] without the wells, as shown by the detail of Fig. 10.24.

For music performers on stage. In order to provide reflections from places near the median plane of each of the performers on the stage, the back wall on the stage is carefully designed as shown in the lower left in Fig. 10.23. The tilted back wall consists of six sub-walls with angles adjusted to provide appropriate reflections within the median plane of the performer. It is worth noting that the tilted sidewalls on the stage provide good reflections to the audience sitting close to the stage, at the same time resulting in a decrease of the IACC. Also, the sidewall on the stage of the left-hand side looking from the audience may provide useful reflections arriving from the back for a piano soloist.

Stage floor structure. For the purpose of suppressing the vibration [10.70] of the stage floor and anomalous sound radiation from the stage floor during a performance, the joists form triangles without any neighboring parallel structure, as shown in Fig. 10.25. The thickness of the floor is designed to be relatively thin (27 mm) in order to radiate sound effectively from the vibration of instruments such as the cello and contrabass. During rehearsal, music performers may control the radiation power somewhat by adjusting their position or by the use of a rubber pad between the floor and the instrument.

Amazon EX1014 Page 382

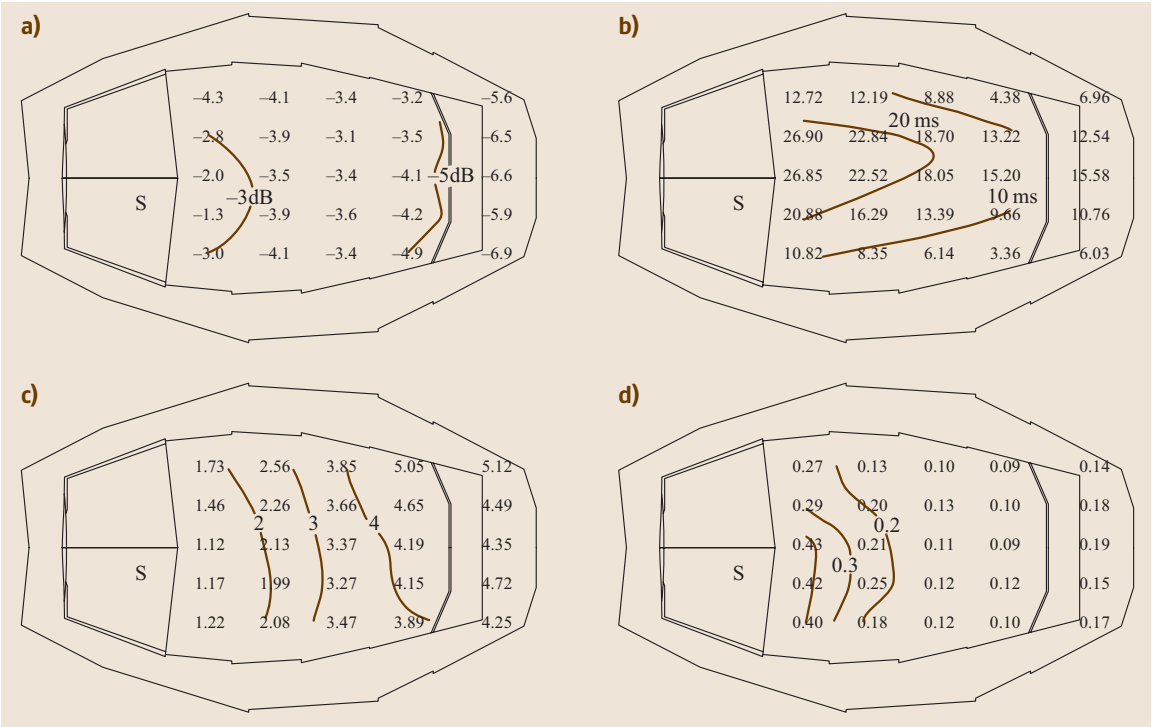


Fig. 10.22a–d Calculated orthogonal factors at each seat with a performing position *S* for an initial design of the hall. In the final design of the hall, the width was enlarged by about 1 m to increase the number of seats. The designed reverberation time was about 1.7 s for the 500 Hz band. **(a)** Relative listening level; **(b)** initial time delay gap between the direct sound and the first reflection Δt_1 [ms]; **(c)** *A*-value, the total amplitude of reflections; **(d)** IACC for white noise

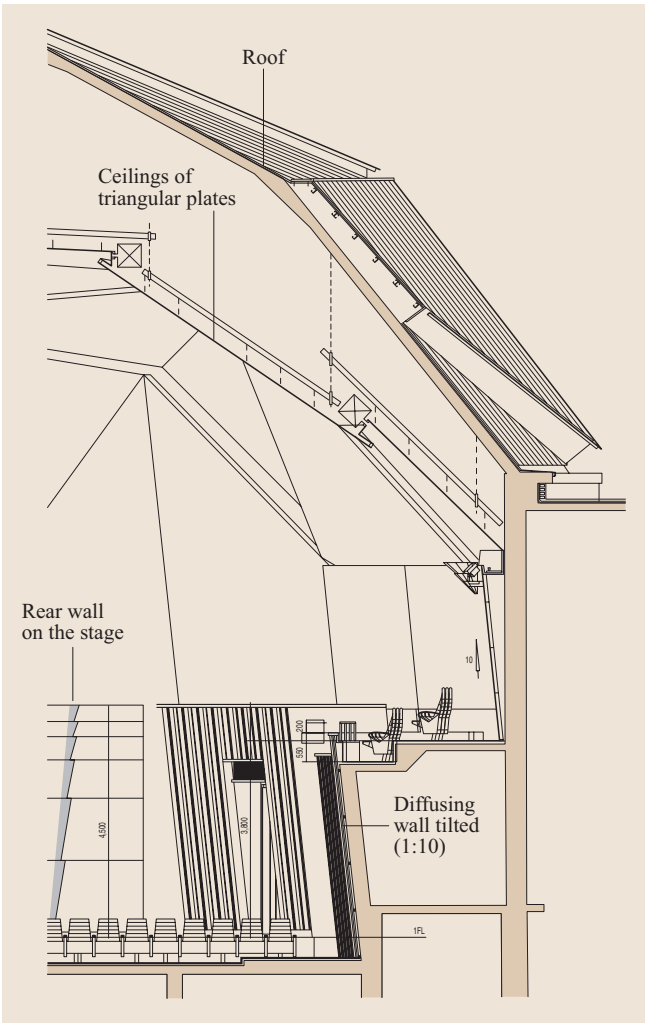


Fig. 10.23 Details of the cross section, including a sectional detail of the rear wall on the stage at the lower left of the figure

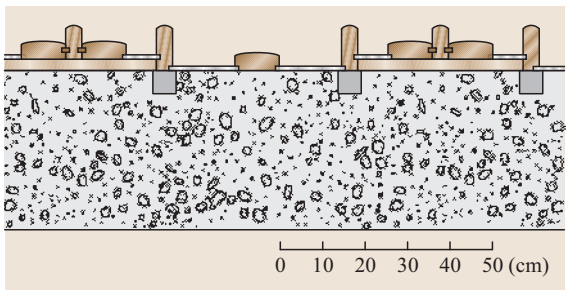


Fig. 10.24 Detail of the diffusing sidewalls effective for the higher-frequency range above 1.5 kHz, avoiding image shift of the sound source on the stage. The surface is deformed from the Schroeder diffuser by removal of the well partitions. (After [10.69])

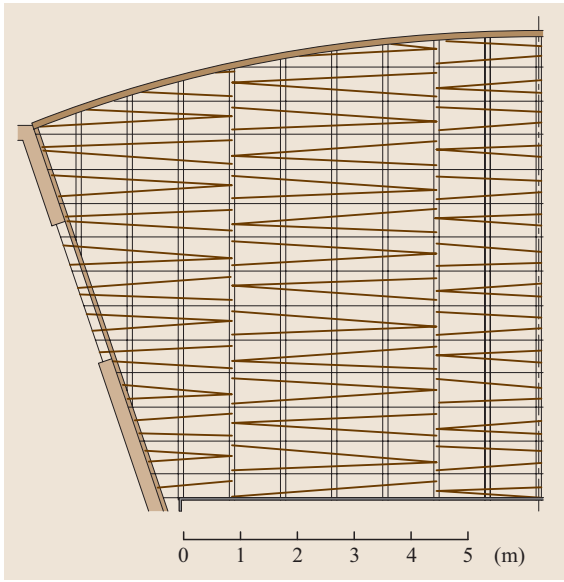


Fig. 10.25 Detail of the triangular joist arrangement for the stage floor, avoiding anomalous radiation due to the normal modes of vibration from certain music instruments that touch the floor

10.3 Individual Preferences of a Listener and a Performer

The minimum unit of society to be satisfied by the environment is one individual, which leads to a unique personal existence. Here, we demonstrate that the individual subjective preferences of each listener and each cellist may be described by the theory in Sect. 10.1, which resulted from observing a number of subjects.

10.3.1 Individual Subjective Preference of Each Listener

In order to enhance individual satisfaction for each listener, a special facility for seat selection, testing each listener's own subjective preference [10.59,71], was first

introduced at the Kirishima International Concert Hall in 1994. The sound simulation system employed multiple loudspeakers. It used arrows for testing the subjective preference of four listeners at the same time. Since the four orthogonal factors of the sound field influence the preference judgment almost independently [10.1], each single factor is varied, while the other three are fixed at the preferred condition for the average listener. Results of testing acousticians who participated in the International Symposium on Music and Concert Hall Acoustics (MCHA95), which was held in Kirishima in May 1995, are presented here [10.1].

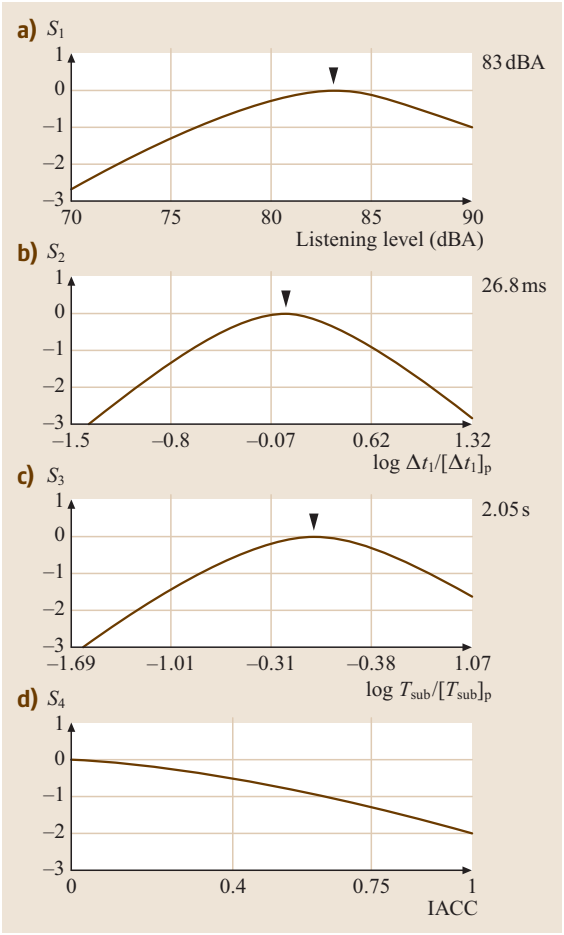


Fig. 10.26a–d Scale values of preference obtained by paired-comparison tests for the four orthogonal factors, subject BL. **(a)** The most preferred listening level was 83 dBA, the individual weighting coefficient in (10.9): $\alpha_1 = 0.06$; **(b)** the preferred initial time delay gap between the direct sound and first reflection was 26.8 ms, the individual weighting coefficient in (10.9): $\alpha_2 = 1.86$, where $[\Delta t_1]_p$ calculated by (10.4) with $(\tau_e)_{min}$ =62 ms for the music used ($A = 4$) is 24.8 ms; **(c)** the preferred subsequent reverberation time is 2.05 s, the individual weighting coefficient in (10.9): $\alpha_3 = 1.46$, where $[T_{sub}]_p$ calculated by (10.6) with $(\tau_e)_{min}$ =62 ms for the music used, is 1.43 s; **(d)** individual weighting coefficient in (10.9) for IACC: $\alpha_4 = 1.96$.

Individual Preference and Seat Selection

The music source was orchestral, the *Water Music* by Händel; the effective duration of the ACF was 62 ms [10.11]. The total number of listeners participating was 106. Typical examples of the test results for listener BL as a function of each factor are shown in Fig. 10.26. Scale values of this listener were rather close to the averages for subjects previously collected: the most preferred $[LL]_p$ was 83 dBA, the value

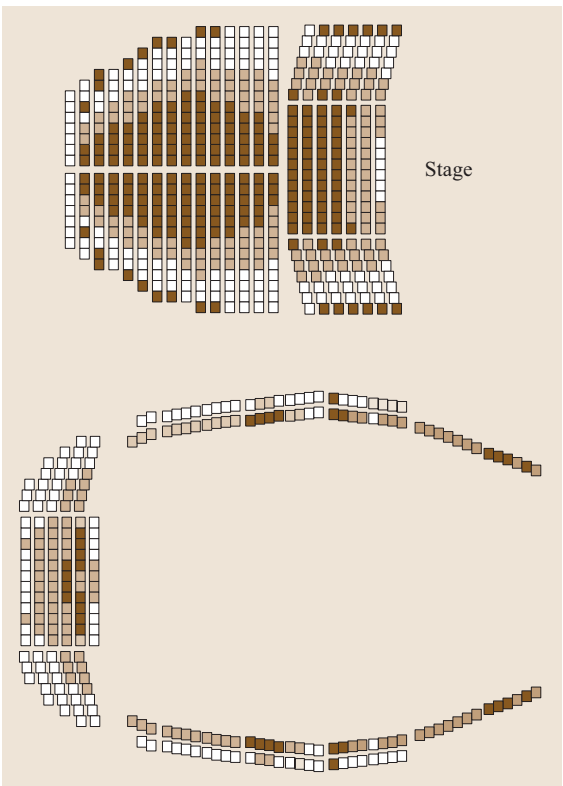


Fig. 10.27 Preferred seating area calculated for subject BL. The seats are classified in three parts according to the scale value of subjective preference calculated by the summation of S_1–S_4 (10.8) together with (10.9). Black indicates preferred seating areas, about one third of all seats in this concert hall, for subject BL

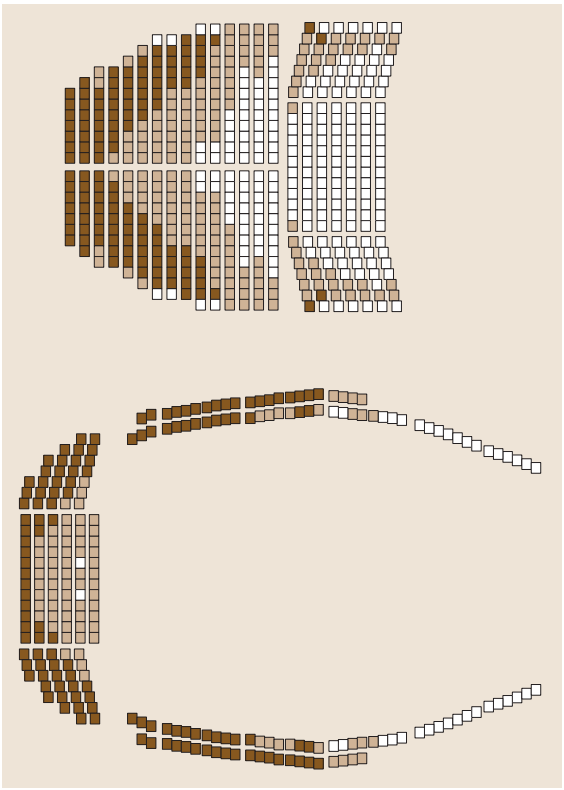


Fig. 10.28 Preferred seating area calculated for subject KH

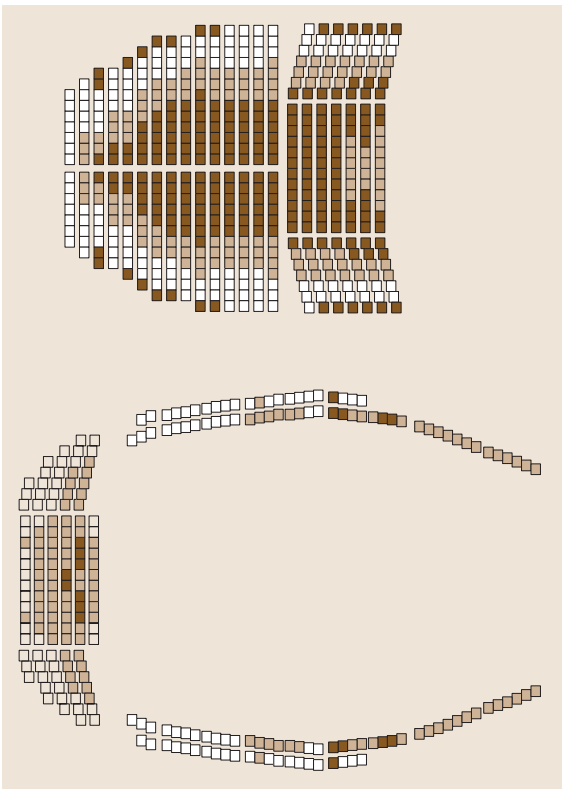


Fig. 10.29 Preferred seating area calculated for subject KK

$[\Delta t_1]_p \approx [(1 - \log_{10} A)(\tau_e)_{min}]$ was 26.8 ms (the global preferred value calculated by (10.4) with the total sound pressure as was simulated $A = 4.0$ is 24.8 ms. And the most preferred reverberation time is 2.05 s [the global preferred value calculated by (10.6) is 1.43 s]. Thus, as was designed, the center area of seats was preferred for listener BL (Fig. 10.27). With regard to the IACC, the result for all listeners was that the scale value of prefer-

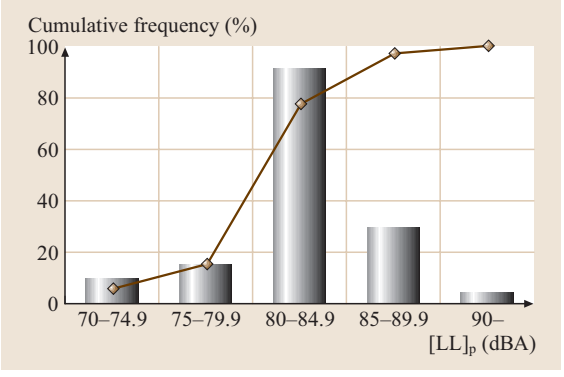


Fig. 10.30 Cumulative frequency of preferred listening level $[LL]_p$ (106 subjects). About 60% of subjects preferred the range of 80–84.9 dBA

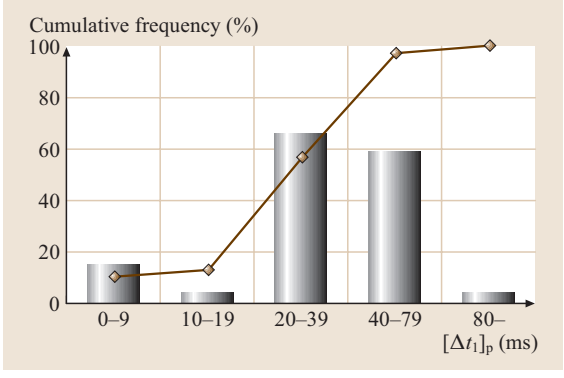


Fig. 10.31 Cumulative frequency of the preferred initial time delay gap between the direct sound and the first reflection $[\Delta t_1]_p$ (106 subjects). About 45% of subjects preferred the range of 20–39 ms. The value of $[\Delta t_1]_p$ calculated using (10.4) is 24.8 ms

Amazon EX1014 Page 386

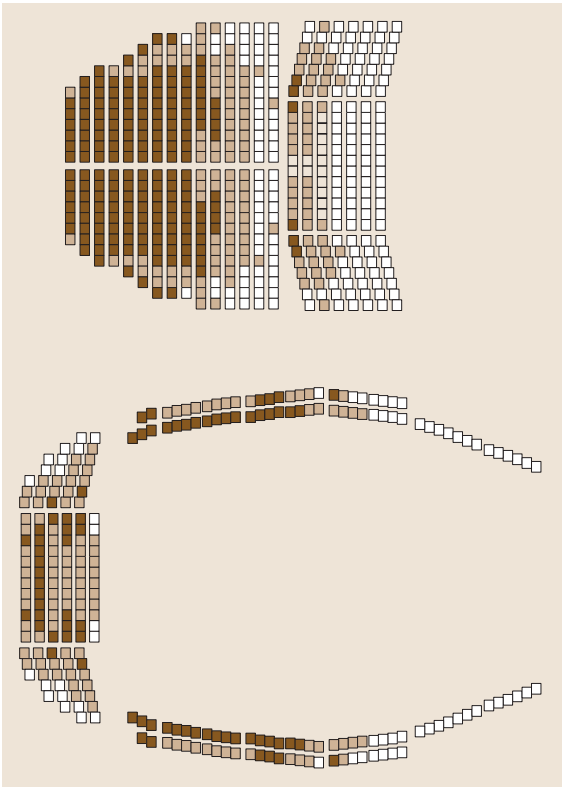


Fig. 10.32 Preferred seating area calculated for subject DP

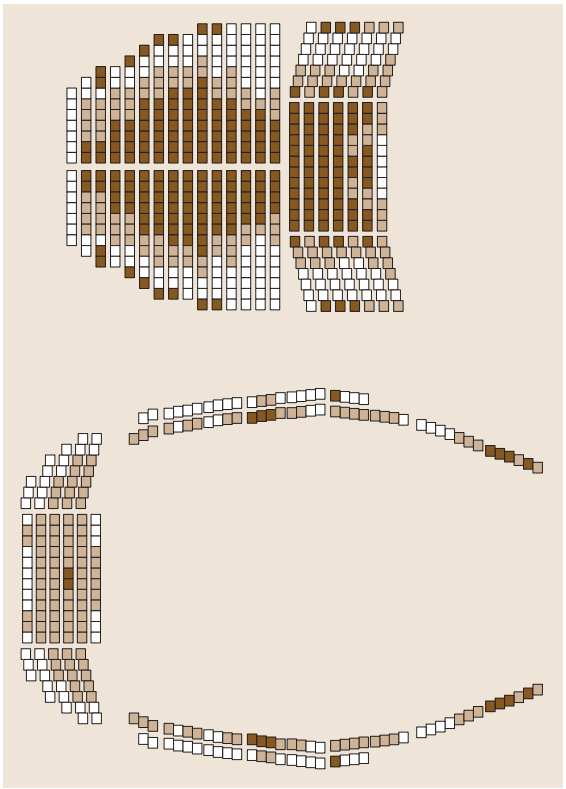


Fig. 10.33 Preferred seating area calculated for subject CA

ence increased with decreasing the IACC value. Since listener KH preferred a very short delay time of Δt_1, his preferred seats were located close to the boundary wall, as shown in Fig. 10.28. Listener KK indicated a preferred listening level exceeding 90 dBA. For this listener, the front seating area close to the stage was preferable, as shown in Fig. 10.29. For listener DP, whose preferred listening level was rather weak (76.0 dBA) and the preferred initial delay time was short (15.0 ms), the preferred seat was in the rear part of hall, as shown in Fig. 10.32. The preferred initial time delay gap for listener CA exceeds 100.0 ms, but was not critical. Thus, any initial delay times were acceptable, but the IACC was critical. Therefore, the preferred areas of seats were located as shown in Fig. 10.33.

Cumulative Frequency of Preferred Values

Cumulative frequencies of the preferred values with 106 listeners are shown in Fig. 10.30 through Fig. 10.34 for three factors. As indicated in Fig. 10.30, about 60% of listeners preferred the range 80–84.9 dBA when listening to music, but some listeners indicated that the most preferred LL was above 90 dBA, and the total range of the preferred LL was scattered, exceeding a 20 dB range. As shown in Fig. 10.31, about 45% of listeners preferred initial delay times of 20–39 ms, which were

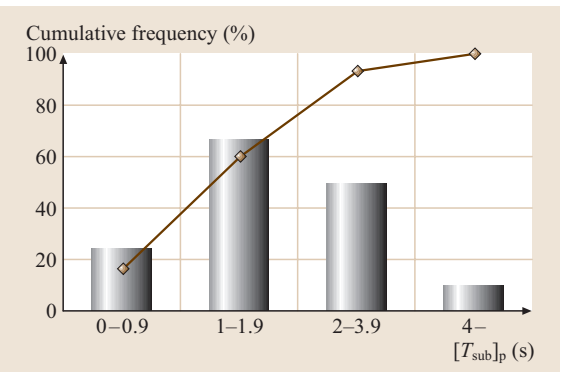


Fig. 10.34 Cumulative frequency of the preferred subsequent reverberation time $[T_{sub}]_p$ (106 subjects). About 45% of subjects preferred the range 1.0–1.9 s. The value of $[T_{sub}]_p$ calculated using (10.6) is 1.43 s

around the calculated preference of 24.8 ms (10.4) with $k = 0.1$, $c = 1$ and $A = 4.0$; however, some listeners indicated 0–9 ms and others more than 80 ms. With regard to the reverberation time, as shown in Fig. 10.34, about 45% of listeners preferred 1.0–1.9 s, which is centered on the preferred value of 1.43 s calculated by (10.6), but some listeners indicated preferences lower than 0.9 s or more than 4.0 s.

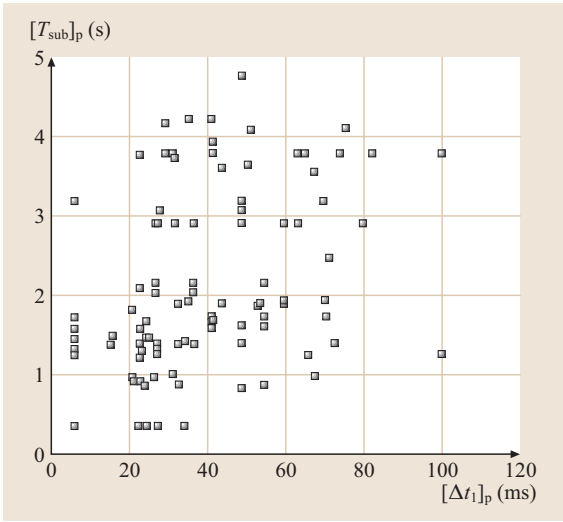


Fig. 10.35 The relationship between the preferred values of $[\Delta t_1]_p$ and $[T_{sub}]_p$ for each subject. No significant correlation between values was achieved

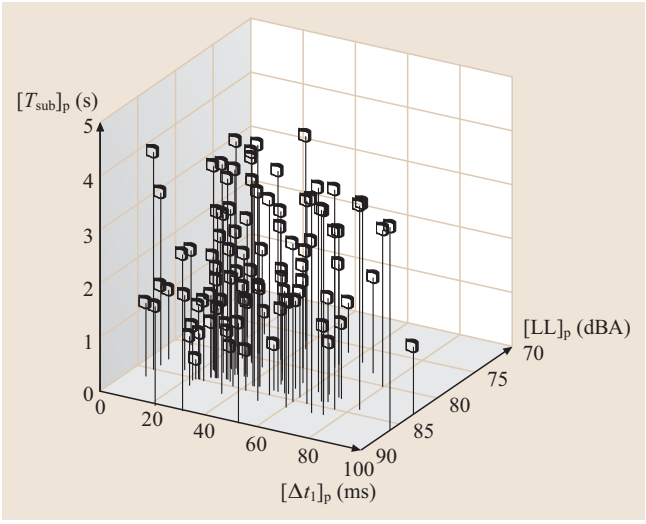


Fig. 10.36 Three-dimensional illustration of the three preferred orthogonal factors for the sound field for each individual subject (*squares*). All 106 listeners tested indicated that a smaller value of the IACC was preferred, and these data are not included in this figure. Preferred conditions are distributed in a certain range of each factor, so that subjects could not be classified into any specific groups

Independence of the Preferred Conditions

It was thought that both the initial delay time and the subsequent reverberation time appear to be mutually related, due to a kind of *liveness* of the sound field. Also, it was assumed that there is a strong interdependence between these factors for each individual. However, as shown in Fig. 10.35, there was little correlation between preference values of $[\Delta t_1]_p$ and $[T_{sub}]_p$ (the correlation is 0.06). The same is true for the correlation between values of $[T_{sub}]_p$ and $[LL]_p$ and for that between values of $[LL]_p$ and $[\Delta t_1]_p$, a correlation of less than 0.11. Figure 10.36 shows the three-dimensional plots of the preferred values of $[LL]_p$, $[\Delta t_1]_p$ and $[T_{sub}]_p$ for the 106 listeners. Looking at a continuous distribution in preferred values, no specific groupings of individuals can be classified from the data.

Another important fact is that there was no correlation between the weighting coefficients α_i and α_j, $i \neq j$ (i, $j = 1, 2, 3, 4$) in (10.9) for each individual listener [10.1]. A listener indicating a relatively small value of one factor will not always indicate a relatively small value for another factor. Thus, a listener can be critical about a preferred condition as a function of a certain factor, while insensitive to other factors, resulting in characteristic differences from other listeners.

10.3.2 Individual Subjective Preference of Each Cellist

To realize an excellent concert, we need to know the optimal conditions not only in the stage enclosure design for performers, but also in the style of the performance. The primary issue is that the stage enclosure is designed to provide a sound field in which performers can play easily. *Marshall* et al. [10.72] investigated the effects of stage size on the playing of an ensemble. The parameters related to stage size in their study were the delay time and the amplitude of reflections. *Gade* [10.73] performed a laboratory experiment to investigate the preferred conditions for the total amplitude of the reflections of performers. *Nakayama* [10.62] showed a relationship between the preferred delay time and the effective duration of the long-time ACF of the source signal for alto-recorder soloists (data was rearranged [10.1]). When we listen to a wide range of music signals containing a large fluctuation, it is more accurately expressed

Amazon EX1014 Page 388

by the minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of the source signals [10.39]. For individual singers, *Noson* et al. [10.74, 75] reported that the most preferred condition of the single reflection for an individual singer may be described by $(\tau_e)_{min}$ and a modified amplitude of reflection according to the overestimate and bone conduction effects (for control of $(\tau_e)_{min}$ see [10.41, 42]).

Preferred Delay Time of the Single Reflection for Cellists

As a good example, the preferred delay time of the single reflection for individual cello soloist is described by the minimum value of the effective duration of the running ACF of the music motifs played by that cellist [10.76]. The same music motifs (motifs I and II) used in the experiment by Nakayama were applied here [10.62]. The tempo of motif I was faster than that of motif II, as shown in Fig. 10.37. A microphone in front of the cellist picked up the music signals performed three times by each of five cellists.

Figure 10.38 shows an example of the regression curve for the scale value of preference fitted by (10.9). The peak of this curve denotes the most preferred delay time $[\Delta t_1]_p$. The values for individual cellists are listed in Table 10.5. Global and individual results (except for that of subject E) for music motif II were longer than those for music motif I.

The most preferred delay time of a single reflection is approximately expressed by the duration τ'_p of the ACF as similar to that of listeners (10.4), so that

$$[\Delta t_1]_p = \tau'_p \approx \left[\log_{10}(1/k') - c' \log_{10} A'\right](\tau_e)_{min} , \quad (10.15)$$

where the values k' and c' are constants that depend on the individual performer and musical instrument used. A substantial difference from (10.4) of listeners is that the amplitude of the reflection A' is defined by $A' = 1$ relative to -10 dB of the direct sound as measured at the ear's entrance. This is due to the phenomenon of *missing reflection* (i. e., a performer overestimating the reflection) [10.1].

Individual Subjective Preference

Using the quasi-Newton method, the resulting constants on average were $k' \approx 1/2$ and $c' \approx 1$ for the five cellists. (It is worth noting that the coefficients k' and c' for alto-recorder soloists were 2/3 and 1/4, respectively [10.1].) After setting $k' = 1/2$, the coefficient c' for each individual was figured out as listed in Table 10.6. The average value of the coefficient c' for the five cellists

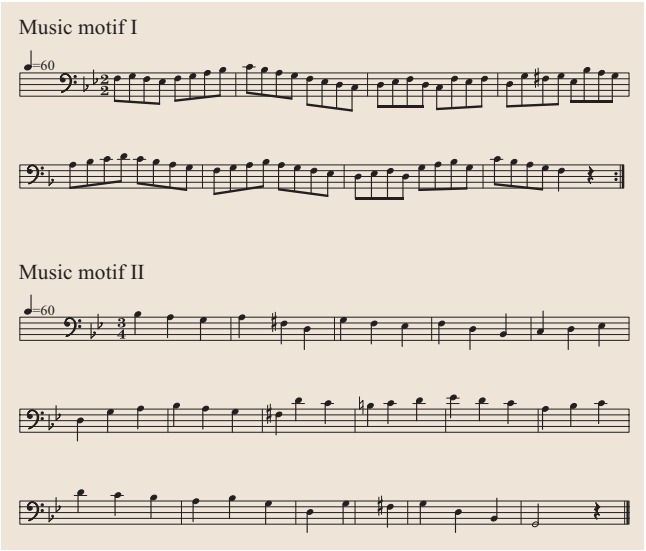


Fig. 10.37 Music motifs II and I composed by Tsuneko Okamoto for the investigation [10.62]

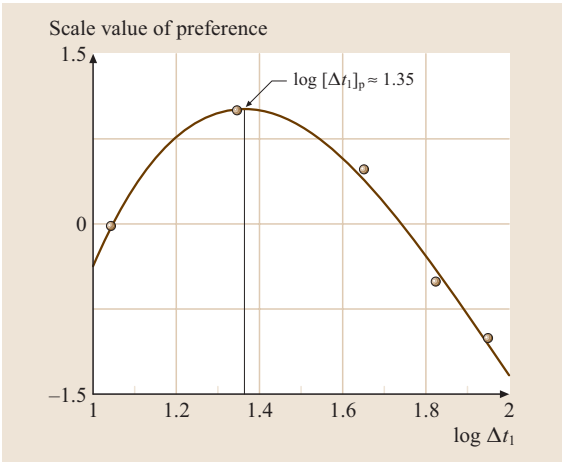


Fig. 10.38 An example of the regression curve for the preferred delay time (subject D, music motif I, −15 dB), $\log_{10}[\Delta t_1]_p \approx 1.35$, thus $[\Delta t_1]_p \approx 22.6$ ms

obtained was 1.03. The relationship between the most preferred delay time $[\Delta t_1]_p$ obtained by the judgment and the duration τ'_p $(= [\Delta t_1]_p)$ of the ACF calculated by (10.15) is shown in Fig. 10.40. Different symbols indicate values obtained in different test series with two music motifs. The correlation coefficient between calculated values of $[\Delta t_1]_p$ and measured values is 0.91 $(p < 0.01)$. The scale values of preference for each of the five cellists as a function of the delay time of the sin-

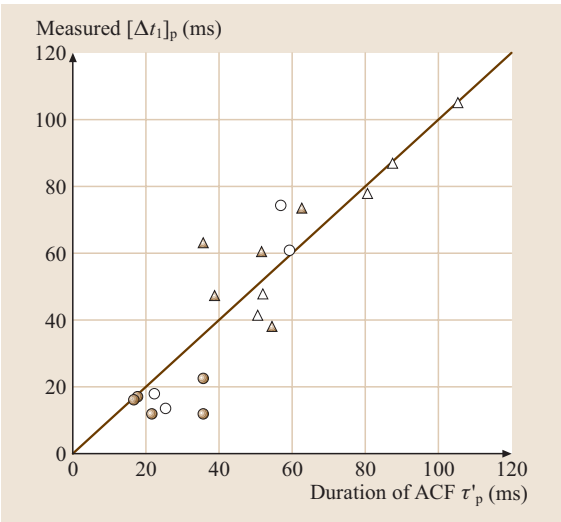


Fig. 10.39 The relationship between the most preferred delay time $[\Delta t_1]_p$ measured and the duration $[\Delta t_1]_p = \tau'_p$ calculated using (10.15). Correlation coefficient, $r = 0.91$ ($p < 0.01$). ●: music motif I, −15 dB; ○: music motif I, −21 dB; ▲: music motif II, −15 dB; △: music motif II, −21 dB

gle reflection normalized by the calculated $[\Delta t_1]_p$ are shown in Fig. 10.40.

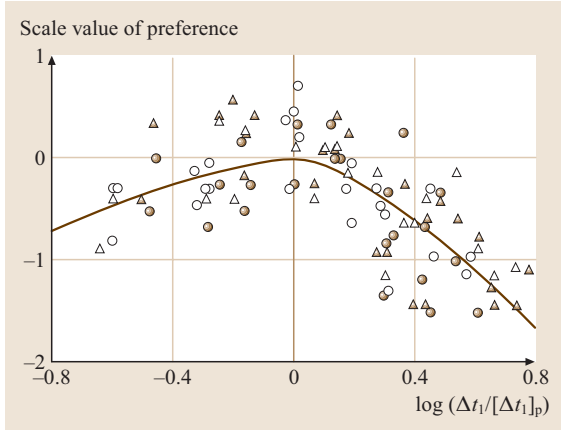


Fig. 10.40 Scale values of preference for each of five cellists as a function of the delay time of the single reflection normalized by its most preferred delay time calculated by (10.15). ●: music motif I, −15 dB; ○: music motif I, −21 dB; ▲: music motif II, −15 dB; △: music motif II, −21 dB. The regression curve is expressed by (10.9), $i = 2$

Table 10.5 Judged and calculated preferred delay times of a single reflection for each cello soloist. Calculated values of $[\Delta t_1]_p$ are obtained by (10.15) using the amplitude of the reflection A'_1 and $(\tau_e)_{min}$ for music signals performed by each cellist

A(dB)	A'(dB) (= $A+10$)	A'	Cellist	Judged $[\Delta t_1]_p$(ms)		Calculated $[\Delta t_1]_p$(ms)	
				Motif I	Motif II	Motif I	Motif II
−15	−5	0.56	A	16.2	47.9	16.3	38.5
			B	<12.0	73.8	35.2	62.7
			C	<12.0	60.8	211.3	51.3
			D	22.6	38.2	35.1	53.9
			E	17.6	63.6	17.3	35.2
			Global	18.0	48.3	24.3	47.5
−21	−11	0.28	A	18.1	48.4	21.8	51.5
			B	61.2	105.0	59.3	105.6
			C	–	77.9	–	80.6
			D	74.6	86.8	56.9	87.4
			E	<14.0	42.2	24.8	50.2
			Global	30.4	71.8	37.6	73.4

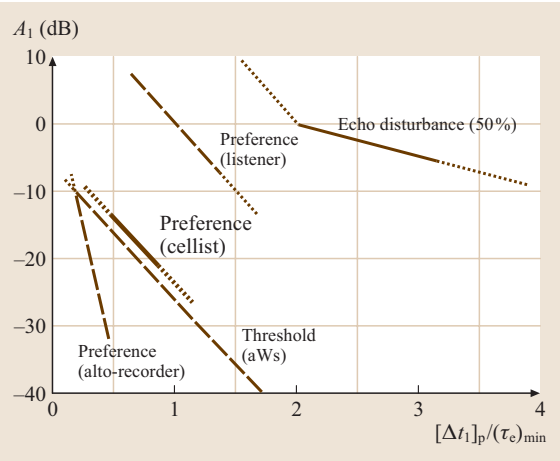
Table 10.6 The coefficient c' for each cellist in (10.15), calculating the preferred delay time of reflection for individual results and the global result, with $k' = 1/2$ (fixed)

	Cellist					Averaged
	A	B	C	D	E	(global)
Coefficient c'	0.47	1.61	1.10	1.30	0.67	≈ 1.03

Amazon EX1014 Page 390

Subjective Responses as a Function of $[\Delta t_1]_p/(\tau_e)_{min}$

Figure 10.41 shows the relative amplitude of the single reflection to that of the direct sound for the preference of cello soloists as a function of the delay time of the single reflection normalized by the minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of the source signal. Several other subjective responses in terms of the amplitude are shown together as a function of the delay time of the single reflection normalized by the value of the effective duration τ_e of the long-time ACF. All these values can generally be expressed by (10.4) in relation to the effective duration of the ACF with the constants k and c, which depend on different subjective responses. An alto-recorder soloist's preference is also plotted in this figure [10.1]. The values for performers are below or close to the threshold of perception of listeners [10.77]. These reconfirm the phenomenon of *missing reflection* for performers.



In order to blend the source music under performance and the sound field in a given concert hall, a performer, to some extent, can control the value of $(\tau_e)_{min}$ by introducing vibrato. Such an introduction of vibrato may decease the value of $(\tau_e)_{min}$ to obtain a more preferred condition for listeners as well, even though there is a short reverberation time in a given concert hall [10.41, 42].

Fig. 10.41 Relative amplitude of the single reflection for the subjective preference of cello soloists as a function of the delay time of a single reflection normalized by the value of $(\tau_e)_{min}$. Also, the amplitudes of several subjective responses as a function of the delay time of the single reflection normalized by the value of τ_e of the source signal. Note that threshold has been rearranged using the typical ACF of the speech signal [10.1]

10.4 Acoustical Measurements of the Sound Fields in Rooms

Acoustical measurements were made in an existing hall for the purpose of testing acoustic factors that were calculated using the architectural scheme at the design stage. Also, subjective preference judgments for different source locations on the stage were performed by the paired-comparison test at each set of seats. The relationship between the resulting scale values of subjective preference and the physical factors obtained by simulation using architectural plan drawings was examined by factor analysis. The accumulation and understanding of the field data, in turn, may improve details of future methods for calculating acoustic factors.

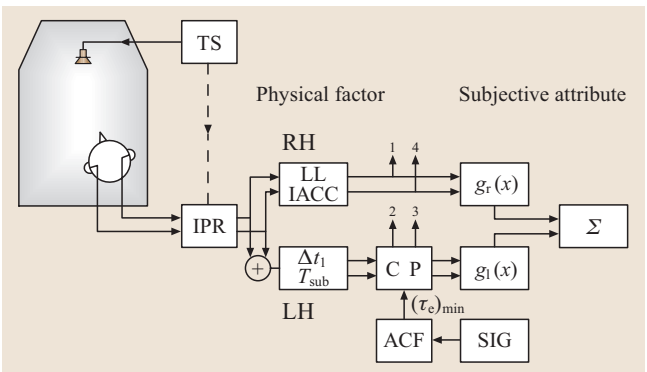


Fig. 10.42 A system for measuring the four orthogonal factors of sound fields and evaluating subjective qualities at each seat in a room. TS: test signal (maximum-length sequence signal); IPR: impulse response analyzer; RH: right-hemispheric factors (listening level and the IACC); LH: left-hemispheric factors (Δt_1, T_{sub}, and the A value); CP: comparators with the most preferred condition based on the effective duration of ACF, $(\tau_e)_{min}$; ACF: autocorrelation function; SIG: source signals; $g_r(x)$: scale values from the right-hemispheric factors; $g_l(x)$: scale values from the left-hemispheric factors; Σ: total scale value of subjective preference

Amazon EX1014 Page 391

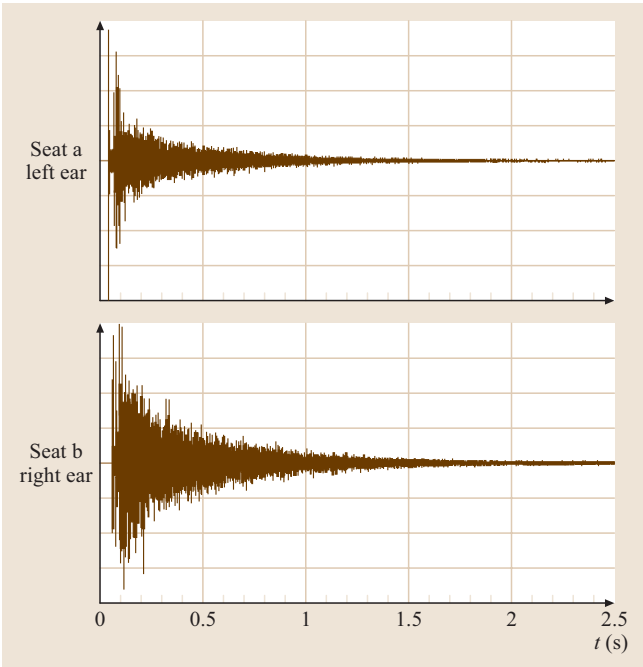
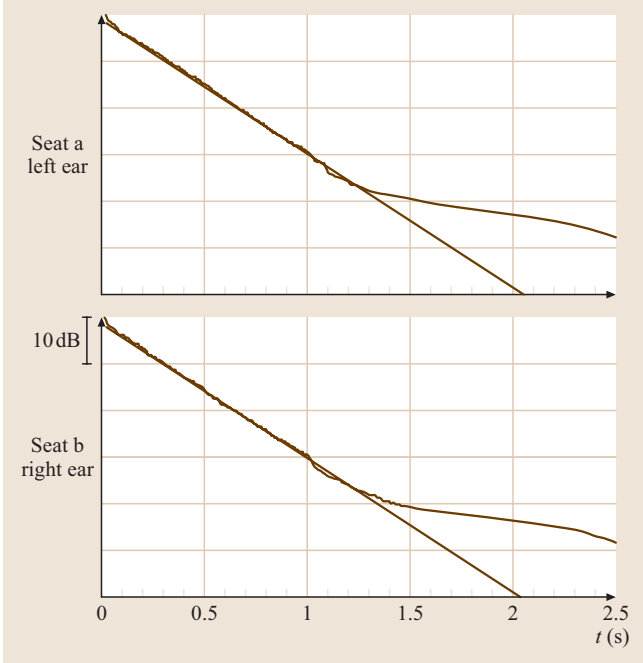


Fig. 10.43a,b Examples of the impulse response measured at seats: **(a)** (near to the stage) and **(b)** (far from the stage) in the Kirishima International Concert Hall. The amplitude of the impulse response measured at seat **(a)** is attenuated because of a strong direct sound.



10.4.1 Acoustic Test Techniques

Binaural Impulse Response

A diagnostic system for measuring the impulse response at the two ear entrances determining the four orthogonal factors, and for further evaluation of the subjective attributes of sound field at each seat in a hall is shown in Fig. 10.42. A test signal is radiated from the loudspeaker to measure impulse responses using two small microphones placed at the ear entrances of a real head (1.1 m above the floor). Then spatial factors associated with the right-hemisphere specialization (LL and IACC) and the temporal factors of left-hemisphere specialization (Δt_1 and T_{sub}) are analyzed. When the effective duration of the ACF of the source signal (τ_e) is calculated, the total scale value may be obtained by adding the scale values of the orthogonal factors referred to the most preferred conditions. The value of $(\tau_e)_{min}$ is used to determine the most preferred temporal values for $[\Delta t_1]_p$ and $[T_{sub}]_p$ (Sect. 10.1.2). If the source signal is fed into the ACF processor, then outputs (numbered 1–4) may be used to control the sound field simultaneously with an electroacoustic system, without any manual adjustment, preserving the preferred conditions of the four factors.

Examples of measuring binaural impulse responses at a seat close to the stage (seat a, left ear) and at a rear seat b (right ear) in the Kirishima International Concert Hall are demonstrated in Fig. 10.43. In this measurement, an omnidirectional dodecahedron loudspeaker with 12 full-range drivers was placed on the stage 1.5 m above the floor for a sound source. The total amplitude of reflections A at a seat close to the stage is usually smaller than that at seat far from the stage.

Another powerful signal to be radiated from the loudspeaker to measure the impulse binaural responses is the pulse signal generated by inverse Fourier transformation [10.78].

Reverberation Time

After the impulse response is obtained, the reverberation time is measured by *Schroeder's* method [10.79,80]. The integrated decay curve as a function of time may be obtained by squaring and integrating the impulse response of the sound field in a room, such that

$$\langle s^2(t)\rangle = \mathrm{K}\int_{t+T}^{t} h^2(x)\,\mathrm{d}x \qquad (10.16)$$

Fig. 10.44 Integrated decay curves obtained from the impulse responses at the two ear entrances, at seat **(a)** in the Kirishima International Concert Hall

Amazon EX1014 Page 392

where the time T should be chosen sufficiently longer than the reverberation time.

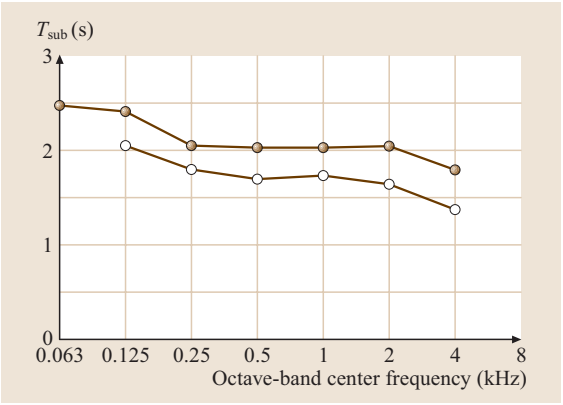


Fig. 10.45 Reverberation time measured in the Kirishima International Concert Hall. ●: measured values without audience; ○: estimated with full audience

For the 500 Hz octave band, examples of the measured decay curve and the decay rate of both left and right ears at seating position ***a*** are shown in Fig. 10.47. The reverberation times measured are both 2.07 s. The measured reverberation times with octave band filters in the Kirishima International Concert Hall (without audience) are plotted as filled circles in Fig. 10.45. The empty circles are estimated values of the reverberation time for a full audience. It is worth noticing that *Jordan* [10.81] showed that the values of the early decay time (EDT) measured over the first 10 dB of decay are close to values of the reverberation time averaged with the interval of −5 to −35 dB.

The total amplitude of reflection A, defined by (10.5), is obtained as its square

$$A^2 = \frac{\int_{\infty}^{\varepsilon} h^2(x)\,\mathrm{d}x}{\int_{\varepsilon}^{0} h^2(x)\,\mathrm{d}x}\,, \qquad (10.17)$$

where ε signifies a small delay time just large enough to cover the duration of the direct sound.

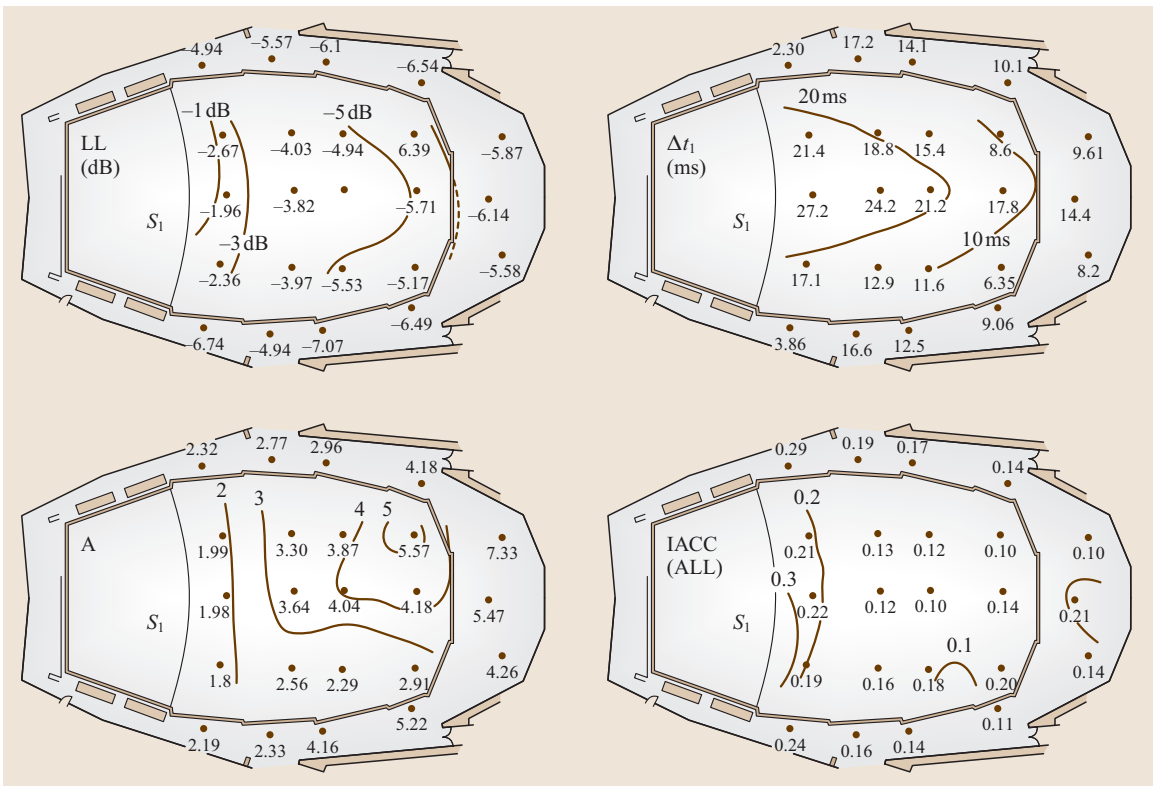


Fig. 10.46a–d Orthogonal factors measured at each seat in the Kirishima International Concert Hall, other than the reverberation time, which is almost constant throughout the hall: **(a)** listening level; **(b)** Δt_1; **(c)** A value, the total amplitude of reflections; **(d)** IACC

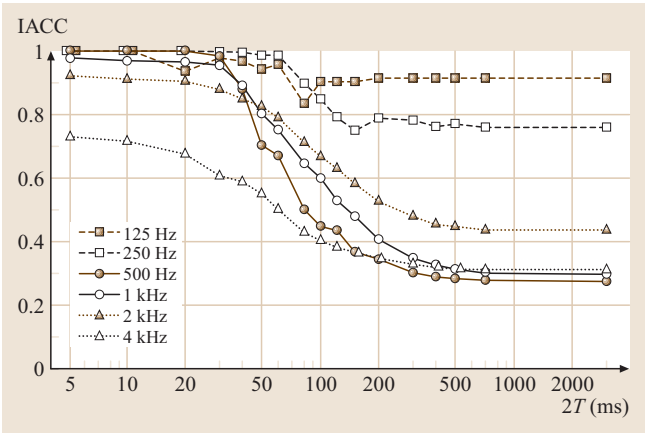


Fig. 10.47 Measured IACC as a function of the integration interval $2T$ of the impulse responses for each octave band range. The value of IACC converged for $2T > 200$ ms (After [10.82])

Measurement of Acoustic Factors at Each Seat in a Concert Hall

In Sect. 10.3.1 we discussed the seat selection system, designed for the purpose of enhancing individual satisfaction. To begin with, four orthogonal factors are measured at each seat in a concert hall [10.67, 68]. Measured values of the listening level (LL), the total amplitude of reflection (A), the initial time delay gap (Δt_1) between the direct sound and the first reflection excluding the reflection from the floor, and the IACC at each seat in the Kirishima International Concert Hall

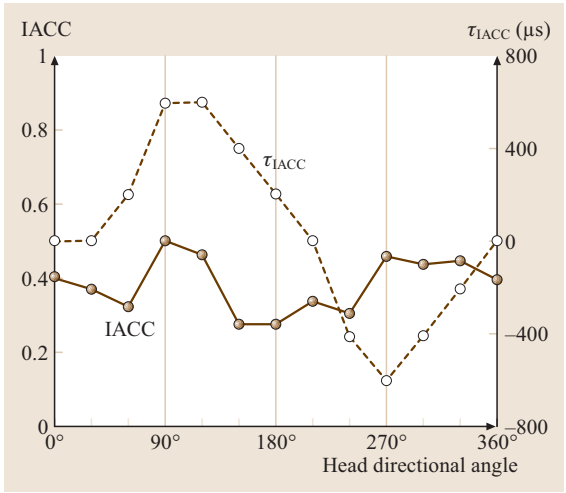


Fig. 10.48 Measured IACC and τ_{IACC} as a function of head direction relative to the sound source in a narrow hotel atrium (After [10.83])

are shown in Fig. 10.46. The reverberation times at all the seats had almost the same value, about 2.05 s for the 500 Hz band.

Even though the final scheme of the concert hall was changed in terms of the width of the hall (one meter wider) from the scheme at the design stage, values of each physical factor measured as shown in Fig. 10.46 are not very different from the values calculated (Fig. 10.22).

Recommended Method for IACC Measurement

There are two purposes for measuring IACC, as needed for subjective evaluations and acoustic comparison of existing halls:

1. In order to evaluate the subjective quality of the sound field in an existing hall, the IACF (with values of IACC, τ_{IACC} and W_{IACC} defined in Fig. 10.4 as well as LL) together with the other three factors Δt_1, T_{sub} and A are measured. Without any octave band filtering, measurements must be performed after passage of the music or the speech signal through an A-weighting network, under identical conditions with subjective judgment.
2. In order to compare values of the IACC as well as T_{sub} for the sound field in existing halls, measurements with octave band filtering are performed. With a fixed sound source on the stage, the IACC is defined by a long integration interval, which includes the effects of the direct sound and all reflections, including reverberation, without any temporal subdivisions.

A typical example of measuring the IACC as a function of the integration interval, which was performed in Symphony Hall, Boston, is shown in Fig. 10.47 [10.82]. It is remarkable that the measured values of IACC almost converged for $2T \approx 200$ ms, and the values are not so different for longer intervals.

If a room is used for performing dance, ice-skating or a party, then the listeners face in various directions. In this case, the values of IACC and τ_{IACC} are measured as a function of the direction of the head. The measured results with the 500 Hz octave band noise in an oblong atrium of a hotel at the distance 10 m from the source position are demonstrated in Fig. 10.48 [10.83]. When the listener is facing the sound source, then IACC = 0.41, and $\tau_{\mathrm{IACC}} = 0$, and thus no image shift occurs. These values are nearly unchanged for head directional angles less than 30° when the listener is facing the lateral side at 90°, then the IACC is greater than 0.50, and τ_{IACC} is about 600 μs, due to the interaural delay time.

Amazon EX1014 Page 394

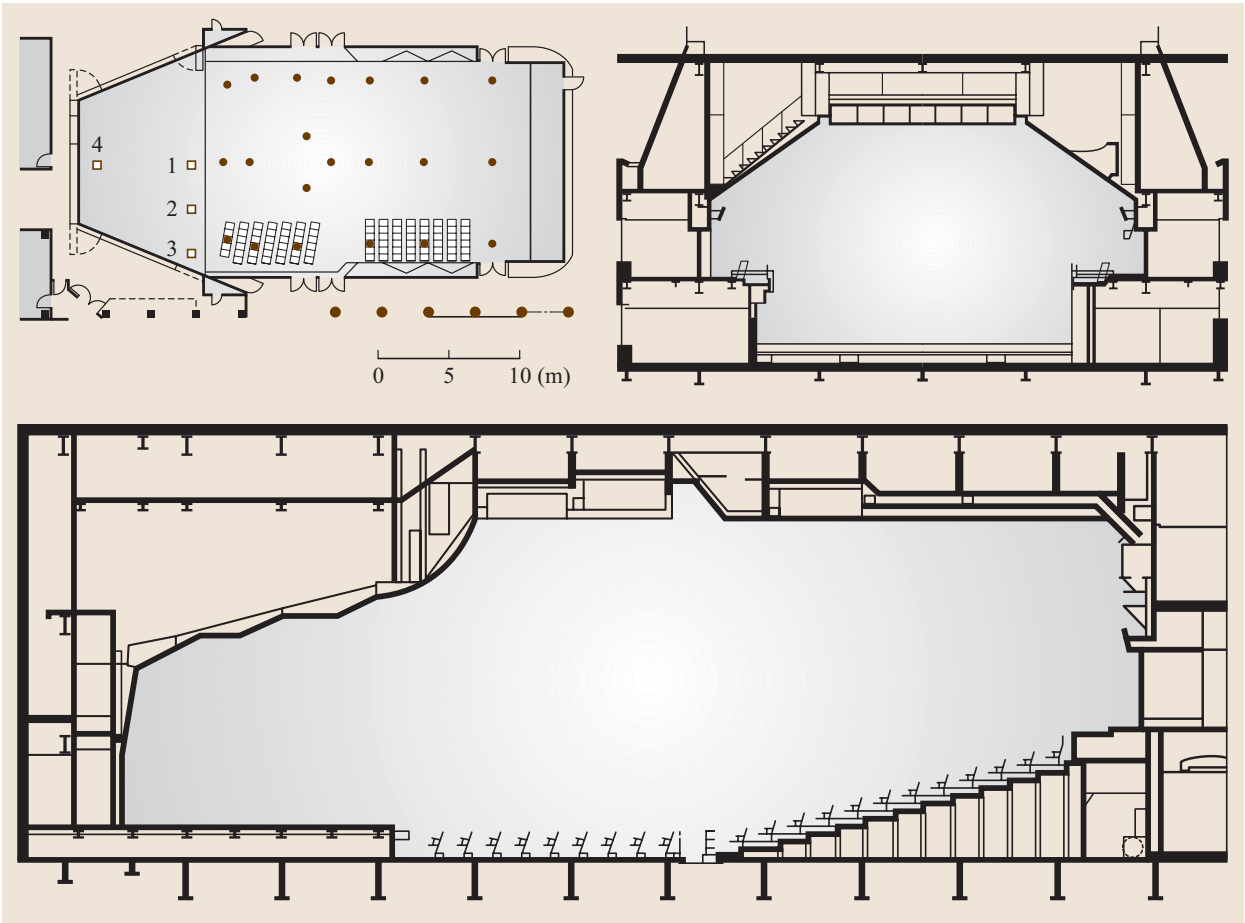


Fig. 10.49 *Left panel* plan; *right* and *bottom panel* cross sections of the Uhara Hall, Kobe. Four source positions, 1, 2, 3 and 4 on the stage, which were switched in the paired-comparison test of subjective preference without moving subjects from seat to seat. There were 21 listening positions including neighboring seats

10.4.2 Subjective Preference Test in an Existing Hall

The subjective preference judgments for different source locations on the stage at Uhara Hall, Kobe, were performed by a paired-comparison test at each set of seats. The relationship between the resulting scale value of subjective preference and the physical factors obtained by calculation, using architectural plan drawings, was examined by factor analysis [10.84]. Calculated scale values of subjective preference were reconfirmed for the Uhara Hall (Fig. 10.49). The physical factors at each set of seats for four source locations on the stage were calculated. In the simulation, the directional characteristics of the four loudspeakers used in preference tests were not taken into consideration for the sake of convenience. The simulation calculation was performed up to the third order of reflection. Due to a floor structure with a fair amount of acoustic transparency, the floor reflection was not taken into account for the calculation, and part of the diffuser ceiling was regarded as a nonreflective plane for the sake of convenience. In the calculation of the IACC, the listeners faced toward the center of the stage, so that the IACC was not always a maximum at the interaural time delay $\tau = 0$.

The hall contains 650 seats with a volume of 4870 m^3. Four identical loudspeakers with the same characteristics were placed 0.8 m above the stage floor,

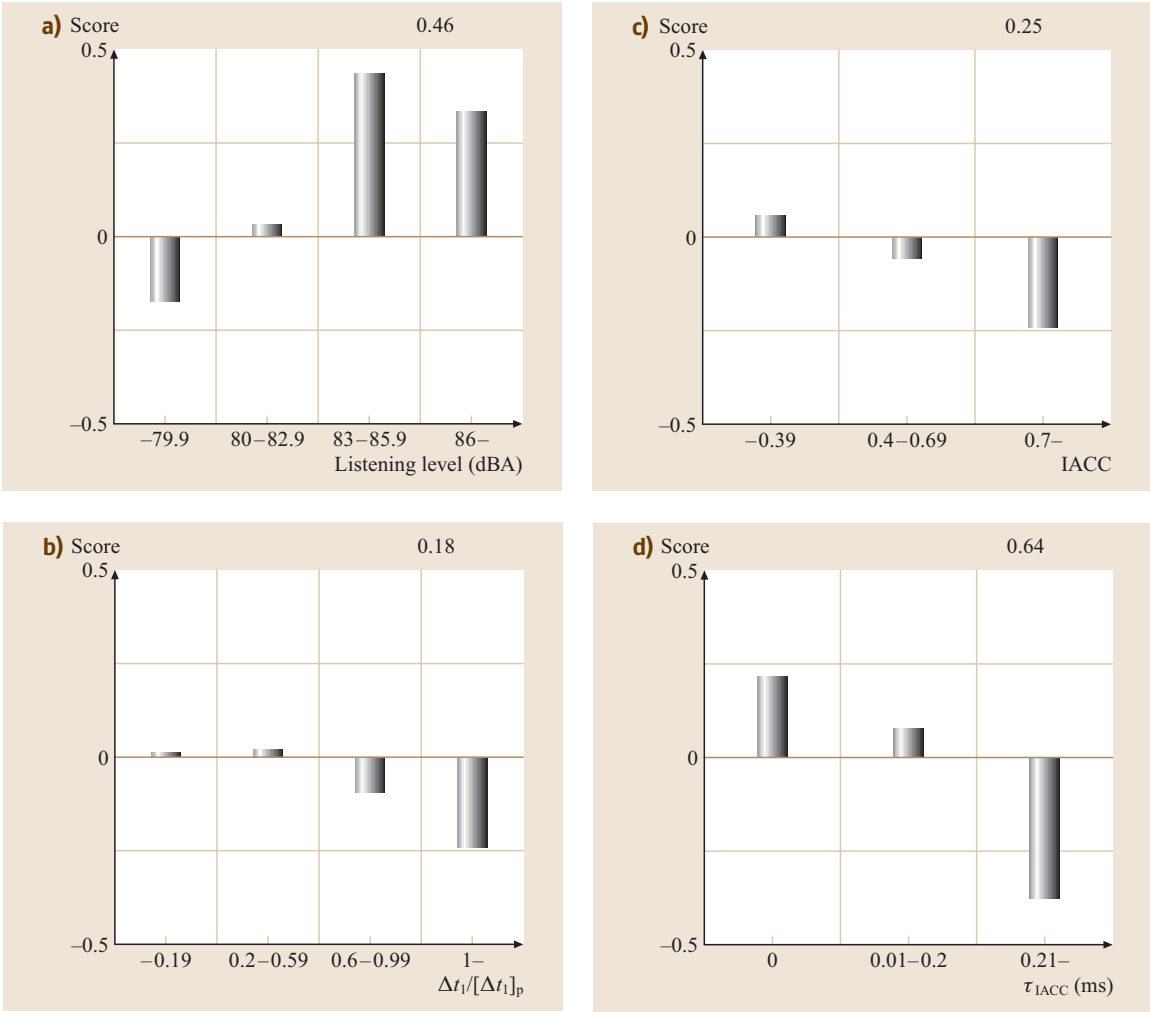


Fig. 10.50a–d Scores for each category of four physical factors obtained by the factor analyses. The number indicated at the upper left part of each figure signifies the partial correlation coefficient between the score and each factor. **(a)** Listening level; **(b)** normalized initial time delay gap between the direct sound and the first reflection; **(c)** IACC; **(d)** interaural time delay of the IACC, τ_{IACC}, found as the most significant factor in this investigation with loudspeaker reproductions on the stage. Tendencies obtained here are similar to those of the scale value shown in Fig. 10.7, which were obtained from the simulated sound field

and sixty-four listeners, divided into 21 groups, were seated in the specified set of seats. Without moving from seat to seat and excluding the effects of other physical factors such as visual and tactual senses on judgments, subjective preference tests by the paired-comparison method were conducted, switching only the loudspeakers on the stage. As a source signal, music motif B was selected in the tests. Scale values of preference were obtained by applying the law of comparative judgment and were reconfirmed by the goodness of the fit [10.85, 86]. The session was repeated five times, exchanging seats, and thus data for 14–16 subjects in total were obtained for each set of seats.

Results of Multiple-Dimensional Analyses

In order to examine the relationship between scale values of subjective preference and physical factors obtained

by simulation of an architectural scheme, the data were analyzed by factor analysis [10.79, 80, 87].

Of the four orthogonal factors, the reverberation time was almost constant for the source location and the seat location throughout the hall, and thus was not involved in the analysis. As previously discussed, as a condition for calculating the scale value of preference, the maximum value of the interaural cross-correlation function must be maintained at $\tau_{IACC} = 0$ to ensure frontal localization of the sound source. However, the IACC was not always maintained at $\tau = 0$ due to the loudspeaker locations, because the subjects were facing the center of stage. In this analysis, therefore, the effect of the interaural time delay of the IACC was added as an additional factor. Thus, the outside variable to be predicted with factor analysis was the scale value obtained by subjective judgments, and the explanatory factors were: (1) the listening level, (2) the initial time delay gap, (3) the IACC, and (4) the interaural time delay (τ_{IACC}).

Scores for Each Factor

The scores for each category of the factors obtained from the factor analysis are shown in Fig. 10.50. As shown in Fig. 10.50a, the scores for the listening level indicate a peak at the subcategory of 83–85.9 dB, with decreasing scores moving away from the preferred listening level. For the IACC, the preference score increases with a decrease in the IACC (Fig. 10.50c). It is worth noting that the scores for the aforementioned two factors are in good agreement with preference scale values obtained from preference judgments for a simulated sound field (Fig. 10.7a and 10.7d). The scores of the initial time delay gap normalized to the optimum value ($\Delta t_1/[\Delta t_1]_p$) peaked at smaller values (Fig. 10.50b) than the most preferred value of the initial time delay gap obtained from the simulated sound fields (Fig. 10.7b). It is considered that, due to the limited range of the Δt_1 in the existing concert hall and the limited data for the short range of the Δt_1, the effects of the Δt_1 of the sound fields was rather minor in this investigation. Concerning τ_{IACC}, as shown in Fig. 10.50d, the score decreases monotonically as the delay is increased. This may be caused by an image shift without balancing of the sound field.

Measured and Calculated Subjective Preference Values

The relationship between the scale value obtained by subjective judgments and the total score at each center of three or four seats is shown in Fig. 10.51. The scale values of preference are well predicted with the total score for four loudspeaker locations ($r = 0.70$, $p < 0.01$). In some cases there is a certain degree of

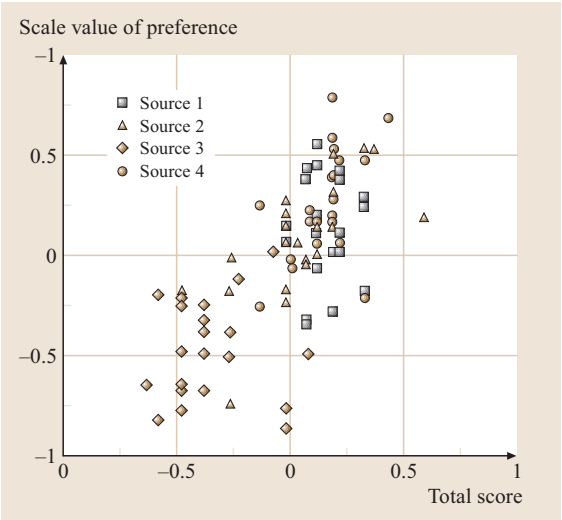


Fig. 10.51 Relationship between the scale value of subjective preference obtained by the paired-comparison test in the existing hall and the total score calculated by using the scores shown in Fig. 10.50. The correlation coefficient was $r = 0.70$ ($p < 0.01$)

apparent coherence between physical factors, for example, the calculated listening level and the IACC for sound fields in existing concert halls. However, these factors are theoretically orthogonal, and therefore the preference scores obtained were in good agreement with the calculated preference scale values obtained by the simulation.

So far the subjective preference of source locations on the stage has been examined at each set of seats. The rear source position (#4) on the stage is more preferred than that of the other source locations. The side source position (#3) indicates a low preference, due to the interaural time delay. The initial time delay gap resulted in a small influence on the total score because of its limited range.

10.4.3 Conclusions

Results of the analysis demonstrate that the theory of calculating subjective preference by the use of orthogonal parameters obtained in the laboratory is supported by experiments in a real hall. This may hold only when the maximum value of the interaural cross-correlation is maintained at $\tau = 0$. However, this condition is usually realized by introducing certain diffusing elements on the sidewalls in a real concert hall when the listeners face a performer.

Amazon EX1014 Page 397

References

10.1 Y. Ando: *Architectural Acoustics, Blending Sound Sources, Sound Fields, and Listeners* (AIP/Springer, New York 1998)

10.2 W.C. Sabine: *Reverberation (The American Architect and the Engineering Record, 1900), Prefaced by L.L. Beranek: Collected papers on acoustics* (Peninsula, Los Altos 1992)

10.3 H. Haas: Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache, Acustica **1**, 49–58 (1951)

10.4 L.L. Beranek: *Music, Acoustics and Architecture* (Wiley, New York 1962)

10.5 P. Damaske: Subjektive Untersuchungen von Schallfeldern, Acustica **19**, 199–213 (1967/68)

10.6 M.V. Keet: *The influence of early lateral reflections on the spatial impression* (Proc. 6th Int. Congress of Acoustics, Tokyo 1968), Tech. Dig., paper E-2-4

10.7 A.H. Marshall: Acoustical determinants for the architectural design of concert halls, Architect. Sci. Rev. (Australia) **11**, 81–87 (1968)

10.8 M. Barron: The subjective effects of first reflections in concert halls – The need for lateral reflections, J. Sound Vibr. **15**, 475–494 (1971)

10.9 P. Damaske, Y. Ando: Interaural crosscorrelation for multichannel loudspeaker reproduction, Acustica **27**, 232–238 (1972)

10.10 M.R. Schroeder, D. Gottlob, K.F. Siebrasse: Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters, J. Acoust. Soc. Am. **65**, 958–963 (1974)

10.11 Y. Ando: Subjective preference in relation to objective parameters of music sound fields with a single echo, J. Acoust. Soc. Am. **62**, 1436–1441 (1977)

10.12 Y. Ando: Calculation of subjective preference at each seat in a concert hall, J. Acoust. Soc. Am. **74**, 873–887 (1983)

10.13 Y. Ando: *Concert Hall Acoustics* (Springer, Berlin, Heidelberg 1985)

10.14 A. Cocchi, A. Farina, L. Rocco: Reliability of scale-model research: A concert hall case, Appl. Acoust. **30**, 1–13 (1990)

10.15 S. Sato, Y. Mori, Y. Ando: The subjective evaluation of source locations on the stage by listeners, music and concert hall acoustics. In: *Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997) pp. 117–123, Chap. 12

10.16 T. Hotehama, S. Sato, Y. Ando: Dissimilarity judgments in relation to temporal and spatial factors for the sound fields in an existing hall, J. Sound Vibr. **258**, 429–441 (2002)

10.17 H. Sakai, P.K. Singh, Y. Ando: Inter-individual differences in subjective preference judgments of sound fields. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997) pp. 125–130, Chap 13

10.18 M. Sakurai, Y. Korenaga, Y. Ando: A sound simulation system for seat selection. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997) pp. 51–59, Chap. 6

10.19 Y. Ando: Evoked potentials relating to the subjective preference of sound fields, Acustica **76**, 292–296 (1992)

10.20 S. Sato, K. Nishio, Y. Ando: Propagation of alpha waves corresponding to subjective preference from the right hemisphere to the left with change in the IACC of a sound field, J. Temporal Des. Architect. Environ. **3**, 60–69 (2003)

10.21 Y. Soeta, S. Nakagawa, M. Tonoike, Y. Ando: Magnetoencephalographic responses corresponding to individual subjective preference of sound fields, J. Sound Vibr. **258**, 419–428 (2002)

10.22 Y. Soeta, S. Nakagawa, M. Tonoike, Y. Ando: Spatial analysis of magnetoencephalographic alpha waves in relation to subjective preference of a sound field, J. Temporal Des. Architect. Environ. **3**, 28–35 (2003)

10.23 L.A. Jeffress: A place theory of sound localization, J. Comp. Physiol. Psychol. **41**, 35–39 (1948)

10.24 J.C.R. Licklider: A duplex theory of pitch perception, Experientia **7**, 128–134 (1951)

10.25 Y. Ando, K. Yamamoto, H. Nagamastu, S.H. Kang: Auditory brainstem response (ABR) in relation to the horizontal angle of sound incidence, Acoust. Lett. **15**, 57–64 (1991)

10.26 H.E. Secker-Walker, C.L. Searle: Time domain analysis of auditory-nerve-fiber firing rates, J. Acoust. Soc. Am. **88**, 1427–1436 (1990)

10.27 P.A. Cariani, B. Delgutte: Neural correlates of the pitch of complex tones. I. Pitch and pitch salience, J. Neurophysiol. **76**, 1698–1716 (1996)

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

10.29 M. Inoue, Y. Ando, T. Taguti: The frequency range applicable to pitch identification based upon the auto-correlation function model, J. Sound Vibr. **241**, 105–116 (2001)

10.30 S. Sato, T. Kitamura, H. Sakai, Y. Ando: The loudness of "complex noise" in relation to the factors extracted from the autocorrelation function, J. Sound Vibr. **241**, 97–103 (2001)

10.31 K. Saifuddin, T. Matsushima, Y. Ando: Duration sensation when Listening to pure tone and complex tone, J. Temporal Des. Architect. Environ. **2**, 42–47 (2002)

10.32 Y. Ando, H. Sakai, S. Sato: Formulae describing subjective attributes for sound fields based on

Amazon EX1014 Page 398

a model of the auditory-brain system, J. Sound Vibr. **232**, 101–127 (2000)
10.33 Y. Ando: A theory of primary sensations measuring environmental noise, J. Sound Vibr. **241**, 3–18 (2001)
10.34 Y. Ando, Y. Kurihara: Nonlinear response in evaluating the subjective diffuseness of sound field, J. Acoust. Soc. Am. **80**, 833–836 (1986)
10.35 Y. Ando, S.H. Kang, H. Nagamatsu: On the auditory-evoked potentials in relation to the IACC of sound field, J. Acoust. Soc. Jpn. (E) **8**, 183–190 (1987)
10.36 Y. Ando, S. Sato, H. Sakai: Fundamental subjective attributes of sound fields based on the model of auditory – brain system. In: *Computational Acoustics in Architecture*, ed. by J.J. Sendra (WIT, Southampton 1999)
10.37 Y. Ando, R. Pompoli: Factors to be measured of environmental noise and its subjective responses based on the model of auditory-brain system, J. Temporal Des. Architect. Environ. **2**, 2–12 (2002)
10.38 S. Sato, Y. Ando, V. Mellert: Cues for localization in the median plane extracted from the autocorrelation function, J. Sound Vibr. **241**, 53–56 (2001)
10.39 Y. Ando, T. Okano, Y. Takezoe: The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields, J. Acoust. Soc. Am. **86**, 644–649 (1989)
10.40 Y. Ando, K. Kageyama: Subjective preference of sound with a single early reflection, Acustica **37**, 111–117 (1977)
10.41 K. Kato, Y. Ando: A study of the blending of vocal music with the sound field by different singing styles, J. Sound Vibr. **258**, 463–472 (2002)
10.42 K. Kato, K. Fujii, K. Kawai, Y. Ando, T. Yano: *Blending vocal music with the sound field – the effective duration of autocorrelation function of Western professional singing voices with different vowels and pitches, Proc. Int. Symp. Musical Acoust., ISMA* (Acoustical Society of Japan, Kyoto 2004) pp. 37–40
10.43 Y. Ando, M. Sakamoto: Superposition of geometries of surface for desired directional reflections in a concert hall, J. Acoust. Soc. Am. **84**, 1734–1740 (1988)
10.44 Y. Ando: Investigations on cerebral hemisphere activities related to subjective preference of the sound field, published for 1983 – 2003, J. Temporal Design Architect. Environ. **3**, 2–27 (2003)
10.45 S. Sato, K. Nishio, Y. Ando: Propagation of alpha waves corresponding to subjective preference from the right hemisphere to the left with change in the IACC of a sound field, J. Temporal Des. Architect. Environ. **3**, 60–69 (2003)
10.46 K. Mouri, K. Akiyama, Y. Ando: Preliminary study on recommended time duration of source signals to be analyzed, in relation to its effective duration of autocorrelation function, J. Sound Vibr. **241**, 87–95 (2001)
10.47 T. Taguti, Y. Ando: Characteristics of the short-term autocorrelation function of sound signals in piano performances. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997) pp. 233–238, Chap. 23
10.48 Y. Ando, S.H. Kang, Morita: On the relationship between auditory-evoked potentials and subjective preference for sound field, J. Acoust. Soc. Jpn. (E) **8**, 197–204 (1987)
10.49 Y. Ando, S.H. Kang, H. Nagamatsu: On the auditory-evoked potential in relation to the IACC of sound field, J. Acoust. Soc. Jpn. (E) **8**, 183–190 (1987)
10.50 Y. Ando, C. Chen: On the analysis of autocorrelation function of α-waves on the left and right cerebral hemispheres in relation to the delay time of single sound reflection, J. Architect. Planning Environ. Eng., Architectural Institute of Japan (AIJ) **488**, 67–73 (1996), (in Japanese)
10.51 C. Chen, Y. Ando: On the relationship between the autocorrelation function of the α-waves on the left and right cerebral hemispheres and subjective preference for the reverberation time of music sound field, J. Architect. Planning Environ. Eng., Architectural Institute of Japan (AIJ) **489**, 73–80 (1996), (in Japanese)
10.52 Y. Soeta, S. Nakagawa, M. Tonoike, Y. Ando: Magnetoencephalographic responses corresponding to individual subjective preference of sound fields, J. Sound Vibr. **258**, 419–428 (2002)
10.53 Y. Soeta, S. Nakagawa, M. Tonoike, Y. Ando: Spatial analysis of magnetoencephalographic alpha waves in relation to subjective preference of a sound field, J. Temporal Des. Architect. Environ. **3**, 28–35 (2003)
10.54 D. Kimura: The asymmetry of the human brain, Sci. Am. **228**, 70–78 (1973)
10.55 R.W. Sperry: Lateral specialization in the surgically separated hemispheres. In: *The Neurosciences: Third study program*, ed. by F.O. Schmitt, F.C. Worden (MIT, Cambridge 1974), Chap. 1
10.56 J.H. Holland: *Adaptation in Natural and Artificial Systems* (Univ. Michigan Press, Ann Arbor 1975)
10.57 S. Sato, K. Otori, A. Takizawa, H. Sakai, Y. Ando, H. Kawamura: Applying genetic algorithms to the optimum design of a concert hall, J. Sound Vibr. **258**, 517–526 (2002)
10.58 S. Sato, T. Hayashi, A. Takizawa, A. Tani, H. Kawamura, Y. Ando: Acoustic design of theatres applying genetic algorithms, J. Temporal Des. Architect. Environ. **4**, 41–51 (2004)
10.59 H. Sakai, P. K. Singh, Y. Ando: Inter-individual differences in subjective preference judgments of sound fields. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997), Chap. 13
10.60 F. Maki: Sound and figure: concert hall design. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997), Chap. 1

Amazon EX1014 Page 399

10.61 Y. Ikeda: Designing a contemporary classic concert hall using computer graphics. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997), Chap. 2

10.62 I. Nakayama: Preferred time delay of a single reflection for performers, Acustica **54**, 217–221 (1984)

10.63 I. Nakayama, T. Uehara: Preferred direction of a single reflection for a performer, Acustica **65**, 205–208 (1988)

10.64 J. Meyer: *Influence of Communication on Stage on the Musical Quality* (Proc. 15th Int. Congress Acoustics, Trondheim 1995) pp. 573–576

10.65 Y. Ando, B. P. Johnson, T. Bosworth: Theory of planning physical environments incorporating spatial and temporal values, Memoir of Graduate School Science and Technology **14-A**, 67–92 (1996)

10.66 Y. Ando: On the temporal design of environments, J. Temporal Des. Architect. Environ. **4**, 2–14 (2004)

10.67 Y. Ando, S. Sato, T. Nakajima, M. Sakurai: Acoustic design of a concert hall applying the theory of subjective preference, and the acoustic measurement after construction, Acustica Acta Acustica. **83**, 635–643 (1997)

10.68 T. Nakajima, Y. Ando: Calculation and measurement of acoustic factors at each seat in the Kirishima International Concert Hall. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997), Chap. 5

10.69 M.R. Schroeder: *Number Theory in Science and Communication* (Springer, Berlin, Heidelberg 1984)

10.70 P.M. Morse: *Vibration and Sound* (McGraw-Hill, New York 1948)

10.71 Y. Ando, P.K. Singh: Global subjective evaluations for design of sound fields and individual subjective preference for seat selection. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997), Chap. 4

10.72 A.H. Marshall, D. Gottlob, H. Alrutz: Acoustical conditions preferred for ensemble, J. Acoust. Soc. Am. **64**, 1437–1442 (1978)

10.73 A.C. Gade: Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and laboratory experiments, Acustica **69**, 193–203 (1989)

10.74 D. Noson, S. Sato, H. Sakai, Y. Ando: Singer responses to sound fields with a simulated reflection, J. Sound Vibr. **232**, 39–51 (2000)

10.75 D. Noson, S. Sato, H. Sakai, Y. Ando: Melisma singing and preferred stage acoustics for singers, J. Sound Vibr. **258**, 473–485 (2002)

10.76 S. Sato, S. Ohta, Y. Ando: Subjective preference of cellists for the delay time of a single reflection in a performance, J. Sound Vibr. **232**, 27–37 (2000)

10.77 H.P. Seraphim: Über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall, Acustica **11**, 80–91 (1961)

10.78 N. Aoshima: Computer-generated pulse signal applied for sound measurement, J. Acoust. Soc. Am. **69**, 1484–1488 (1981)

10.79 C. Hayashi: Multidimensional quantification I., Proc. Jpn. Acad. **30**, 61–65 (1954)

10.80 C. Hayashi: Multidimensional quantification II., Proc. Jpn. Acad. **30**, 165–169 (1954)

10.81 V.L. Jordan: Acoustical criteria for auditoriums and their relation to model techniques, J. Acoust. Soc. Am. **47**, 408–412 (1969)

10.82 T. Hidaka: Personal communication (1996)

10.83 Y. Kobayasi, H. Tokuhiro, M. Owaki, K. Okuno, S. Yamada, Y. Ando: Acoustical design and characteristics of the atrium for music performance in a hotel. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997), Chap. 27

10.84 S. Sato, Y. Mori, Y. Ando: The subjective evaluation of source locations on the stage by listeners. In: *Music and Concert Hall Acoustics, Conf. Proc. MCHA 1995*, ed. by Y. Ando, D. Noson (Academic, London 1997), Chap. 12

10.85 L.L. Thurstone: A law of comparative judgment, Psychol. Rev **34**, 273–289 (1951)

10.86 F. Mosteller: Remarks on the method of paired comparison, III. Psychometrika **16**, 207–218 (1951)

10.87 C. Hayashi: On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, Ann. Inst. Statist. Math., 69–98 (1952)

Amazon EX1014 Page 400

11. Building Acoustics

This chapter summarizes and explains key concepts of building acoustics. These issues include the behavior of sound waves in rooms, the most commonly used rating systems for sound and sound control in buildings, the most common noise sources found in buildings, practical noise control methods for these sources, and the specific topic of office acoustics. Common noise issues for multi-dwelling units can be derived from most of the sections of this chapter. Books can be and have been written on each of these topics, so the purpose of this chapter is to summarize this information and provide appropriate resources for further exploration of each topic.

11.1 **Room Acoustics** 387
11.1.1 Room Modes 388
11.1.2 Sound Fields in Rooms 389
11.1.3 Sound Absorption 390
11.1.4 Reverberation 394
11.1.5 Effects of Room Shapes 394
11.1.6 Sound Insulation 395

11.2 **General Noise Reduction Methods** 400
11.2.1 Space Planning 400
11.2.2 Enclosures 400
11.2.3 Barriers 402
11.2.4 Mufflers 402
11.2.5 Absorptive Treatment 402
11.2.6 Direct Impact and Vibration Isolation 402
11.2.7 Active Noise Control 402
11.2.8 Masking 403

11.3 **Noise Ratings for Steady Background Sound Levels** 403

11.4 **Noise Sources in Buildings** 405
11.4.1 HVAC Systems 405
11.4.2 Plumbing Systems 406
11.4.3 Electrical Systems 406
11.4.4 Exterior Sources 406

11.5 **Noise Control Methods for Building Systems** 407
11.5.1 Walls, Floor/Ceilings, Window and Door Assemblies 407
11.5.2 HVAC Systems 412
11.5.3 Plumbing Systems 415
11.5.4 Electrical Systems 416
11.5.5 Exterior Sources 417

11.6 **Acoustical Privacy in Buildings** 419
11.6.1 Office Acoustics Concerns 419
11.6.2 Metrics for Speech Privacy 419
11.6.3 Fully Enclosed Offices 422
11.6.4 Open-Plan Offices 422

11.7 **Relevant Standards** 424

References 425

11.1 Room Acoustics

When a sound wave in a room encounters a boundary surface, some of its energy is reflected back into the room, as illustrated in Fig. 11.1. The physical laws regarding reflected sound energy are analogous to those for the laws of optics. Just as light bounces off a mirror at the same angle as its angle of incidence, sound waves have equal angles of incidence and reflection. Acoustically reflective surfaces are typically smooth and hard.

Some common room acoustics problems caused by reflection are echoes and room resonance. Echoes result from the limitations of our hearing mechanisms to process sounds. When the difference in arrival times between two sounds is less than 60 ms, we hear the combination of the two sounds as a single sound. However, when this difference exceeds 60 ms, we hear the two distinct sounds (Fig. 11.2). When these two sounds are generated by the same source, this effect (which is called an echo) can cause difficulty in understanding speech, especially when arrival times differ by more than 100 ms. These kinds of delays occur when a person hears a sound wave coming directly from a source and another coming from a reflecting surface. Given that the speed of sound in air is roughly 300 m per second, the 100 ms delay translates to a distance of roughly 30 m. There-

Amazon EX1014 Page 401

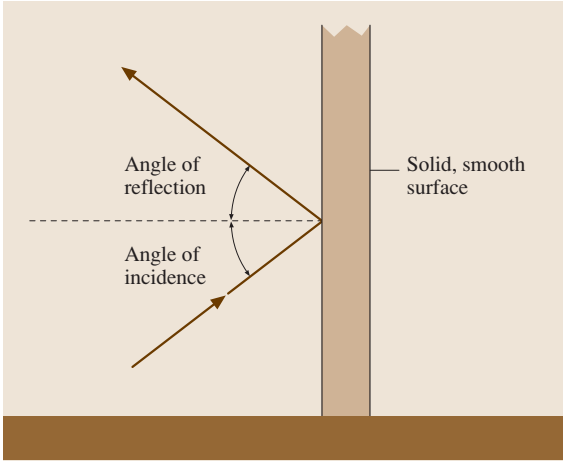


Fig. 11.1 Reflection of a sound wave off a hard, smooth surface

fore, a difference in sound travel path of more than 30 m between a sound wave traveling directly from a source to a listener and a sound wave traveling from a source to a reflecting surface and then to a listener causes an echo.

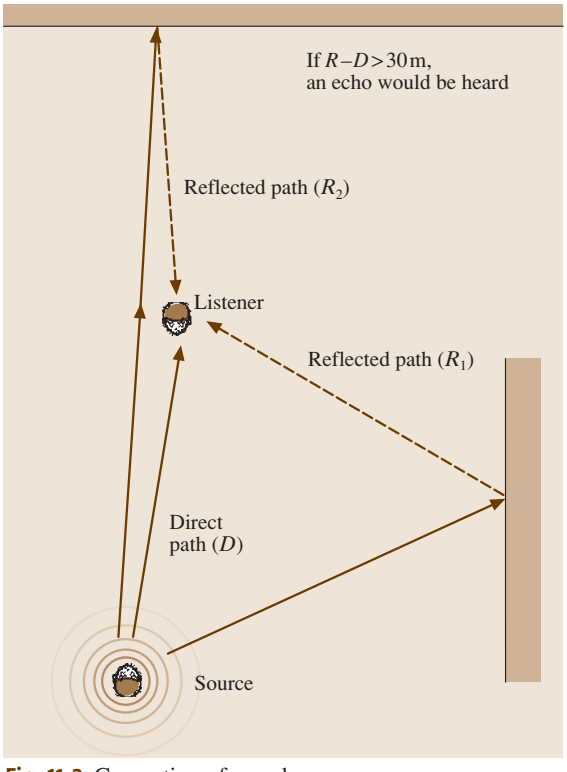


Fig. 11.2 Generation of an echo

When parallel reflective surfaces are tall and more than 10 m from each other, a rapid succession of mid-frequency echoes can occur, known as flutter echo, resulting in a flapping sound similar to that generated by birds or bats. When these surfaces are fairly close to each other, as in a narrow hallway or stairwell, flutter echo sounds more like a high-pitched buzzing since the smaller spacing accommodates higher frequencies (shorter wavelengths).

Echoes are normally perceived as discrete reflections that can be clearly heard and identified, but many reflections off all surfaces within a room can combine to produce the phenomenon known as reverberation. Reverberation can raise the sound level in a room and can also reduce speech intelligibility, but it is desirable for certain types of music. More discussion on reverberation and its control can be found later in this chapter.

11.1.1 Room Modes

Room modes occur at specific frequencies when two reflective walls are parallel to each other. Since the surfaces reflect the sound, their mirror images reflect off each wall to set up a stationary pressure pattern in a room. This phenomenon is called a single-dimensional (or axial) standing wave, as shown in Fig. 11.3, and it is the simplest form of room modes.

Axial standing waves occur at frequencies described by the equation below:

$$f_n = nc/2d\,, \tag{11.1}$$

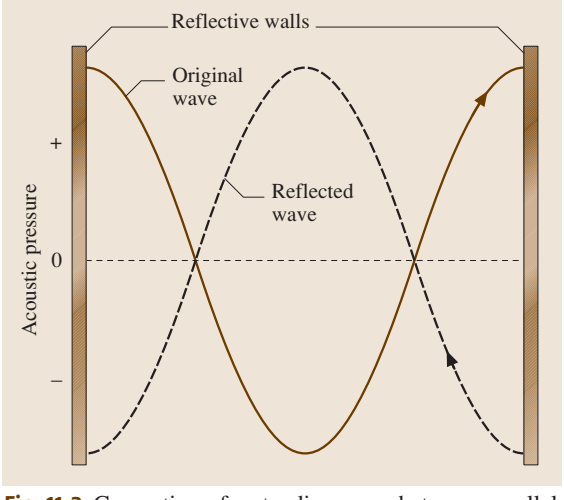


Fig. 11.3 Generation of a standing wave between parallel, reflective surfaces

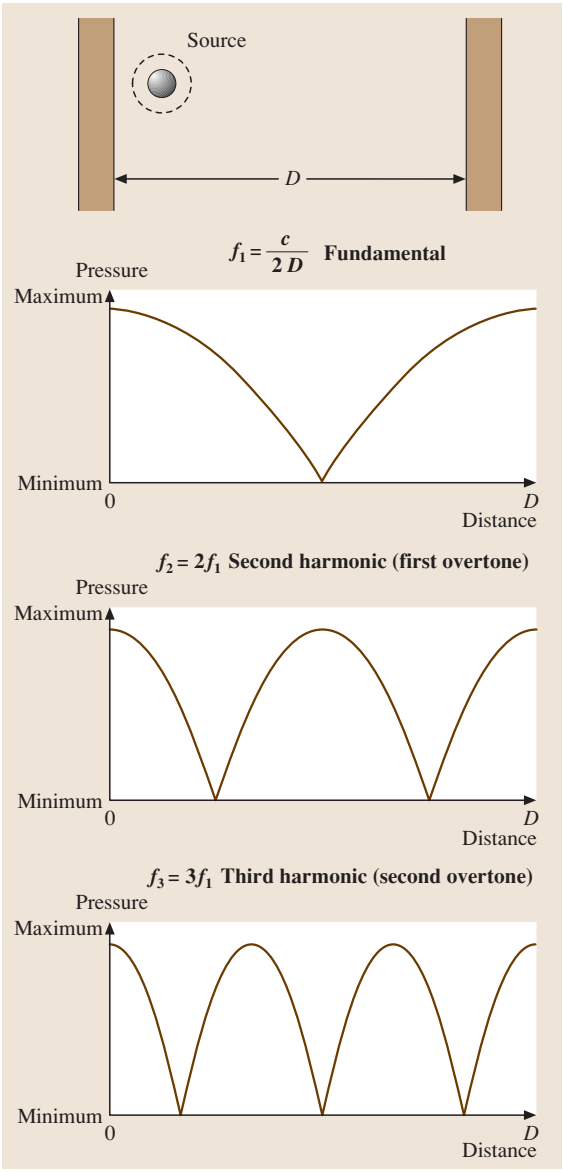


Fig. 11.4 Pressure patterns of the lowest axial standing wave modes in a room with parallel reflective surfaces

where $n = 1, 2, 3, \ldots$, c is the speed of sound, and d is the distance between the parallel reflective surfaces. This means that axial standing wave room modes occur at integer multiples of specific 1/2-wavelengths of sound, as is shown in Fig. 11.4. Standing waves can become more complex in two and three dimensions, where they are known as tangential and oblique modes, respectively.

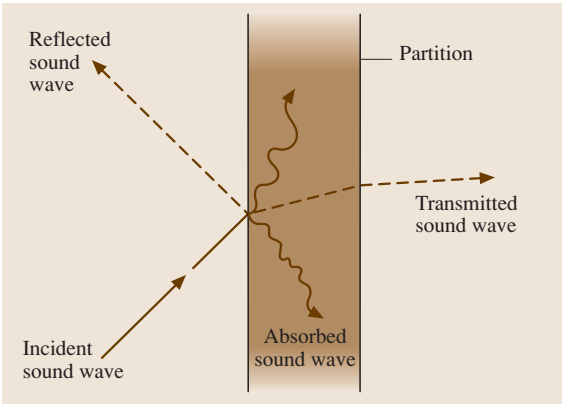


Fig. 11.5 The interaction of a sound wave with a partition

A design problem with standing waves is that they generate uneven sound distribution patterns. Some areas will have higher levels of sound (because a standing wave is reinforcing the pressure at those locations) and some areas will have lower levels of sound (because the standing wave is canceling much of the pressure at those locations), but only at specific frequencies.

The topic of building acoustics stresses the importance of controlling sound. When one talks about controlling sound, it is often assumed that one is referring to the reduction of sound. However, there are cases (e.g., auditorium or concert hall design) in which we want to preserve the sound energy but we would like to control its spatial spreading characteristics. The primary ways to reduce sound are through absorption and insulation. Using absorption on an auditorium's side walls may eliminate unwanted reflections but may also eliminate the possibility of some people hearing sound coming from the stage. We therefore must clarify how we plan to control the sound. Redirection and diffusion can have favorable acoustic results for even sound distribution in large rooms.

Discussions about noise control usually refer to the reduction of sound (since noise is defined as unwanted sound). Figure 11.5 shows what happens to a sound wave that interacts with a room's surface. Part is absorbed, part is redirected, and the rest is transmitted through the surface.

11.1.2 Sound Fields in Rooms

Several different fields, or acoustic environments, exist in all rooms. The size of these fields depends on the dimensions of the room, the reflective qualities of the room's surfaces, and the frequency of interest. The

Amazon EX1014 Page 403

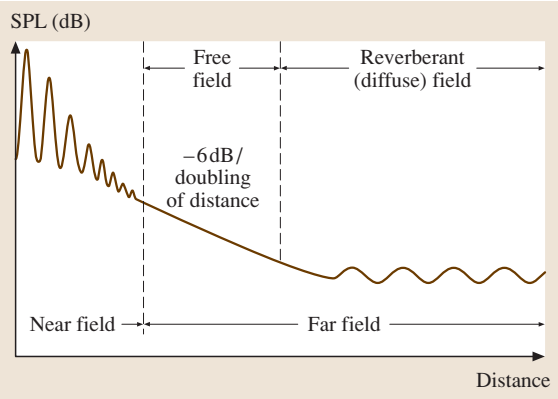


Fig. 11.6 Sound fields in rooms

near field is the region within 1/4-wavelength (of the lowest frequency of interest) of a sound source or large reflective surface. Sound pressure levels can fluctuate dramatically in the near field of a source and sound pressures cancel and enhance each other near large reflective surfaces, so sound pressure level measurements should be avoided in near fields.

The far field is the region beyond the near field. It is contrasted with the near field to designate the region that is appropriate for recording sound pressure level measurements from a sound source. As one moves out of the near field and into the far field, sound pressure levels drop off at a rate of 6 dB per doubling of distance, for a point source, in accordance with the inverse square law that is simplified in the following equation:

$$\mathrm{SPL}_2 = \mathrm{SPL}_1 - [20 \times \log(d_2/d_1)] \,, \qquad (11.2)$$

where SPL_1 is the sound pressure level at the location closer to the sound source, SPL_2 is the sound pressure level at the location farther from the sound source, d_1 is the distance from the source at which SPL_1 is measured, and d_2 is the distance from the source at which SPL_2 is measured.

The region in which (11.2) is valid is known as the free field. Free-field conditions exist in large open outdoor spaces or in rooms having highly absorptive surfaces, in which there are no obstructions in the sound travel path between the source and listener. The free field is sometimes referred to as the direct field when source measurements are taking place that only consider the sound wave traveling from the source to the listener with no influence by reflected sound coming from room surfaces.

As sound pressure levels decay from sources within a room, they eventually drop to a relatively constant level which is determined by the amount of reflected sound within a room. This lower limit is known as the reverberant or diffuse field. The size of the reverberant field depends on the size of the room and the size and characteristics of its reflective surfaces. Within the reverberant field, sound pressures are similar independent of location. Figure 11.6 puts all of the fields described above into perspective.

11.1.3 Sound Absorption

Absorption converts sound energy into heat energy. It is useful for reducing sound levels *within* rooms but not *between* rooms. Each material with which a sound wave interacts absorbs some sound. The most common measurement of that is the absorption coefficient, typically denoted by the Greek letter α. The absorption coefficient is a ratio of absorbed to incident sound energy. If a material does not absorb any sound incident upon it, its absorption coefficient is 0. In other words, a material with an absorption coefficient of 0 reflects all sound incident upon it. In practice, all materials absorb some sound, so this is a theoretical limit. If a material absorbs all sound incident upon it, its absorption coefficient is 1. As with the lower limit for absorption coefficients, all materials reflect some sound, so this is also a theoretical limit. Therefore, absorption coefficients range between 0 and 1.

Absorption coefficients vary with frequency. Typical absorptive materials are porous and characterized by absorption coefficients that increase with frequency. They therefore have limited effectiveness for lower frequencies, especially below 250 Hz. There are absorbers that have been designed to absorb these lower frequencies, and these will be discussed shortly. However, for typical cases, it is convenient to use a single number (incorporating multiple frequency components) to describe the absorption characteristics of a material. This value has been defined by the American Society for Testing and Materials (ASTM) in Standard C 423 as the noise reduction coefficient (NRC). The NRC is the arithmetic (as opposed to the logarithmic) average of a material's absorption coefficients at 250, 500, 1000, and 2000 Hz, rounded to the closest 0.05.

Table 11.1 lists absorption coefficients and NRC values for common materials. Note that the values listed in Table 11.1 are for general reference purposes only, and specific values should be based on manufacturers specifications. Also note that absorption coefficients and NRC values have no units associated with them. In general, materials with NRC values less than 0.20 are considered

Amazon EX1014 Page 404

Table 11.1 Absorption coefficients and NRC values for common materials

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	NRC
Painted drywall	0.10	0.08	0.05	0.03	0.03	0.03	0.05
Plaster	0.02	0.03	0.04	0.05	0.04	0.03	0.05
Smooth concrete	0.10	0.05	0.06	0.07	0.09	0.08	0.05
Coarse concrete	0.36	0.44	0.31	0.29	0.39	0.25	0.35
Smooth brick	0.03	0.03	0.03	0.04	0.05	0.07	0.05
Glass	0.05	0.03	0.02	0.02	0.03	0.02	0.05
Plywood	0.58	0.22	0.07	0.04	0.03	0.07	0.10
Metal blinds	0.06	0.05	0.07	0.15	0.13	0.17	0.10
Thick panel	0.25	0.47	0.71	0.79	0.81	0.78	0.70
Light drapery	0.03	0.04	0.11	0.17	0.24	0.35	0.15
Heavy drapery	0.14	0.35	0.55	0.72	0.70	0.65	0.60
Helmholtz resonator	0.20	0.95	0.85	0.49	0.53	0.50	0.70
Ceramic tile	0.01	0.01	0.01	0.01	0.02	0.02	0.00
Linoleum	0.02	0.03	0.03	0.03	0.03	0.02	0.05
Carpet	0.05	0.05	0.10	0.20	0.30	0.40	0.15
Carpet on concrete	0.05	0.10	0.15	0.30	0.50	0.55	0.25
Carpet on rubber	0.05	0.15	0.13	0.40	0.50	0.60	0.30
Upholstered seats	0.19	0.37	0.56	0.67	0.61	0.59	0.55
Occupied seats	0.39	0.57	0.80	0.94	0.92	0.87	0.80
Water surface	0.01	0.01	0.01	0.01	0.02	0.03	0.00
Soil	0.15	0.25	0.40	0.55	0.60	0.60	0.45
Grass	0.11	0.26	0.60	0.69	0.92	0.99	0.60
Cellulose spray (1″)	0.08	0.29	0.75	0.98	0.93	0.76	0.75

to be reflective while those with NRC values greater than 0.40 are considered to be absorptive. When significant sound energy must be absorbed, as may be the case for eliminating echoes or standing waves, materials having higher absorption coefficients are usually recommended.

A few cautionary notes apply. NRC values are convenient to use for rating the absorption characteristics of a material. However, they should only be used when the sound sources of interest are within the 200–2000 Hz range. For sources outside of this range, and especially below this range, materials effective for the specific frequency of interest must be used. Also note that some manufacturers specify NRC and absorption coefficient values that are greater than 1.0. Methods used to measure absorption coefficients can artificially raise their values above 1.0; yet such values inaccurately imply that more energy is absorbed by a material than is incident upon it, which is a physical impossibility. Therefore, any published absorption coefficient or NRC values greater than 1.0 should be not be considered as greater than 1.0.

Another absorption metric gaining increasing application is the sound absorption average (SAA). The SAA is a single number rating that is the average, rounded off to the nearest 0.01, of the sound absorption coefficients of a material for the twelve one-third octave bands from 200 through 2500 Hz. Although the SAA replaces the NRC rating, as directed by the ASTM C 423 since the year 2000, most product literature still uses NRC values.

The method by which these materials are mounted affects their sound absorption effectiveness, especially for low frequencies (below 500 Hz). Figure 11.7 shows

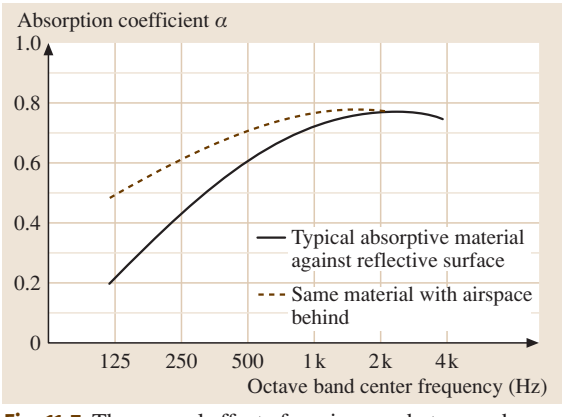


Fig. 11.7 The general effect of an air space between absorptive material and its mounting surface

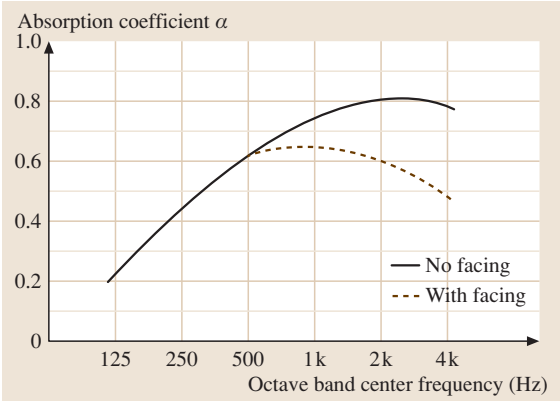


Fig. 11.8 The general effect of non-acoustically transparent facings on absorption

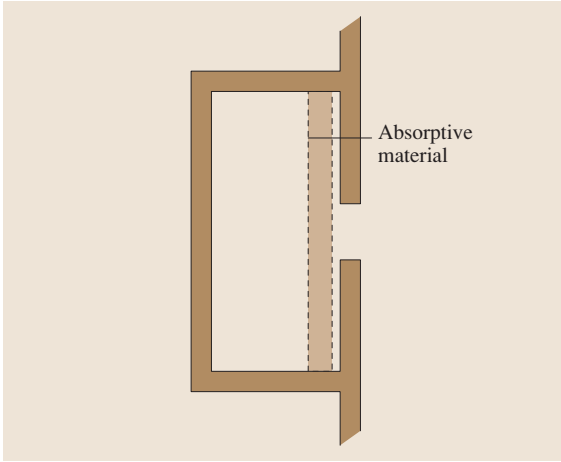


Fig. 11.10 Cross section of a Helmholtz resonator concrete masonry unit. The absorptive material helps to spread the absorptive properties over a larger frequency range than would otherwise occur

the effect of air spaces between an absorptive material and a mounting surface.

Facings on absorptive materials can degrade their absorptive properties, especially for higher frequencies (above 2000 Hz). Acoustically transparent facings, such as grill cloth and open-weave fabrics, will have little effect on absorption properties, but facing materials that are not acoustically transparent, such as perforated metal and wood slats, can produce the kind of effects shown in Fig. 11.8.

Of critical importance with these facings is the open area ratio. Since there are so many perforated metal designs, it is best to check with the manufacturer for the most appropriate opening ratio and design for the situation. Figure 11.9, however, offers some guidelines for wood-slat ceiling designs.

As Table 11.1 shows, the absorption coefficients of most porous materials increase with increasing frequency. This means that they are not as effective at low frequencies as at higher ones. If absorption is required for frequencies below 250 Hz, special materials must be used. Each of these materials has an air space behind its light or open surface to provide the extra absorption. Two common materials used for this purpose are Helmholtz resonators and diaphragmatic absorbers. Helmholtz resonators have narrow openings that lead to the outside on one end and into a larger air cavity on the other, as is shown in Fig. 11.10.

As viewed by an incoming sound wave, the air in the narrow neck functions as a mass and the air in the larger cavity functions as a spring. A mass on a spring will resonate at a frequency appropriate to that mass

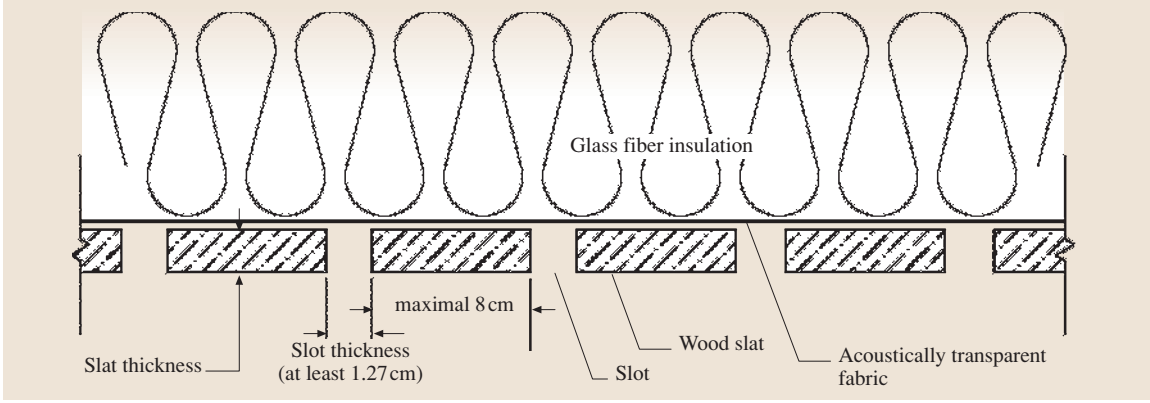


Fig. 11.9 General design guidelines for absorptive wood-slat ceilings

Amazon EX1014 Page 406

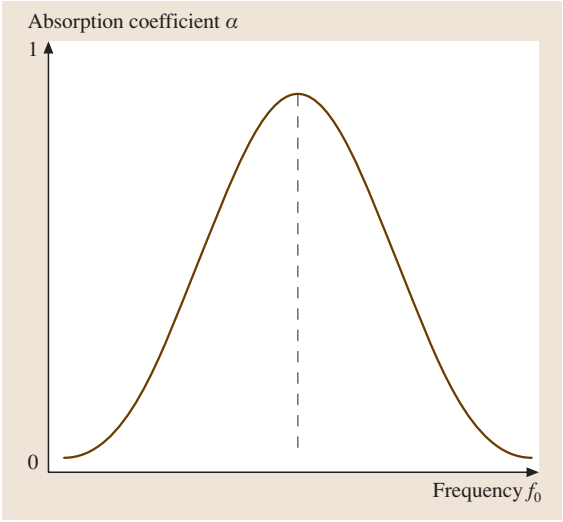


Fig. 11.11 General absorption characteristics of Helmholtz resonators not filled with absorptive material

and spring stiffness. Thus, at and near the resonant frequency of such an acoustic chamber, sound will be absorbed from an incoming sound wave, as is shown in Fig. 11.11.

Such devices incorporated in wall constructions are usually resonant below 250 Hz, depending on the dimensions of the neck and size of the cavity. There are commercially available products that incorporate this design into concrete masonry units. When viewing these products installed as partitions, the wall surfaces have slots in them, as are shown in Fig. 11.12. Table 11.1 has a listing for these products which shows their superior absorption in a narrow low-frequency range. The porous and coarse nature of the surface provides modest absorption at higher frequencies also.

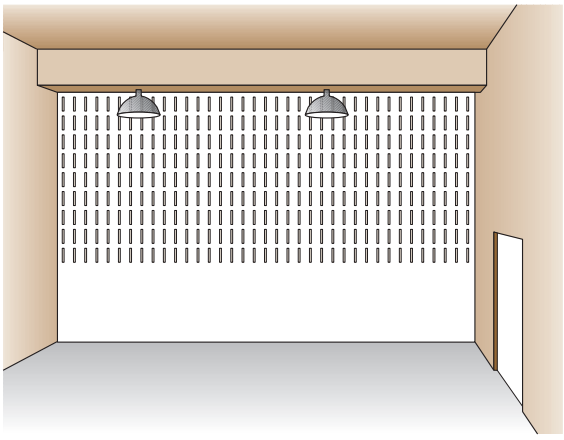


Fig. 11.12 Wall built from Helmholtz resonator concrete masonry units

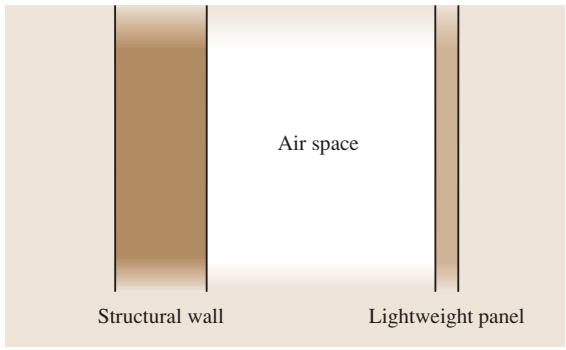


Fig. 11.13 Cross section of a generalized diaphragmatic absorber

Diaphragmatic absorbers work according to similar principles to Helmholtz resonators, except they are comprised of an air space behind a lightweight wall, as shown in Fig. 11.13.

Reverberant Field Noise Reduction

Adding absorptive materials to a room's surfaces, in addition to reducing or eliminating echoes and standing waves, can reduce the overall reverberant field noise level within a room. The mathematical basis for this is to determine the total absorption for all of a room's surfaces using the following equation:

$$A = \sum_{i=1}^{n} \alpha_i S_i = \alpha_1 S_1 + \alpha_2 S_2 + \ldots + \alpha_n S_n \,, \quad (11.3)$$

where α_1, α_2, etc., are the absorption coefficients of each surface material in the room while S_1, S_2, etc. are the surface areas corresponding to the surfaces having the same subscript value as those for the absorption coefficients. The total absorption (A) is in units of sabins, for which one sabin is defined as the total absorption provided by a one square foot piece of material having an absorption coefficient of 1.

The reverberant field noise reduction is defined as

$$\mathrm{NR} = 10 \times \log(A_2/A_1) \,, \quad (11.4)$$

where A_1 is the total absorption within a room before absorptive treatment is applied and A_2 is the total absorption within that same room after absorptive treatment is applied. The practical limits of NR are 12–15 dBA with typical cases yielding 6–8 dBA of noise reduction

Amazon EX1014 Page 407

within a room from the proper use of absorptive materials. Bear in mind that this will not reduce sound pressure levels close to a source but will only reduce reflected sound pressure levels, which limits the effectiveness of the noise reduction to locations in a room's reverberant field.

11.1.4 Reverberation

Absorption is useful in reducing or eliminating unwanted reflections off surfaces. The standing waves discussed in Sect. 11.1.1 can be eliminated by covering one of the parallel surfaces with absorptive material. Absorption can also be used to eliminate echoes. The rear walls of auditoriums are prime candidates for absorptive materials since rear walls have the greatest potential to cause echoes. The most common use of absorption, however, is to control reverberation.

Reverberation is the build up of sound within a room resulting from repeated sound wave reflections off all of its surfaces. Reverberation can increase sound levels within a room by up to 15 dBA, as well as distort speech intelligibility. Reverberation is desirable for rooms in which music is being played, especially classical and cathedral-style music, to add a pleasant persistence to the sounds. Therefore, there are different reverberation characteristics that would be appropriate for different room uses.

Reverberation is described by a parameter known as the reverberation time (denoted RT_{60}). RT_{60} can be defined in two ways – physically and mathematically. Physically, RT_{60} is the time (in seconds) that it takes for a sound source to reduce in sound pressure level (within a room) by a factor of 60 dB after that sound source has been silenced. Mathematically, in what is known as the sabine equation, RT_{60} is directly proportional to the volume of a room and inversely proportional to the absorption of the materials in the room, as follows:

$$RT_{60} = 0.161 \cdot V/A \,, \tag{11.5}$$

where V is the room volume in cubic meters and A is the total absorption of the room's surfaces in metric sabins.

This means that RT_{60} increases as the size of a room increases and as the absorption of the room's surfaces decreases. Conversely, RT_{60} decreases as the size of a room decreases and as the absorption of the room's surfaces increases. There are then two principal ways to control RT_{60}: (1) by changing a room's size and (2) by changing the amount of absorption on its surfaces. Although it is possible to reduce a large room's volume by dividing the space with walls, it is more practical to adjust RT_{60} by adding sound absorptive materials.

Table 11.2 Optimum mid-frequency RT_{60} values for various occupied facilities

Type of facility	Optimum mid-frequency RT_{60} (s)
Broadcast studio	0.5
Classroom	1.0
Lecture/conference room	1.0
Movie/drama theater	1.0
Multipurpose auditorium	1.3 to 1.5
Contemporary church	1.4 to 1.6
Rock concert hall	1.5
Opera house	1.4 to 1.6
Symphony hall	1.8 to 2.0
Cathedral	3.0 or higher

Since the absorption performance of materials varies with frequency, so does RT_{60}. Table 11.2 offers generally accepted ranges of RT_{60} for different uses in the mid-frequency (500–1000 Hz) range. As you can see from the table, lower RT_{60} values are desirable for rooms used mainly for human speech and higher RT_{60} values are desirable for rooms used mainly for music. The optimum mid-frequency RT_{60} for a fully occupied room is different for various types of music. Because contemporary orchestral repertoires emphasize late classical and romantic music, the optimum mid-frequency RT_{60} for a fully occupied concert hall is usually stated as 1.8–2.0 s.

Optimum RT_{60} values generally increase by 10% of the values in Table 11.2 for each halving of frequency below 500 Hz and decrease by 10% of the values in Table 11.2 for each doubling of frequency above 1000 Hz. A room with a low (less than 0.8 s) RT_{60} is called a *dead* room and a room with a high (greater than 1.7 s) RT_{60} is called a *live* room. Multipurpose facilities should have RT_{60} values between the live and dead range limits.

Note that the sabine equation is based on the assumption that the absorptive properties of the surface materials are spread evenly throughout the room. When this is clearly not the case, other (more complex) RT_{60} calculation methods would be more appropriate.

11.1.5 Effects of Room Shapes

Although absorption is necessary in many circumstances, eliminating reflections is not always useful. This

Amazon EX1014 Page 408

is of key importance in rooms where an audience is listening to a performance or lecture. In this type of room, it is desirable that all audience members hear the sound not only clearly, but without preference to seating location. Without an electronic sound system, this can only be accomplished by reflections off side walls and ceilings. Discrete echoes can be eliminated by avoiding smooth, flat reflective surfaces and by having irregular and convex surfaces to diffuse the sound evenly throughout the audience. For smaller rooms such as recording studios that require diffusion, commercially available sound-diffusing panels called quadratic residue diffusers are available. These panels can also be used for larger spaces, as well as irregularly shaped surfaces.

Concave reflective surfaces focus sound in certain areas and defocus sound from others, causing hot spots where sound is concentrated and dead spots where sound cannot be heard. Concave reflective surfaces should be avoided for this reason. If aesthetics dictate the need for a concave surface, it would be best to install an absorptive or diffusive surface (as needed) and cover it with acoustically transparent fabric in the concave shape.

Reflective rear walls in auditoriums are notorious for generating echoes because of their associated large sound travel path differences. For this reason, reflective surfaces should be avoided for rear walls. Reflective surfaces are beneficial, especially for concert halls, when they are close to the stage and along side walls. Reflective surfaces close to the stage assist in several ways, by sending sound into the audience rather than allowing it to be lost behind the stage and by enhancing the sound through lateral reflections off side walls to spread the sound more evenly throughout an audience. Another benefit of reflective surfaces near the stage is that they permit the performers to hear each other, something that is critical to concert performances. These so-called early reflections are usually generated by shells on the stage or by hanging reflective panels.

11.1.6 Sound Insulation

The description of the insulation of sound is similar in many ways to the description of the absorption of sound. As with absorption, there is a transmission coefficient that ranges from the ideal limits of 0 to 1. The transmission coefficient, denoted by the Greek letter τ, is the unitless ratio of transmitted to incident sound energy. Unlike the absorption coefficient, however, the limit of $\tau = 1$ is possible in practice since a transmission coefficient of 1 implies that all of the sound energy is transmitted through a partition. This would be the case

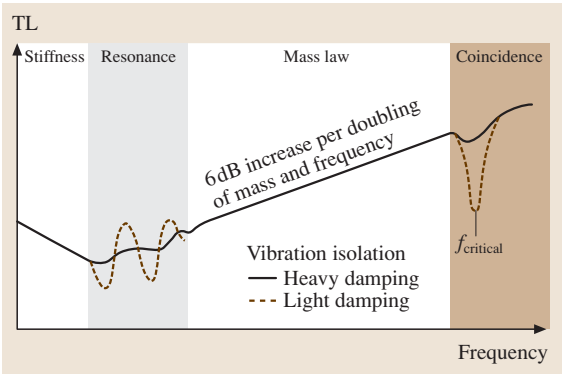


Fig. 11.14 General characteristics of the transmission loss spectrum

for an open window or door, where the sound energy has no obstruction to its path. The other extreme of $\tau = 0$ (implying no sound transmission), however, is not a practical value since some sound will always transmit through a partition.

Unlike absorption, the principal descriptor for sound insulation is a decibel level based on the transmission coefficient. This value is known as the transmission loss (TL), and is based on the following equation:

$$\mathrm{TL} = 10 \cdot \log(1/\tau)\,. \tag{11.6}$$

The transmission loss can be loosely defined as the amount of sound reduced by a partition between a sound source and a listener. The complete sound reduction of a partition between two rooms also takes into account the absorptive characteristics of the listener's room, as follows:

$$\mathrm{SPL_S} - \mathrm{SPL_L} = \mathrm{TL} + 10 \cdot \log(A_\mathrm{L}/S)\,, \tag{11.7}$$

where $\mathrm{SPL_S}$ is the average sound pressure level in the room enclosing the sound source, $\mathrm{SPL_L}$ is the average sound pressure level in the adjacent listener's room, A_L is the total absorption in the listener's room, TL is the transmission loss of the partition between the two rooms, and S is the surface area of the partition between the two rooms. Note that TL is the quantity that is typically reported in manufacturers' literature since it is measured in a laboratory independent of the installation.

Since the logarithm of 1 is 0, the condition in which the transmission coefficient is 1 translates to a TL of 0 dB. This concurs with the notion that an open air space in a wall allows the free passage of sound. Although an open air space itself would have a TL value of 0 dB associated with it, a wall with an open air space in it

Amazon EX1014 Page 409

Table 11.3 Critical frequencies for common building materials

Material	Thickness (cm)	Critical frequency (Hz)
Concrete	8	100
Plywood	1.2	1700
Gypsum wall board	1.2	3100
Steel or aluminum	0.3	4100
Lead	1.2	4400
Glass	0.3	4900
Plexiglass	0.3	9800

would have a TL value greater than 0 (up to 10) dB, depending on the size of the opening and the location with respect to the wall. The practical upper limit of TL is roughly 70 dB.

As for absorption, TL is frequency dependent. Typical partitions have TL values that increase with increasing frequency, as is shown in Fig. 11.14, with the exception of a dip in TL around a bending wave resonance frequency, known as the critical frequency. This effect is known as the coincidence dip. The extent of this dip depends on the vibrational damping of the partition and the frequency at which it occurs depends on the density and thickness of the material.

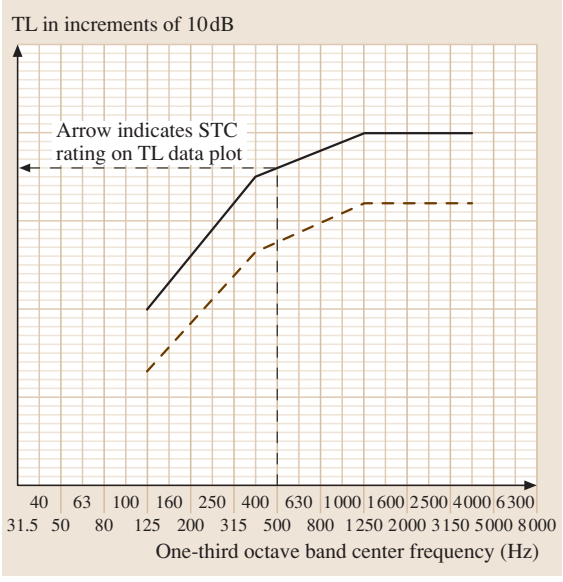


Fig. 11.15 The standard STC calculation curve, from ASTM E 413. STC values are derived by plotting the TL spectrum on this curve and determining the highest curve shape that meets the criteria of: (1) the sum of the differences between individual TL values and the *solid line* is less than 32 dB and (2) no individual 1/3-octave band TL value is more than 8 dB below the *solid line*. The TL value at 500 Hz corresponding to this spectrum is the STC value

Table 11.3 lists critical frequencies for common building materials. Critical frequencies generally decrease at the same rate as the material thickness increases. Therefore, a doubling in material thickness translates to cutting the critical frequency in half for the same homogeneous material. Coincidence effects can be minimized by using multilayered partitions with different materials and different material thicknesses combined into a single partition.

There is a single-number rating for TL that takes the entire frequency spectrum into account, established by ASTM standard E413. This value, known as the sound transmission class (STC), is not derived by the simple averaging method used for NRC values. Instead, the TL frequency spectrum is matched to a standard curve within the limits imposed by the ASTM standard. Figure 11.15 shows how STC is calculated from the TL spectrum. Note that STC values have no units associated with them, but they are based on decibels. Similar to NRC, STC is useful to describe the sound insulation efficiency of a partition over the human speech frequency range of roughly 500–2000 Hz. If sound insulation outside of that frequency range (especially for frequencies below 250 Hz) is required, the TL values relevant to the frequency range of interest must be used.

Table 11.4 lists TL and STC values for common partitions. As with Table 11.1, these values are for general reference purposes only, and specific values should be based on manufacturers' specifications. Table 11.5 gives some meaning to the numbers by listing speech privacy ratings for different ranges of STC values.

As with any construction system, the actual in-field performance seldom matches the theoretical or laboratory STC rating for the construction. This is most often caused by flanking paths, poor construction practices, and/or lack of sealing cracks, gaps and openings. The noise isolation class (NIC) is the metric that considers these conditions. NIC is determined using the same method and template as is used for determining STC (shown on Fig. 11.15), using measured field data instead of using controlled laboratory data. The NIC value is typically 5–10 points less than the STC rating of a given partition design.

STC does not effectively consider the transmission of impact noise on floors from foot traffic as it may affect floors below. For this purpose the impact insulation class (IIC) was established by ASTM standard

Amazon EX1014 Page 410

Table 11.4 Transmission loss and STC values for common partitions

Partition	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	STC
1/2 inch drywall on both sides of wooden studs	17	31	33	40	38	36	33
1/2 inch drywall on wooden studs with 2 inches of insulation	15	30	34	44	46	41	37
Double layer of 1/2 inch drywall on wooden studs	25	34	41	51	48	50	41
1/2 inch drywall on staggered wooden studs	23	28	39	46	54	44	39
1/2 inch drywall on staggered wooden studs with 2 inches of insulation	29	38	45	52	58	50	48
1/2 inch drywall on metal studs	22	27	43	47	37	46	39
1/2 inch drywall on metal studs with 2 inches of insulation	26	41	52	54	45	51	45
8 inch thick concrete masonry units	36	44	50	54	58	56	53
Open-plane office partition	10	12	12	12	12	11	12
4 inch thick brick wall	32	34	40	47	55	61	45
1/2 inch drywall inside/1 inch stucco outside on wooden studs	21	33	41	46	47	51	42
Single-paned 1/8 inch thick glass	18	21	26	31	33	22	26
1/2 inch thick laminated glass	31	34	38	40	37	46	40
Double-paned 1/8 inch thick glass with 2 inch air gap	13	25	35	44	49	43	37
Hollow wooden door, 1 3/4 inch thick	14	19	23	18	17	21	19
Solid wooden door, 1 3/4 inch thick	29	31	31	31	39	43	34
Hollow metal door, 1 3/4 inch thick	24	23	29	31	24	40	28
Filled metal door, 1 3/4 inch thick	26	34	40	48	44	52	43
Wood joist floor/ceiling with 1/2 inch plywood subfloor and 1/2 inch drywall	23	32	36	45	49	56	37
8 inch thick concrete slab floor	32	38	47	52	57	63	50
Wood plank shingled roof	29	33	37	44	55	63	43
Wood plank shingled roof with 1/2 inch drywall ceiling, 4 inches of insulation	35	42	49	62	67	79	53
Corrugated steel roof with 1 inch of sprayed cellulose	17	22	26	30	35	41	30

E 989. The IIC stresses the low-frequency range, as is shown in Fig. 11.16, but note that IIC does not address very low (below 100 Hz) frequencies, which may be associated with resonances of a building's structural components.

One other single-number rating worth noting here is the outdoor–indoor transmission class (OITC), established in ASTM E 1332. The OITC is a single-number rating for the effectiveness of a building façade in reducing noise to interior spaces. The calculation of OITC is based on an A-weighted source spectrum of typical transportation (aircraft, rail, and traffic) sources in the 80–4000 Hz range and the façade's transmission loss spectrum in the same frequency range.

Non-Homogeneous Partitions

Up to this point we have been discussing homogeneous partitions. Actual designs, however, include walls composed of different materials, such as those with windows and doors. Each of these wall components has different associated TL characteristics. Wall components having lower TL characteristics than the rest of the wall can significantly degrade a wall's sound reduction effectiveness. This can be calculated using the composite

Amazon EX1014 Page 411

Table 11.5 Speech privacy associated with STC ratings*

STC range	Sound privacy
0 to 20	No privacy (voices heard clearly between rooms)
20 to 40	Some privacy (voices heard in low** background noise)
40 to 55	Adequate privacy (only raised voices heard in low background noise)
55 to 65	Complete privacy (only high level noise heard in low background noise)
70	Practical limit

* assuming no significant flanking paths or openings in walls
** in the 35 dBA range

Table 11.6 Transmission loss reduction as a function of air opening*

Wall area having air opening (%)	Resultant wall TL (dB)	Resultant reduction in TL (dB)
0.01	39	6
0.1	30	15
0.5	23	22
1	20	25
5	13	32
10	10	35
20	7	38
50	3	42
75	1	44
100	0	45

* based on original wall TL of 45 dB

transmission coefficient, which is defined as:

$$\tau_{\text{comp}} = \left(\sum_{i=1}^{n} \tau_i S_i\right) / S$$
$$= (\tau_1 S_1 + \tau_2 S_2 + \ldots + \tau_n S_n)/S\,, \qquad (11.8)$$

where τ_1, τ_2, etc., are the transmission coefficients of each wall component while S_1, S_2, etc. are the surface areas corresponding to the surfaces having the same subscript value as those for the transmission coefficients, and S is the surface area of the entire partition. The composite transmission coefficient that is calculated using (11.8) can then be replaced in (11.6) to yield the composite transmission loss for the partition.

Air gaps (e.g., around doors, windows or penetrations) are notorious for compromising the sound reduction effectiveness of walls. Table 11.6 shows the TL degradation caused by varying sizes of air gaps in a wall originally rated at a TL of 45 dB. As the table shows, an air gap just one tenth of one percent the size of a wall can lower the TL rating from 45 to 30 dB. Figure 11.17 offers a graphical method for making this determination, not only for openings but also for doors and windows that have lower TL values than the walls in which they are installed. This emphasizes the importance of sealing walls, avoiding air gaps, and placing airtight seals around doors.

A rule that governs most of the TL spectrum of homogeneous partitions, known as the mass law, states that TL increases at a rate of 6 dB with each doubling of mass and with each doubling of frequency (shown in Fig. 11.14). This rule can make it impractical to solve sound privacy issues with homogeneous partitions. For example, if a 1/2 meter-thick concrete wall does not provide enough sound insulation for a specific situation,

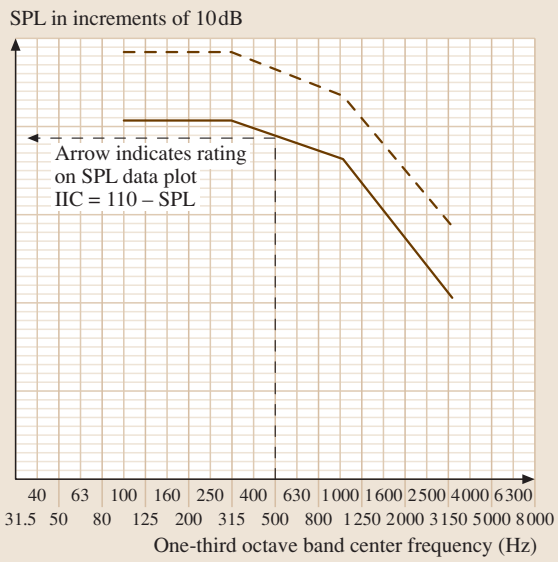


Fig. 11.16 The standard IIC calculation curve, from ASTM E 989. IIC values are derived by plotting the SPL spectrum (from a tapping machine specified in ASTM E 492) on this curve and determining the lowest curve shape that meets the criteria of: (1) the sum of the differences between individual SPL values and the *solid line* is less than 32 dB and (2) no individual 1/3-octave band TL value is more than 8 dB above the *solid line*. The SPL value at 500 Hz corresponding to this spectrum, subtracted from 110, is the IIC value

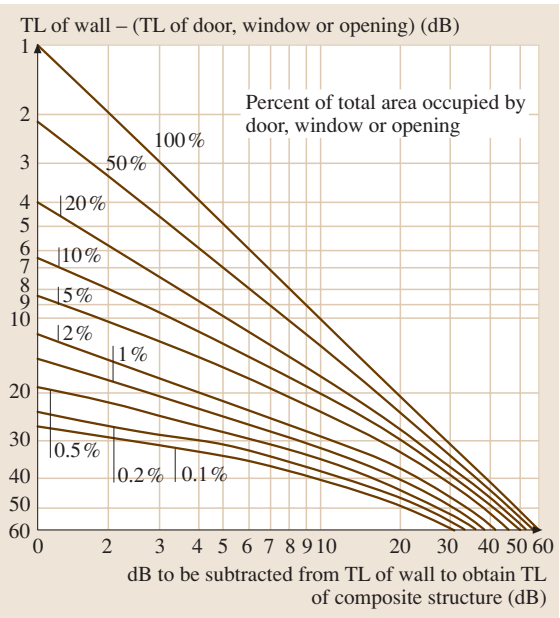


Fig. 11.17 Chart for determining the reduction in TL wall rating by the installation of a door, window, or opening

doubling the thickness of that wall to 1 meter would only offer an additional 6 dB of sound reduction. A more practical way of avoiding this kind of issue is by using multilayered partitions.

Multilayered partitions are comprised of layers of different materials. Each time sound passes through a different material, its level is reduced. Therefore, this method can be used to reduce costs, weight and space restrictions while providing adequate sound reduction. A sharp change in density of material is most effective in raising TL. Air spaces between wall sections and materials are effective by setting up such environments and also breaking any rigid connections between sides of a partition. A rigid connection can provide a vibration channel for sound to pass through with little reduction. For example, the noise reduction effectiveness of a studded wall filled with fiber insulation between studs can be short-circuited because sound may travel unimpeded

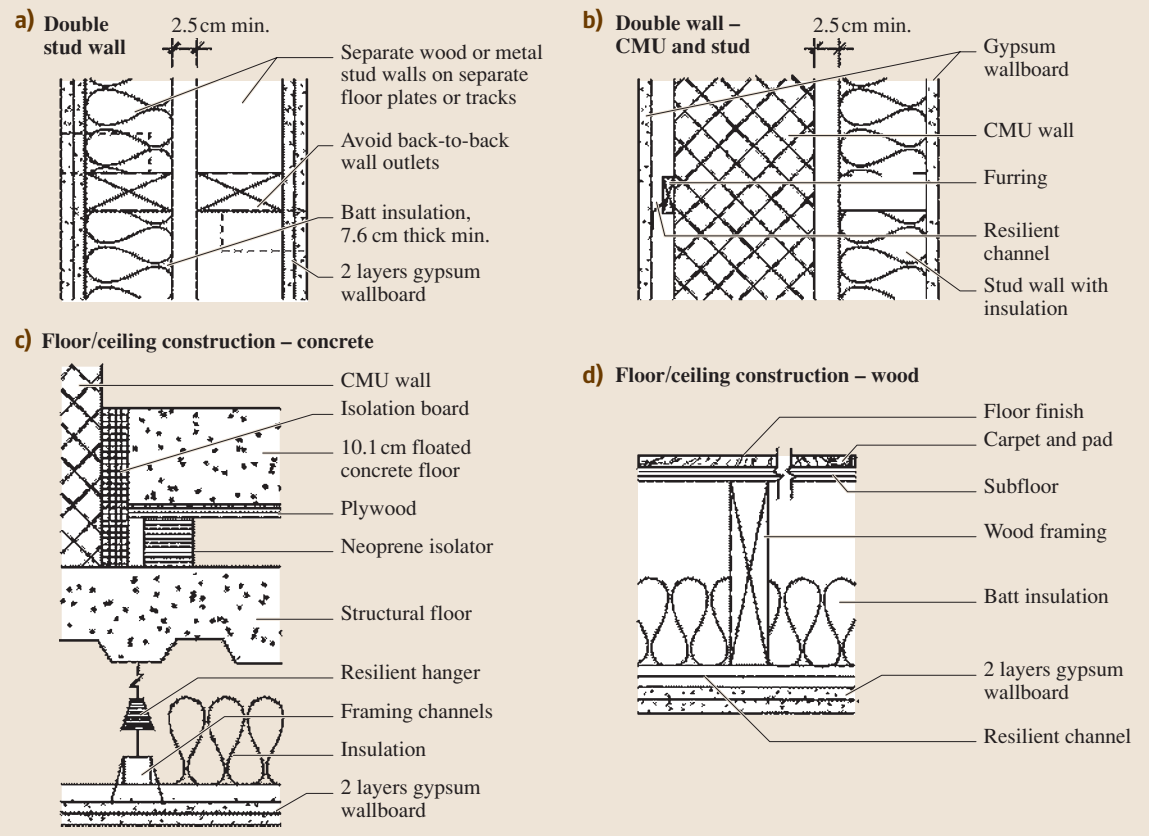


Fig. 11.18 Examples of multilayered partition and floor/ceiling designs that result in high TL

Amazon EX1014 Page 413

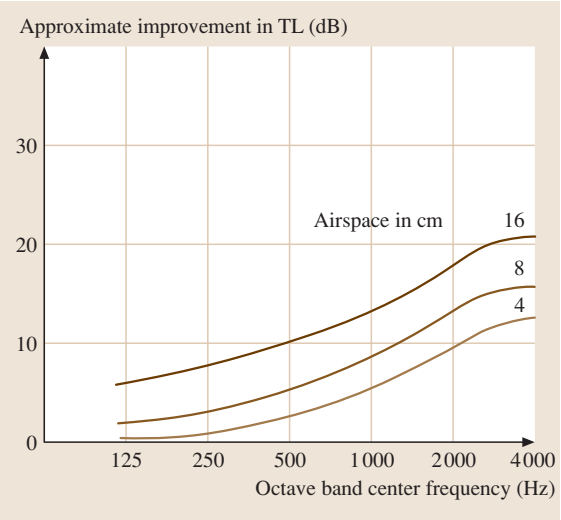


Fig. 11.19 General TL improvement that can be expected from air spaces in partitions

through the studs to the other side of the wall. Figure 11.18 shows examples of multilayered partitions that are effective for sound insulation.

Figure 11.19 shows the TL benefit of different air spaces in partitions. In general, the TL of a partition increases at a rate of 5 dB with each doubling of air space (within a partition) beginning with an air space of 5 cm.

As is mentioned above, transmission loss generally increases with frequency. If significant sound reduction is required for frequencies below 250 Hz, it is most practical to use multilayered partitions with air spaces and staggered studs or resilient channels between components. Resilient channels are shaped metal bars (the S shape is common) or resilient clips that connect two wall components. Only one side of their shape touches each side of the wall, and they effectively reduce vibrational energy from traveling through them. Resilient clips resemble isolation hangers, which are discussed later in this chapter. Another option for greater transmission loss at low frequencies is massive concrete slabs.

11.2 General Noise Reduction Methods

Noise can be controlled at its source, in the path between the source and the listener, or at the listener. Table 11.7 summarizes the general options available. If the noise can be controlled sufficiently well at its source, it is unnecessary to consider the path or listener locations. Likewise, if the noise can be controlled sufficiently well in the path between the source and listener, it is unnecessary to consider the listener's location for noise control measures.

The options for noise control at the source are generally self-explanatory. Most often, due to economic or logistic issues, noise control options are limited to the path between the source and the listener. Given many misconceptions about these options, it is useful to discuss some of them further.

11.2.1 Space Planning

The most effective noise control method for buildings is space planning. This is true both for room layout within a building and for the placement of the building itself with respect to exterior noise sources. Inherently noisy spaces adjacent to spaces inherently requiring quiet should be avoided as much as possible. The noisiest indoor spaces tend to be mechanical rooms, restrooms, and elevator shafts.

For placement with respect to loud exterior sources, buildings should be as far as possible from noisy streets or rail lines, taking advantage of shielding from other buildings and topographical features. If building placement near noisy outdoor sources is unavoidable, special attention should be paid to the design of the façade(s) facing the noisy sources, avoiding operable windows, doors, and penetrations.

Another aspect to consider for outdoor sound is the potential effect of a building's noise sources on the surrounding community. Rooftop mechanical units are the most common noise sources that may need to be insulated from community intrusions, especially when municipal codes are involved. The community noise aspect of a building is often considered in the building's permitting process.

11.2.2 Enclosures

Enclosures can be effective at reducing noise levels as long as they are properly designed. The following points should be considered when designing noise enclosures (also illustrated in Fig. 11.20a–d).

1. The enclosure must completely surround the noise source, with no air gaps. As mentioned earlier in

Table 11.7 Generalized noise reduction options

Control at the source	Control in the path	Control at the listener
Maintenance	Enclosure(s)	Relocate listener
Avoid resonance	Barrier(s)	Enclose listener
Relocate source/ space planning	Muffler(s)	Hearing protection
Remove unnecessary sources	Absorptive treatment	Masking
Use quieter model(s)	Vibration isolation	
Redesign source to be quieter	Active noise control	

this chapter, air gaps can significantly compromise the noise reduction effectiveness of partitions. An enclosure with any side open is not an enclosure but a barrier, and the noise reduction effectiveness of barriers is limited by diffraction to 15 dBA, independent of the barrier material (as long as the material would provide at least 20 dBA of noise reduction on its own). Enclosures, on the other hand, can provide up to 70 dBA of reduction.

2. The enclosure must be isolated from floors or any structural members of a building. An enclosure covering the sides and top of a noise source but not the bottom (since the source is sitting on the floor or ground) can compromise its effectiveness for several reasons. First, the chances of the sides of the enclosure perfectly sealing to the ground are slim, and therefore, air gaps would result. Second, vibrations will be carried along the ground or floor since the source is in direct contact with it. The only way to reduce these vibrations is to vibrationally isolate the source from the ground or floor using tuned springs (appropriate for the source), pads, or a floating floor.
3. The enclosure should be constructed using nonporous materials. Sound absorptive material can be effective in reducing noise when it is used as part of a multilayered enclosure (on the inside); however, absorptive material by itself is not effective in reducing noise.

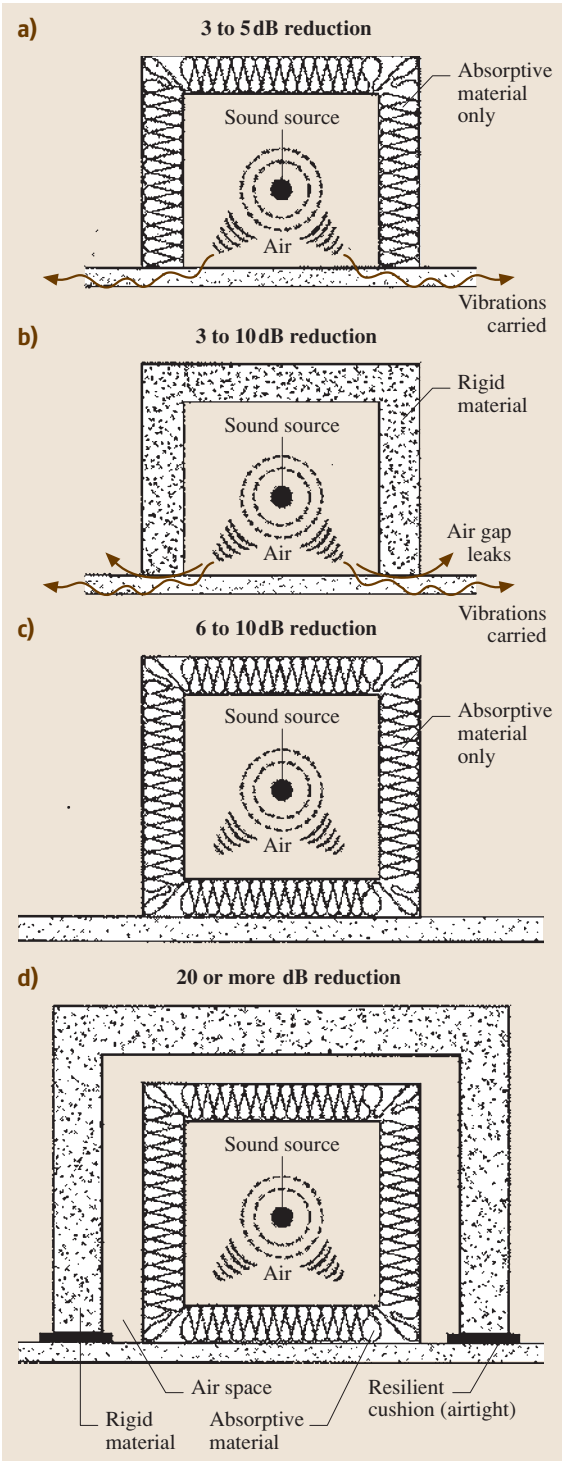


Fig. 11.20a–d Noise reduction effectiveness of enclosures sitting on a solid surface **(a)** enclosing the source on five sides with absorptive material only, **(b)** enclosing the source on five sides with one layer of rigid, nonporous material, **(c)** completely enclosing the source with absorptive material only, and **(d)** surrounding the source in an isolated, multilayered enclosure

Amazon EX1014 Page 415

4. The enclosure designer must consider that some sources require ventilation. This cannot translate to leaving an opening in the enclosure without compromising the noise control effectiveness of the enclosure. Ventilation systems must be developed that minimize noise transmission.
5. The enclosure should be constructed using multilayered construction for maximum efficiency. As is mentioned earlier in this chapter, doubling the mass of an enclosure would add 6 dB to its noise reduction effectiveness. This can easily lead to excessive weight for an effective homogeneous enclosure. As for single partitions, multilayered enclosures can add more than 20 dB of effectiveness (under similar space requirements) to massive enclosures with a fraction of the weight.

11.2.3 Barriers

A barrier is contrasted from an enclosure by being open to the air on at least one side. Because of diffraction, noise barriers are limited to 15 dBA of noise reduction capability, independent of the material. This limited effectiveness is compromised even more if there are reflective ceilings above the barrier because sound reflected off the ceiling minimizes the barrier's effectiveness. Therefore, wherever noise barriers are used indoors, an absorptive ceiling should be installed above them. It is also important to have no air spaces within or under the barriers, since this will compromise their already limited effectiveness.

The noise reduction effectiveness of barriers is typically rated by the insertion loss (denoted by IL). The IL is the reduction in sound pressure level, at a specific location, resulting from the installation of a barrier. The IL is then the difference between conditions with and without a barrier.

To provide any insertion loss, a barrier must break the line of sight between the sound source and listener. In other words, if you can see a sound source on the other side of a barrier, that barrier is providing no sound reduction (from that source) for you. Breaking this line of sight typically provides a minimum of 3–5 dBA of insertion loss, with insertion loss increasing as one goes further into the shadow zone of the barrier.

Bear in mind that a row of trees is ineffective for outdoor noise control. Only solid walls or earth berms breaking the line of sight between the noise source and building occupants will produce any meaningful noise reduction.

11.2.4 Mufflers

Mufflers are devices that are inserted in the path of ductwork or piping with the specific intention of reducing sound traveling through that conduit. The effectiveness of mufflers is typically rated using insertion loss. Mufflers must be designed for each purpose to preserve the required static or back pressure characteristics of the equipment and to reduce noise in the appropriate frequency range(s). For that reason, each muffler is unique to its installation.

11.2.5 Absorptive Treatment

Absorptive treatment within a room can reduce reverberation and, in this process, reduce noise levels by up to 10 dBA. Bear in mind, though, that absorptive treatment is only effective for reducing reverberation within a room and not for reducing transmission of sound between rooms.

11.2.6 Direct Impact and Vibration Isolation

Mechanical equipment can generate vibration that can travel through a building's structural members to affect remote locations within a building. It is therefore prudent to isolate any heavy equipment from any structural members of buildings. This can be accomplished by mounting the equipment on springs, pads, and/or inertia blocks; however, the selection of specific isolating devices (especially springs) should be performed by a specialist trained in vibration analysis. The main reason for this is that each vibration isolation device is tuned to a specific frequency range. If this is not matched properly with the treated equipment, the devices can amplify the vibration and cause more of a problem than would have occurred without any treatment.

Mechanical equipment and, to a lesser degree, footfalls generate direct impacts that should also be addressed by properly isolating floors from building structures.

11.2.7 Active Noise Control

Passive noise control involves all of the noise control methods discussed above, in which the sound field is not directly altered. Active noise control involves electronically altering the character of the sound wave to reduce its level. In this case, a microphone measures the noise and a processor generates a mirror image of

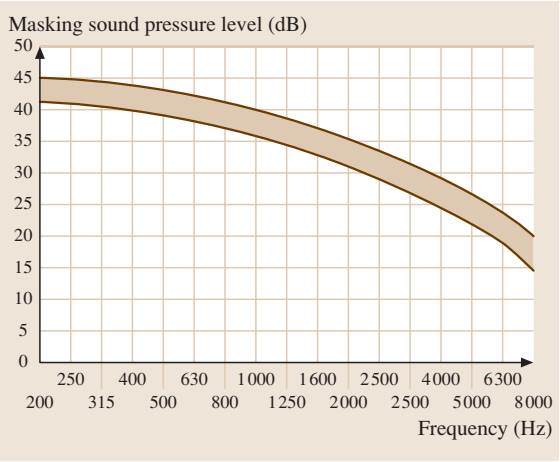


Fig. 11.21 Typical characteristics of effective masking systems (frequency response range in *shaded area*)

(180° out of phase from) that source. This mirror image is then reproduced by a loudspeaker in the path of the original sound. This new sound cancels enough of the original signal to reduce levels by up to 40 dB under the appropriate conditions. Although this is a very powerful noise control tool, active noise control is only practical in local environments and for tones below 500 Hz. Ventilation ducts provide ideal environments for active noise control systems because they are enclosed environments and their noise is often dominated by low frequency tones (associated with the fan characteristics).

11.2.8 Masking

As long as background sound levels are low in a building, one way of easing a noise problem is to add a more pleasing sound to the environment to make the noise less noticeable. This is especially desirable in open-plan offices where people can clearly hear activities in other offices and areas (since their offices are not completely enclosed). Any desirable sound can provide masking, but most often electronic masking systems are used, comprised of loudspeakers placed between

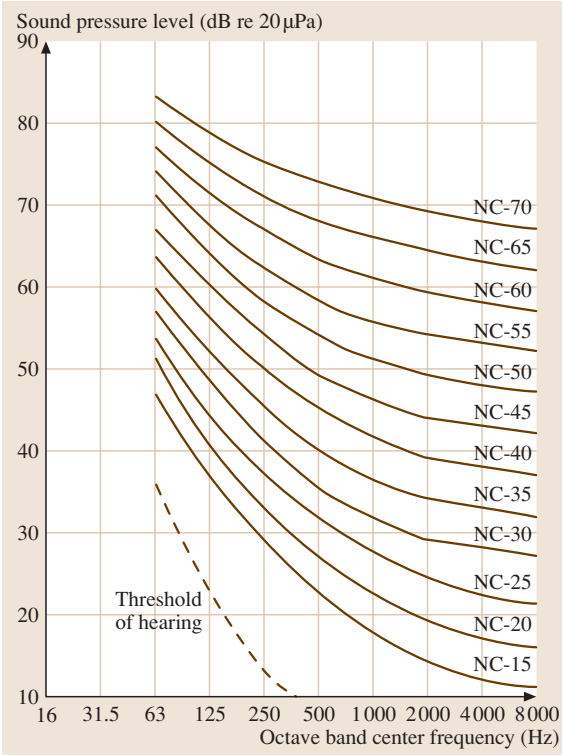


Fig. 11.22 NC curves

dropped ceilings and structural ceilings. These loudspeakers are connected to signal processors that are set to generate sounds similar to those generated by typical heating, ventilating and air conditioning (HVAC) systems. Although many people think of masking system sounds as white noise, which has an equal amount of energy in all audible frequencies, a typical masking system frequency response is more like that shown in Fig. 11.21, where less emphasis is placed on higher frequencies. Whatever the response, it is advisable to have an electronic masking system installed by a contractor with experience in this area. If the system is set at too high a level or with a harsh frequency response, the resulting environment may be more unpleasant with the masking system than without.

11.3 Noise Ratings for Steady Background Sound Levels

There are two standardized methods for rating the background sound levels inside rooms. Each of these methods is referenced in ANSI standard S12.2 and the *American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Handbook*. These are the balanced noise criterion (NCB) and the room criterion (RC)

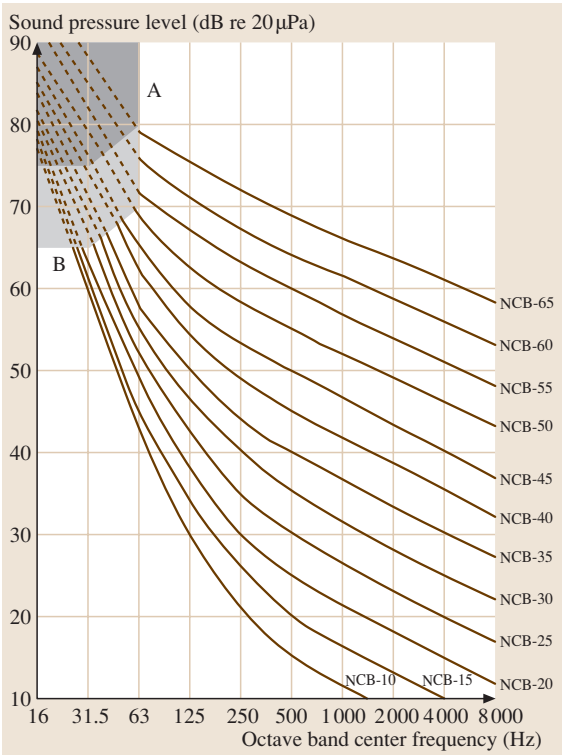


Fig. 11.23 NCB curves. Sound in the *region labeled A* can cause perceptible noise-induced vibration such as rattling of doors, windows, or fixtures. Sound in the *region labeled B* may generate lower levels of these noise-induced vibrations

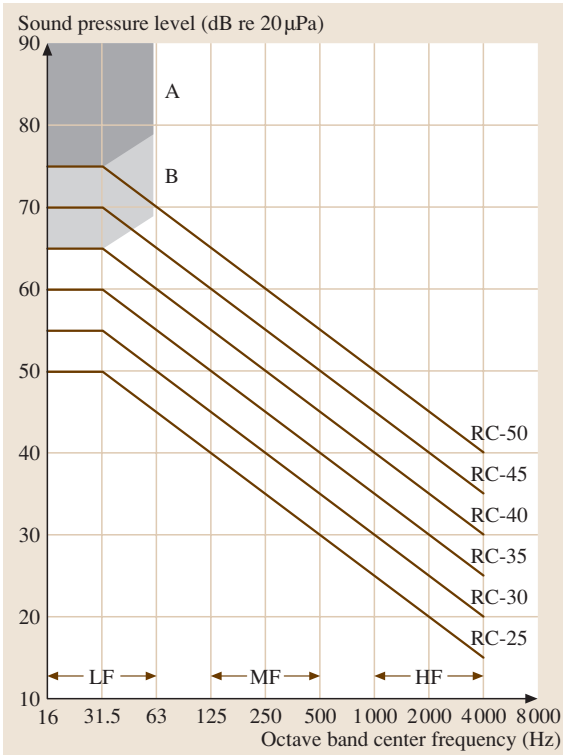


Fig. 11.24 RC curves. The *A and B regions* have the same meanings as the NCB curves in Fig. 11.23. The LF, MF, and HF regions are designated as low-frequency (rumble), medium-frequency (roar), and high-frequency (hiss) ranges, respectively

curves. The NCB curves are updated versions of the widely used noise criterion (NC) curves, different in that they are extended down to the 16 and 31.5 Hz octave bands. Figure 11.22 shows the NC curves and Fig. 11.23 shows the NCB curves.

Table 11.8 Recommended background noise criteria for common spaces

Category of space	Specific uses	NC, NCB or RC(N) range	dBA limit
Sensitive listening spaces	Broadcast and recording studios, concert halls	15 to 20	25 dBA
Performance spaces	Theaters, churches, video and teleconferencing	20 to 25	30 dBA
Presentation spaces	Large conference rooms, small auditoriums, movie theaters, courtrooms, meeting and banquet rooms, executive offices	25 to 30	35 dBA
Private spaces	Offices, small conference rooms, classrooms, private residences, hospitals, hotels, libraries	30 to 35	40 dBA
Public spaces	Restaurants, lobbies, open-plan offices, clinics	35 to 40	45 dBA
Service and support spaces	Computer equipment rooms, public circulation areas, arenas, convention centers	40 to 45	50 dBA

As for the room criterion (RC) method, the NCB method was developed to better consider low-frequency noise generated by HVAC systems. In each case, standardized curves are used in conjunction with measured noise levels in a room to yield a single-number rating for HVAC noise. The NC and NCB rating values result from plotting the measured octave band sound pressure level spectrum on each chart and noting the lowest standard curve that is not exceeded by the measured values. This single number is then rated against design guidelines established in the two references above. Determining the RC rating is more complicated, based on the 1000 Hz value of the standard RC curve (in Fig. 11.24) that matches the arithmetic average of the 500, 1000, and 2000 Hz values of the SPL room spectrum. The quality of the sound is also rated by the RC method, based on the deviations of the measured SPL spectrum from the standard RC curve. A *neutral* spectrum, designated by *N* and used in most RC specifications, implies that levels below 500 Hz do not exceed an RC curve by more than 5 dB and levels above 1000 Hz do not exceed an RC curve by more than 3 dB. When deviations greater than these exist, the RC rating is given the designation of the frequency range (from Fig. 11.24) in which the greatest deviations occur. The levels in the octave bands below 500 Hz are much lower for RC curves than those for the NCB curves. The purpose of this lower limit is to consider fluctuations or surges at the ventilation outlets better. These fluctuations are not considered for NCB ratings.

Table 11.8 lists generally recommended NC, NCB and RC(N) ranges for common spaces.

11.4 Noise Sources in Buildings

The most common noise sources in buildings, other than the inhabitants, are related to HVAC systems, plumbing systems, electrical systems, and exterior sources. Figure 11.25 shows the most common paths for intrusive noise in buildings.

Although there is legislation that often deals with noise disturbances directly related to people as sources, these other sources are typically unregulated and their associated noise levels are governed by the desires of developers, builders, and occupants.

11.4.1 HVAC Systems

Noise sources associated with HVAC systems can be separated into two general categories – mechanical equipment and duct-borne/airflow noise.

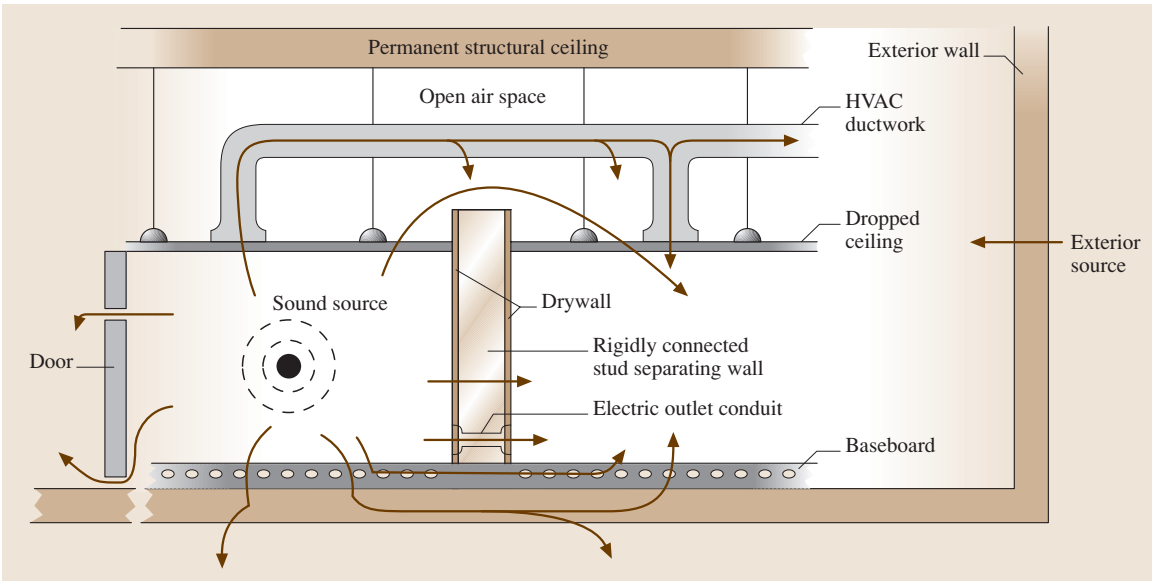


Fig. 11.25 Common noise leaks occur through these nine paths

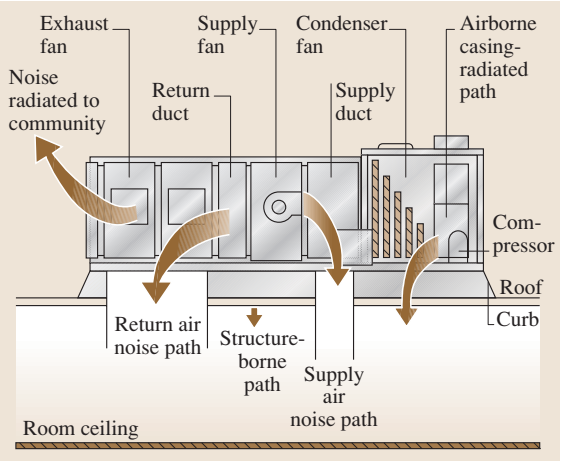


Fig. 11.26 Common noise paths for a typical rooftop HVAC unit

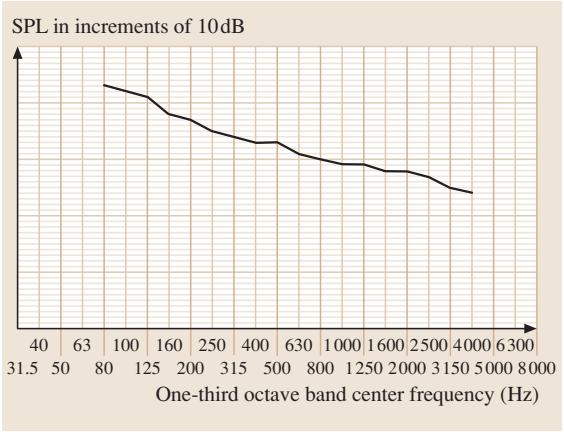


Fig. 11.27 Average transportation noise spectrum shape (adapted from data in ASTM E 1332)

Mechanical Equipment

Common mechanical equipment associated with HVAC systems includes pumps, compressors, chillers, generators, and air handlers. Rotating components of these pieces of equipment, such as gears and fans, generate most of the noise that causes concerns in buildings. Rotating mechanical equipment typically generate tones, their frequencies being associated with their rotational speeds. When mechanical equipment is housed in rooms within buildings, its associated noise can affect rooms throughout a building. When mechanical equipment is placed on rooftops or slabs outside buildings, its associated noise can also affect the surrounding communities. Figure 11.26 shows noise sources and paths for a typical rooftop HVAC unit.

Mechanical equipment not only generates noise in buildings, but also generates vibrations that can propagate throughout a building's structural members if not properly isolated. These vibrations can excite building members far from the sources and cause remote building components to rattle and generate their own noise.

Duct-Borne/Airflow Noise

Air is typically carried throughout a building using a system of ductwork. Ducts carry the tones generated by fans and they cause additional noise by inducing turbulence in the airflow. Some of this noise is carried through the ductwork to rooms in buildings (out of grilles in each room) and some of it is radiated directly from the duct walls (known as breakout noise).

11.4.2 Plumbing Systems

The most common plumbing system noise sources are water flowing through pipes and noise radiating from the walls of pipes.

11.4.3 Electrical Systems

The most common electrical system noise sources are transformers and noise radiated from associated conduit (by carrying vibrations that excite walls).

11.4.4 Exterior Sources

The most common exterior noise sources that affect building inhabitants are those associated with transportation (such as vehicular traffic, rail, and aircraft), industrial operations, and mechanical equipment from nearby buildings or mounted outside the same building. Figure 11.27 shows an average spectrum shape for transportation noise sources used by ASTM E 1332. This shows the concentration of sound in the low frequency (below 500 Hz) range, emphasizing the need for exterior-wall noise control design measures that effectively address low frequencies.

Amazon EX1014 Page 420

11.5 Noise Control Methods for Building Systems

The principles discussed in Sect. 11.2 can be put into practice using the general design concepts listed in this section. Specific dimensions and sound ratings have intentionally been omitted from the drawings in this section because these drawings stress the design principles rather than specific installations. These design concepts can be used to develop specific designs for each project.

11.5.1 Walls, Floor/Ceilings, Window and Door Assemblies

A key issue with all of these designs is to stop the path of sound and vibration through rigid connections between building components. Generous use of air spaces and resilient materials is the most effective method for this to occur. Another key issue is avoiding openings in and around partitions (for the reasons stated in Sect. 11.1.6). All openings must therefore be completely sealed. These two themes will be repeated in most of the designs listed in this section.

Walls

The most effective wall system for noise reduction is a double-wall system, in which wall sections (including studs) are separated by an air space. The larger the depth of the air space, the more effective the wall will be in controlling sound. Double walls can be especially effective for controlling low-frequency sound, as long as the walls are properly sealed and isolated from the building structure.

Figure 11.28 shows the general concept of a double-studded wall that is effective for noise control. Although metal studs are shown in the drawing, wooden studs would provide comparable results in this design because of the air gap. Figure 11.29 shows a full floor-to-structural-ceiling double-wall construction concept.

Note that the wall must be sealed with a non-hardening material at its perimeter to avoid any sound leakage through air gaps (since a hardening material may crack and leave gaps). Figure 11.30 shows a double-wall concept using a concrete masonry unit (CMU) and studded component.

All of these double-wall designs will provide STC values in excess of 55; however, if low-frequency noise

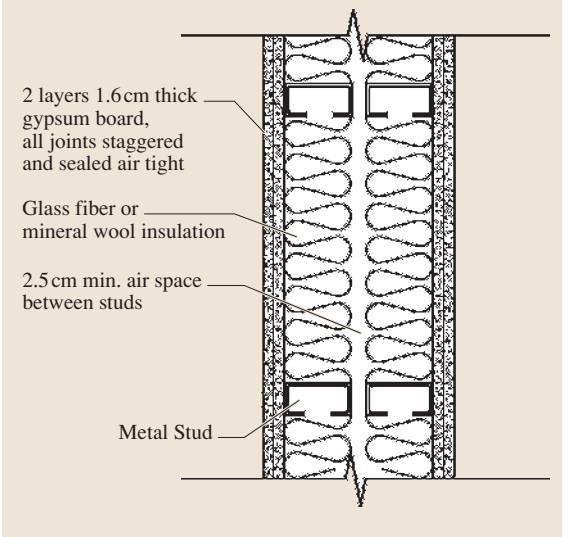


Fig. 11.28 Cross section of a double-wall design

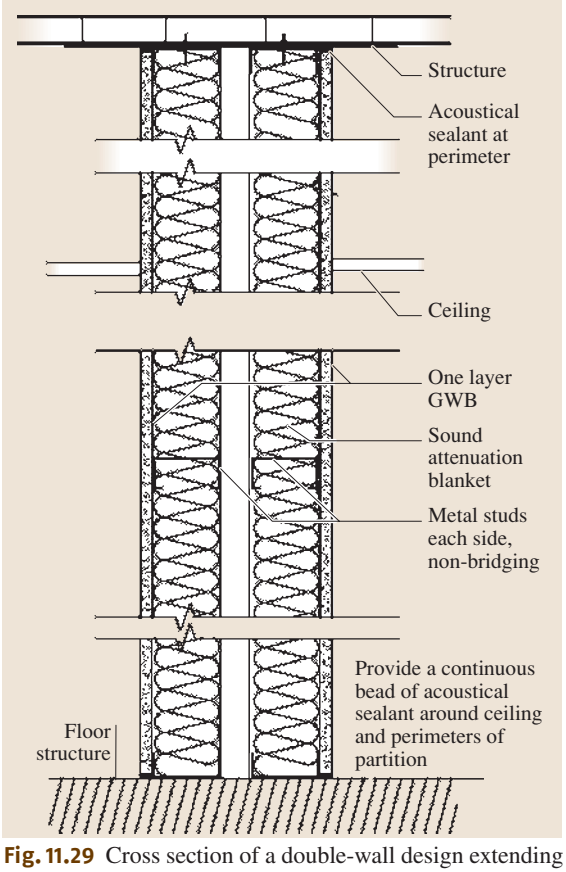


Fig. 11.29 Cross section of a double-wall design extending from floor to structural ceiling

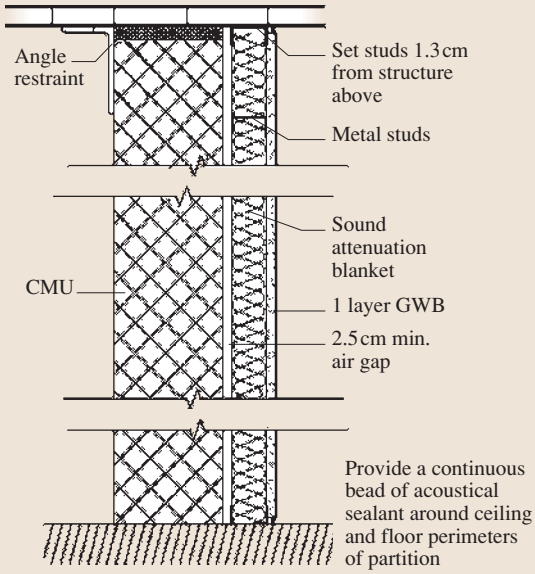
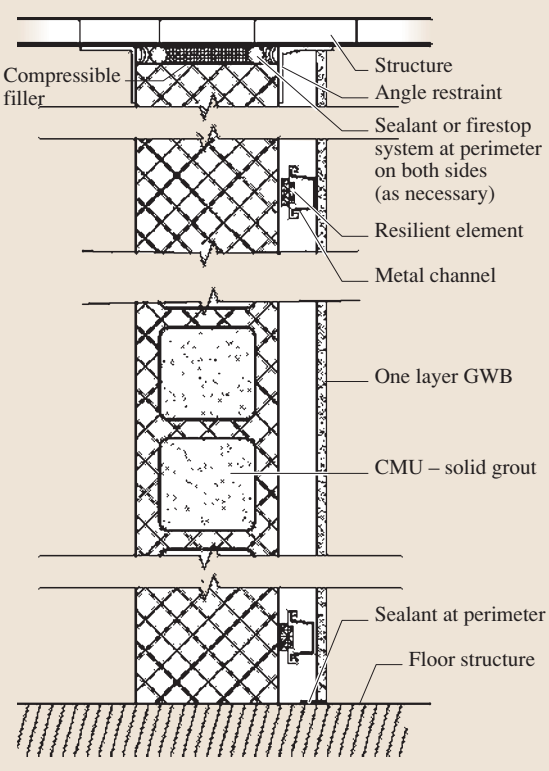


Fig. 11.30 Cross section of a floor to structural ceiling double-wall design with a CMU wall



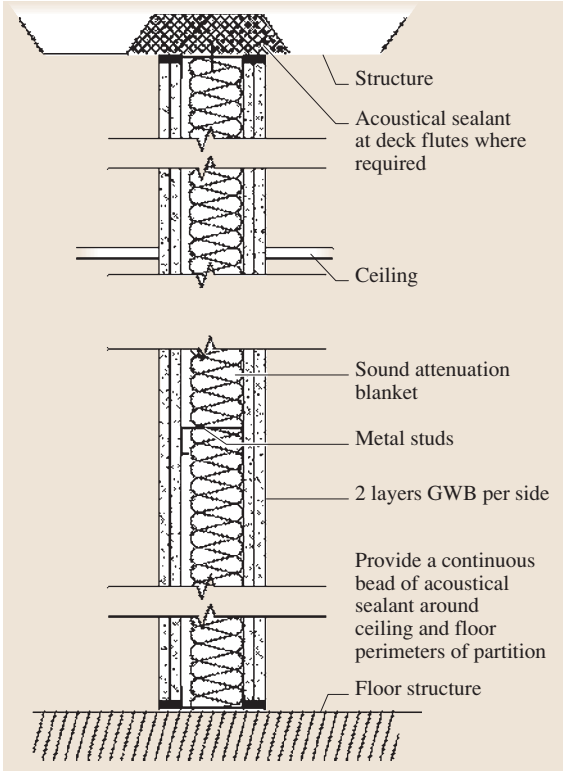


Fig. 11.32 Cross section of a single-studded sound insulating wall meeting a fluted metal structural deck

control is an issue, the STC rating does not address that and larger air spaces between wall components will increase the low-frequency noise control effectiveness.

Double-wall constructions are often impractical because of space restrictions. In these cases, an alternative to the CMU construction in Fig. 11.30 would be that shown in Fig. 11.31, in which the air space is replaced by a resilient element.

When working with gypsum board, Fig. 11.32 shows a design concept that will provide an STC of at least 50 as long as it is sealed properly. Note that a fluted structural deck may provide air gaps at the top of the wall that would need to be sealed to preserve the noise reduction effectiveness of the system.

When a room needs to be isolated from the building structure (which would be the case when structure-borne vibrations cause unwanted noise and vibration in certain rooms in which sensitive activities are tak-

Fig. 11.31 Cross section of alternate construction to a double-wall design in Fig. 11.30

Amazon EX1014 Page 422

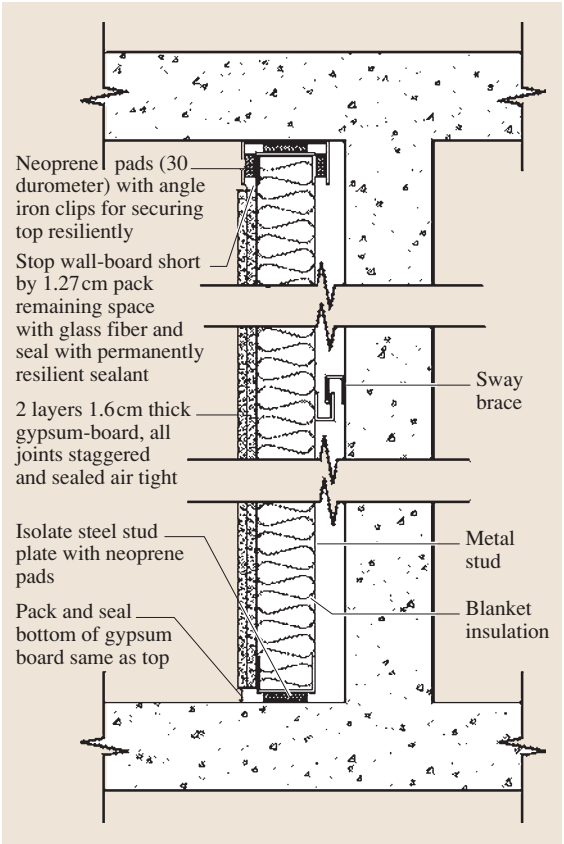


Fig. 11.33 Cross section of a resiliently mounted wall on a concrete building structure

ing place), Fig. 11.33 shows the conceptual design for a resiliently mounted wall. Figure 11.34 shows appropriate designs for intersections of double-studded walls.

A common path for sound leaks between rooms in buildings with large windows as facades is the connection between walls and those windows. Figures 11.35 and 11.36 show designs that effectively minimize sound transmission between rooms having this situation, Fig. 11.35 with a permanent installation and Fig. 11.36 with a curtain wall design.

Floor/Ceilings

Noise control design for floor/ceiling assemblies typically involves a significant amount of vibration isolation to minimize the transmission of footfall and other impact noise between floors. The design concepts that minimize vibration between floors can also be used to minimize sound transmission between floors.

The most common noise control design for floors is the floating floor system, in which a layer of resilient material is placed between the floor structure and the subflooring. More complicated measures are required for effective control of the high vibration and noise levels that can be associated with mechanical equipment in buildings. Figures 11.37–11.43 offer design concepts for resiliently mounted ceilings for mechanical rooms or other noisy spaces. Figure 11.44 shows a sound isolating design for a metal roof deck and Fig. 11.45 shows a sound isolating design for a concrete roof deck.

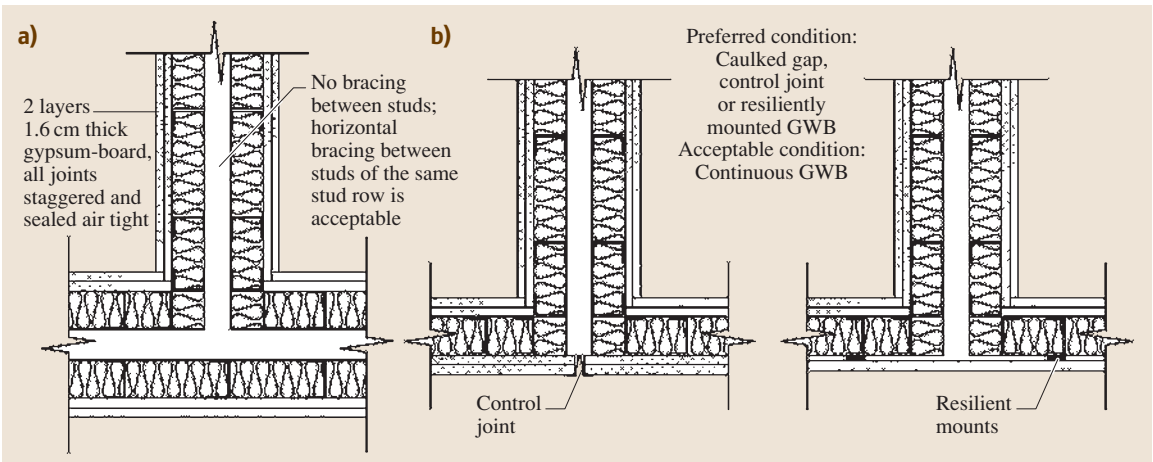


Fig. 11.34 Cross section of effective double-stud wall junction designs

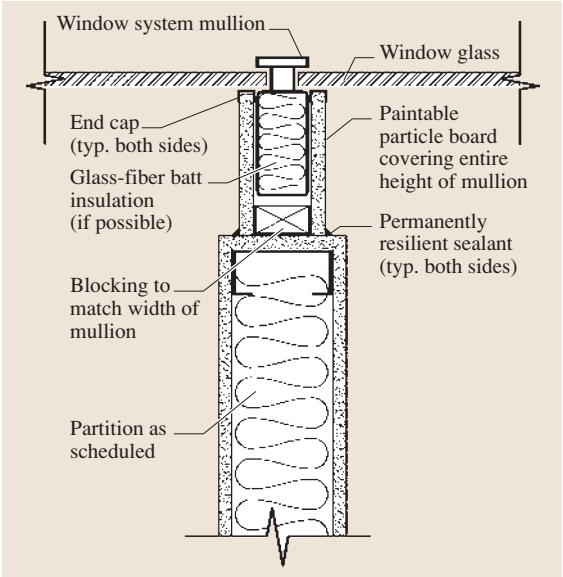


Fig. 11.35 Cross section of an effective mullion/partition junction

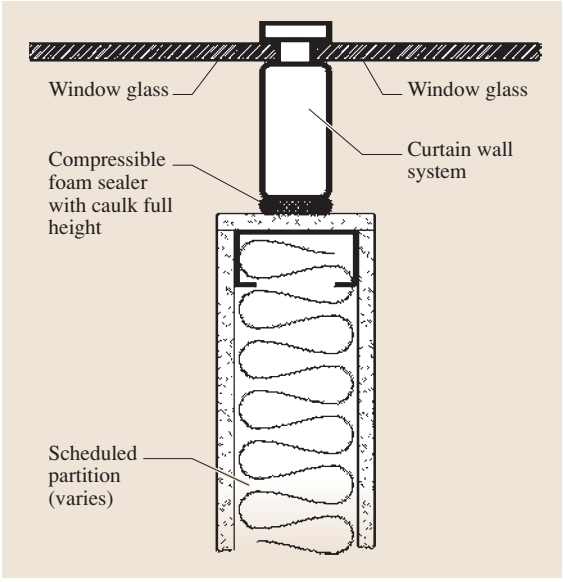


Fig. 11.36 Cross section of an effective mullion/partition junction with a curtain wall system

Windows and Doors

Windows and doors often have the lowest TL ratings of a wall system and are therefore often the sources of noise leaks through partitions. In addition to their lower TL ratings than the main wall, installation with air gaps

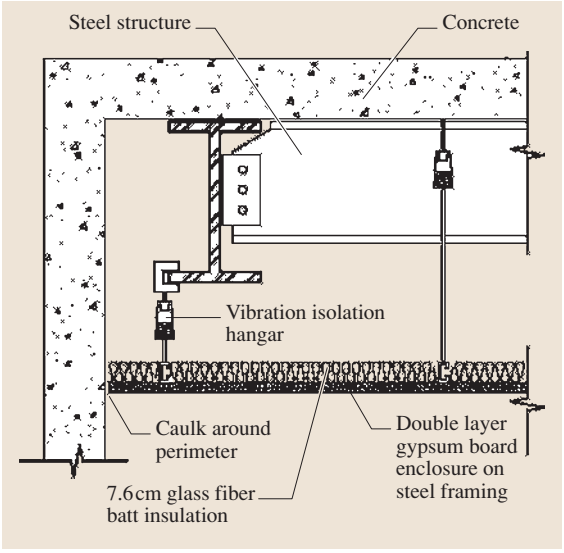


Fig. 11.37 Cross section of a resiliently mounted gypsum-board ceiling

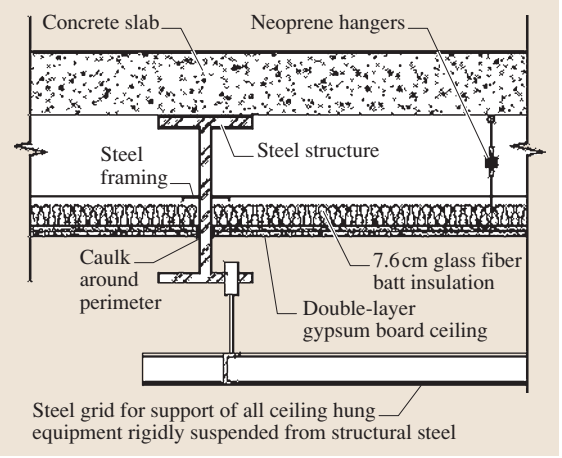


Fig. 11.38 Cross section of a sound insulating gypsum board ceiling framed from I-beams

around them can significantly compromise their noise control effectiveness.

Many window manufacturers provide models that offer significant sound insulation and each of these manufacturers offers its own acoustical test data and design parameters. As for walls, larger air spaces between double-paned designs will provide better sound insulation than single-paned windows of the same thickness. For those wishing to manufacture custom sound-insulating windows, Fig. 11.46 shows some design

Amazon EX1014 Page 424

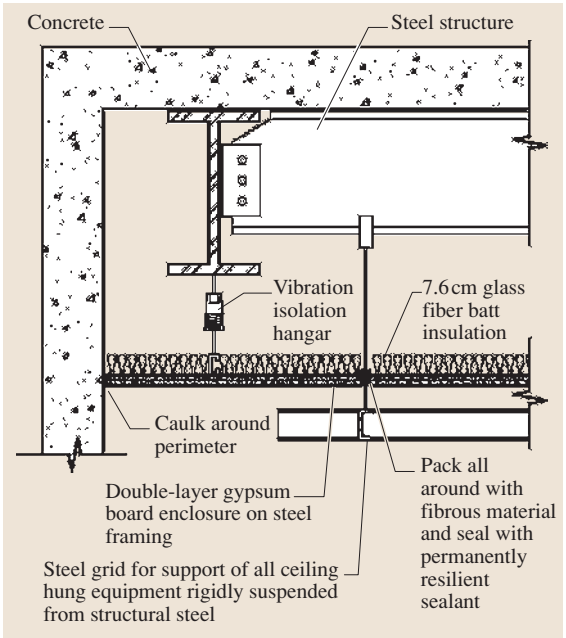


Fig. 11.39 Cross section of an effective sound insulating ceiling for mechanical rooms

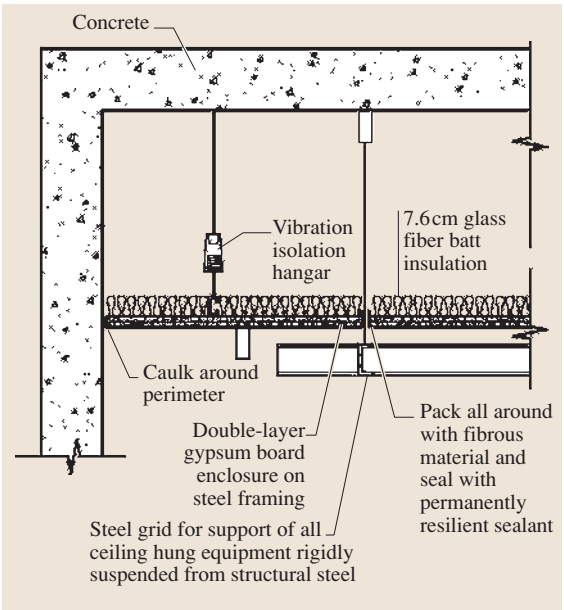


Fig. 11.40 Cross section of a ceiling resiliently mounted from a concrete slab

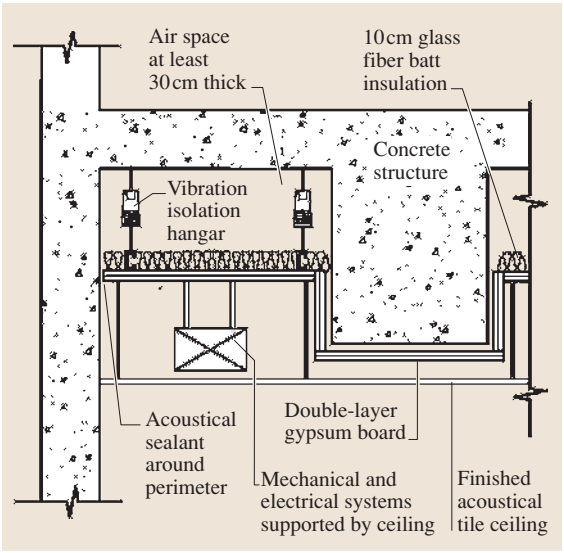


Fig. 11.41 Cross section of a resilient ceiling supporting mechanical and electrical system equipment

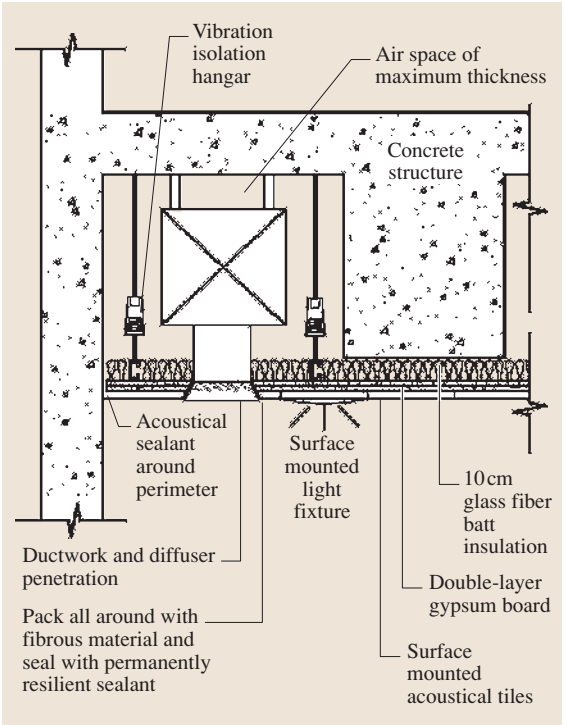


Fig. 11.42 Cross section of a resilient ceiling with HVAC equipment

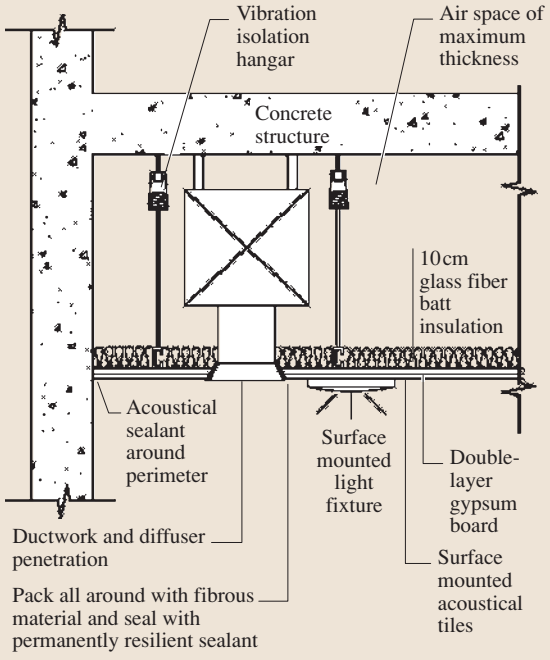


Fig. 11.43 Cross section of an alternative resilient ceiling design

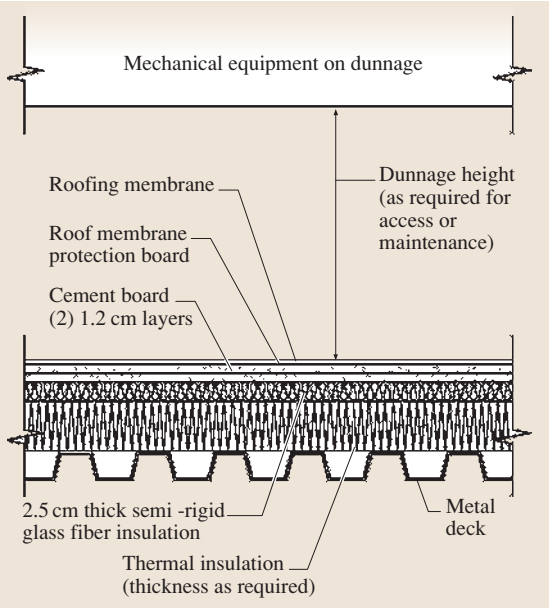


Fig. 11.44 Cross section of a sound isolating sandwich design for a metal roof deck

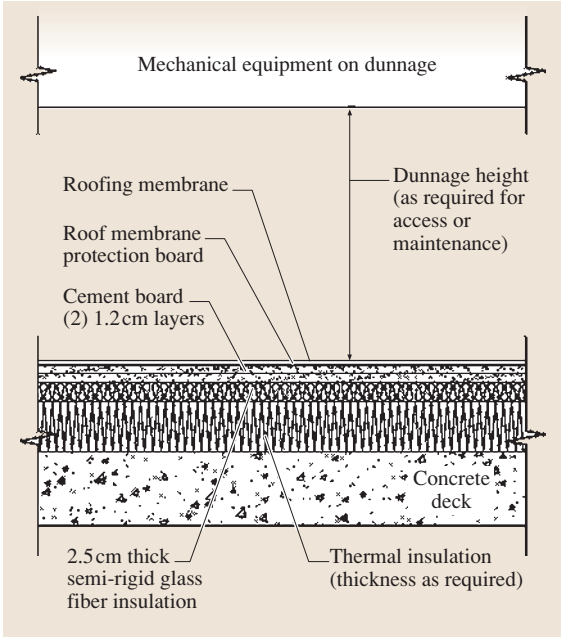


Fig. 11.45 Cross section of a sound isolating sandwich design for a concrete roof deck

guidelines for high-TL interior windows and Fig. 11.47 shows design guidelines for high-TL exterior windows.

As for windows there are many door manufacturers that provide sound insulation data for their prefabricated products. In general, solid-core doors are more effective than hollow-core doors for noise reduction. To maximize the effectiveness of these doors, they should be gasketed around their perimeters and they should either have drop seals or rubber skirts on their undersides, which should contact solid floor thresholds.

11.5.2 HVAC Systems

ASHRAE publishes thorough discussions on HVAC noise sources and their control in their *Fundamentals and Applications Handbooks*. The basic principles to bear in mind for minimizing HVAC noise are minimizing airflow turbulence and isolating vibrations, each of which is able to carry noise far from the source to remote locations in buildings.

Airflow Turbulence

Airflow turbulence noise can be minimized by

- lowering flow rates as much as possible,

Amazon EX1014 Page 426

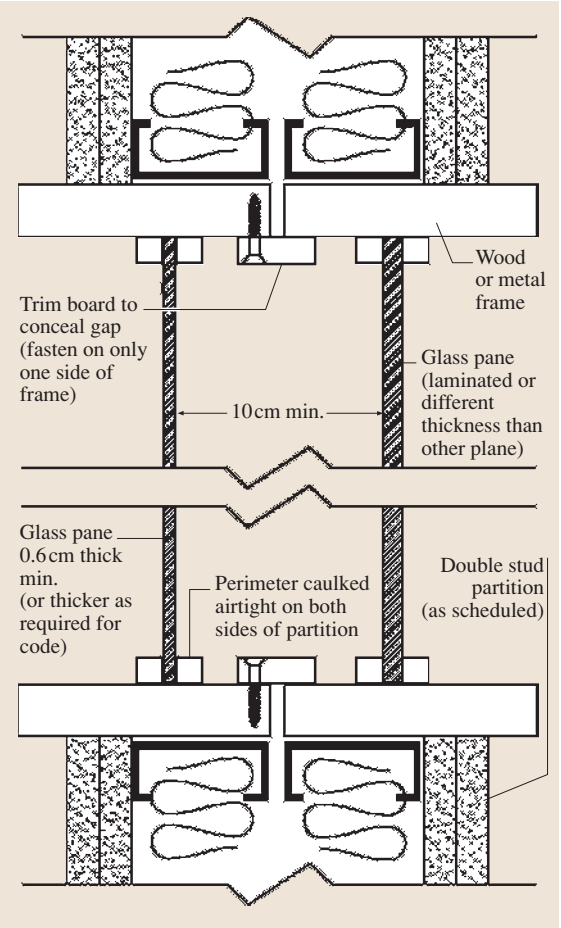


Fig. 11.46 Cross section of a high-TL interior window in a double-wall design

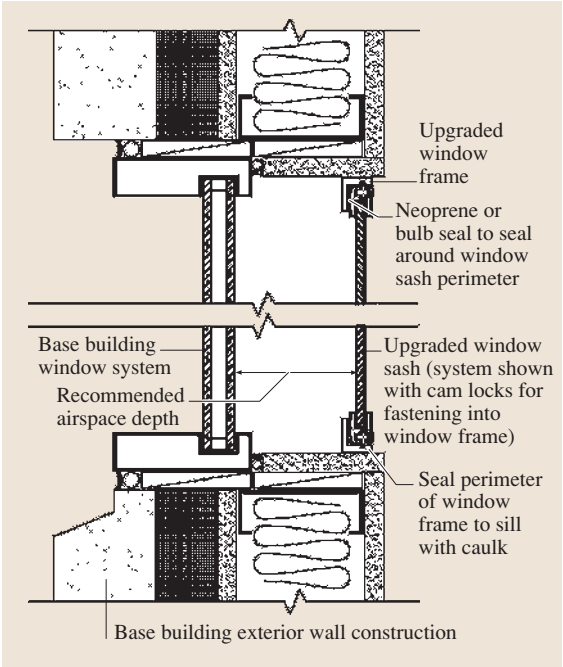


Fig. 11.47 Cross section of an upgraded exterior window for high-TL design

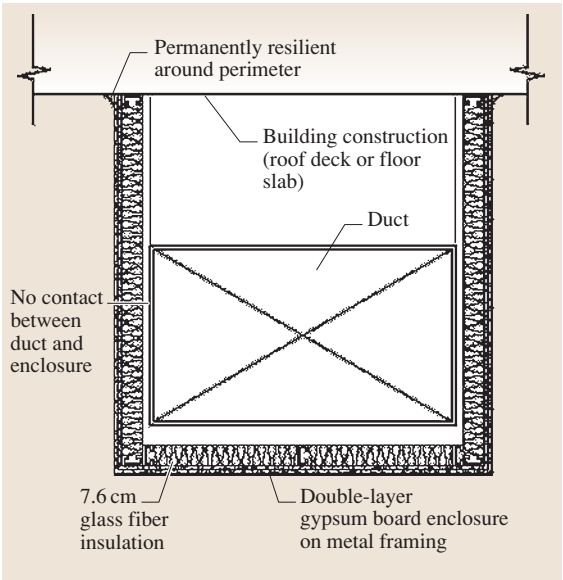


Fig. 11.48 Cross section of a double-layer gypsum-board duct enclosure

- distributing the flow as evenly as possible throughout the ducted system,
- providing long ducted passages and transitions that avoid sharp changes in direction or discontinuities in cross-sectional area,
- installing properly designed silencers and absorptive lining in the flow path that do not unduly compromise the efficiency of the HVAC system,
- wrapping ductwork with lagging materials, and
- enclosing ductwork and HVAC equipment.

Figures 11.48 through 11.50 offer effective noise control designs for duct and hanging HVAC equipment enclosures.

Another issue to consider with ductwork is sound leaks through wall and ceiling duct penetrations. If these penetrations are not sealed properly, they will diminish the noise reduction effectiveness of the wall or ceiling

Amazon EX1014 Page 427

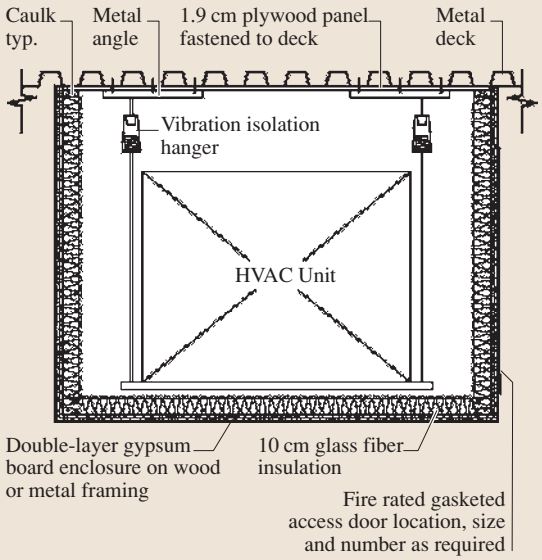


Fig. 11.49 Cross section of a duct enclosure resiliently mounted to a metal deck

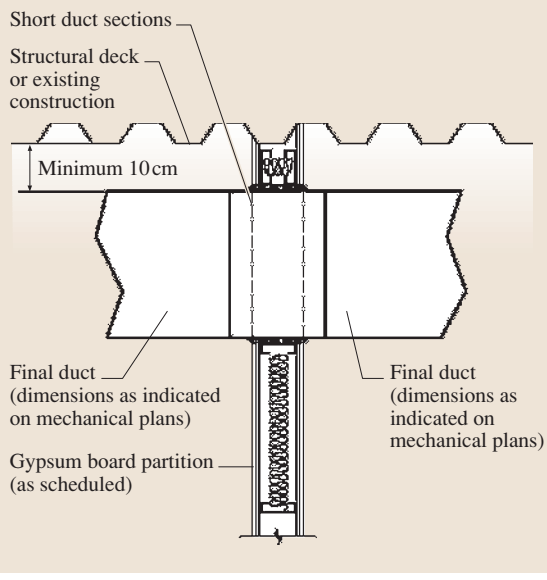


Fig. 11.51 Cross section of a duct penetrating a gypsum-board partition near a deck

that they are penetrating. Figure 11.51 shows an effective design for a single duct penetrating a wall and Fig. 11.52 shows an effective design for multiple duct penetrations. Figure 11.53 shows design considerations for the placement of a duct silencer near a wall penetration.

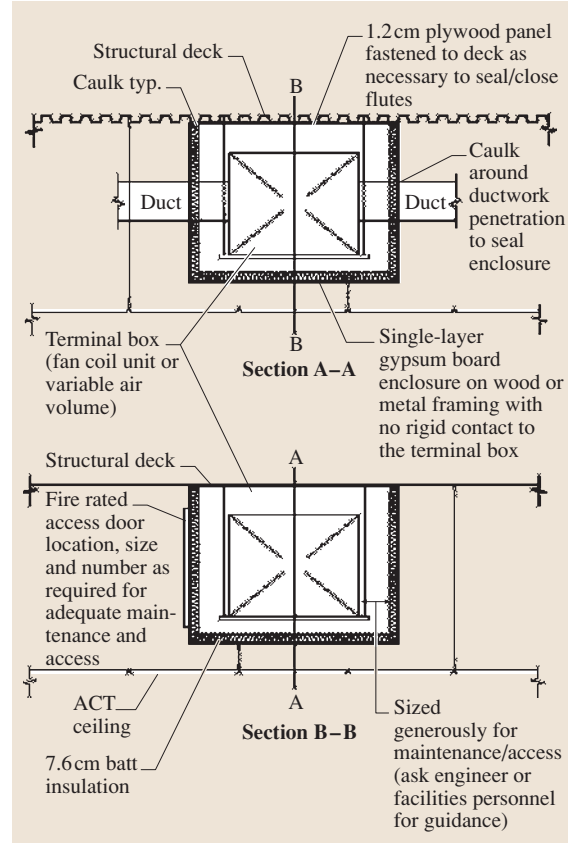


Fig. 11.50 Sections of a ducted terminal box enclosure mounted to a structural deck

Vibration Isolation

Springs and pads are the most common tools used to minimize the transfer of vibration from HVAC units to the building structure. The key here is to eliminate any rigid connections between the units and the structure. Figure 11.54 shows a typical spring isolator properly mounted to reduce this transfer of vibrations effectively.

These benefits, however, will become short-circuited if improper or misaligned isolators are installed. Figure 11.55 shows improperly loaded spring isolators supporting a piece of equipment, one being overcompressed and the other being undercompressed.

Figure 11.56 shows an example of a spring isolating pipe hanger that is misaligned. The rigid connections established by this misalignment compromise the effectiveness of the system.

Figure11.57 shows a common design in which pipe supports are rigidly attached to a floor. Figure 11.58

Amazon EX1014 Page 428

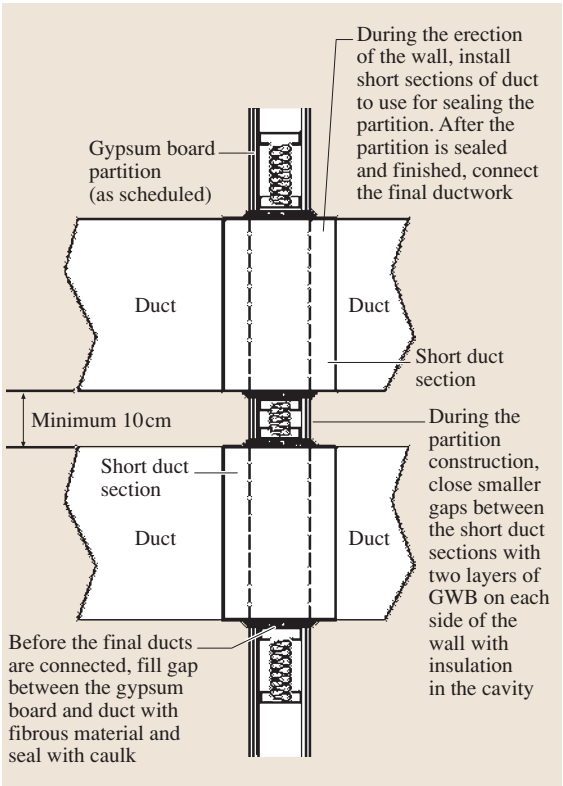


Fig. 11.52 Cross section of sealed multiple duct penetrations in a partition

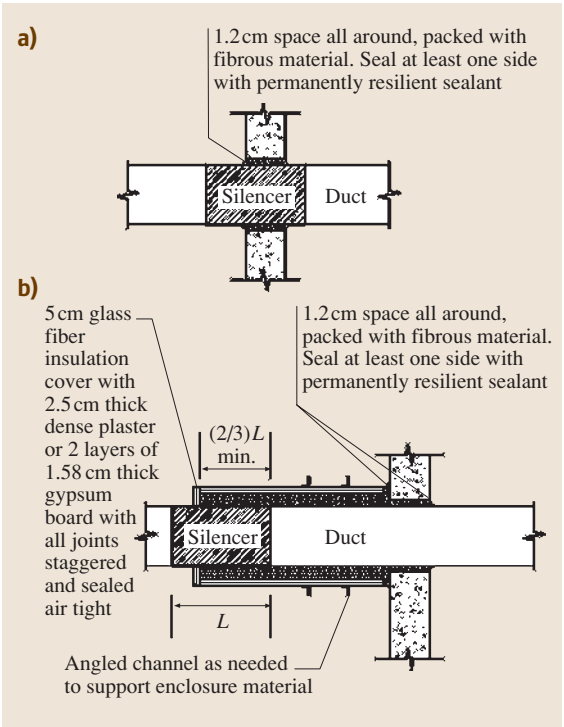


Fig. 11.53a,b Cross section of a duct penetration with appropriate silencer locations. **(a)** preferred location; **(b)** alternate location

shows an effective design to reduce the transfer of vibrations between the piping and the floor using neoprene pads and spring hangers. Note that any spring hangers must be appropriately sized for each situation. When lateral restraints are required for large equipment, Fig. 11.59 offers a general design for effective vibration isolation.

Enclosures for Floor-Based Equipment

When barriers and source treatments do not provide enough noise control, the only practical noise control option for many floor-based units may be a full enclosure. Key issues for enclosures of HVAC equipment are access (for maintenance) and ventilation. A practical design for indoor environments is shown in Fig. 11.60, in which a heavy curtain system (which can be drawn along tracks for access) surrounds the equipment. There are hoods built into the sliding curtains to provide the necessary ventilation for the equipment.

Exterior enclosures are discussed in Sect. 11.5.5 below.

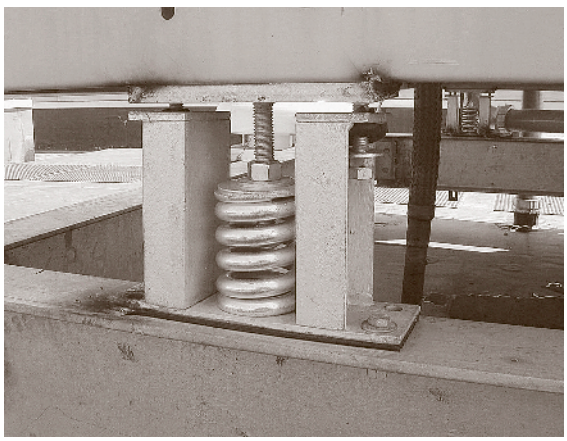


Fig. 11.54 A properly mounted spring isolator

11.5.3 Plumbing Systems

As for HVAC systems, the most common plumbing-system noise issues are related to reducing flow turbulence and vibration isolation. For plumbing sys-

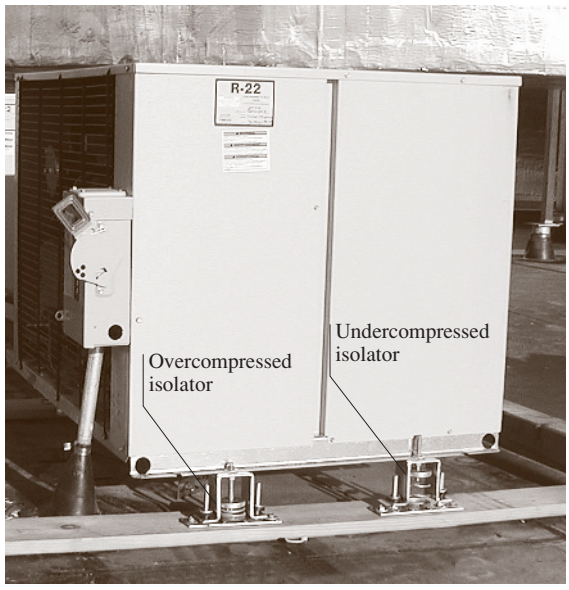


Fig. 11.55 Examples of misaligned and misloaded spring isolators

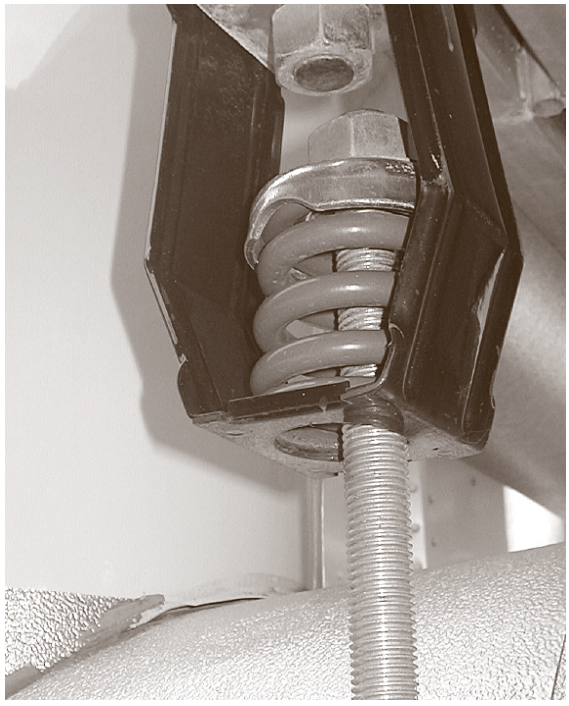


Fig. 11.56 A misaligned isolating pipe hanger. Note that the post is touching the support, thus short-circuiting the isolating effect of the hanger

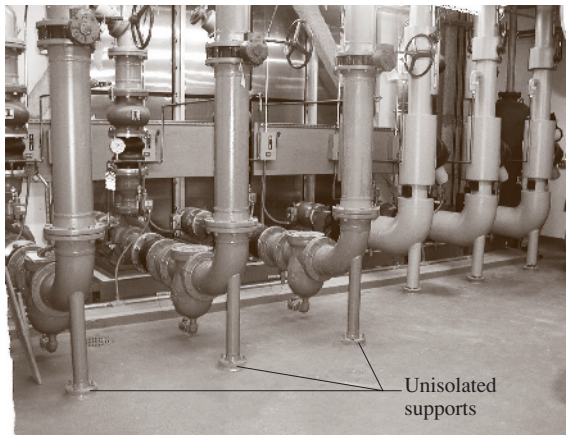


Fig. 11.57 Pipe supports rigidly attached to a structural floor. This can channel vibrations through the building structure

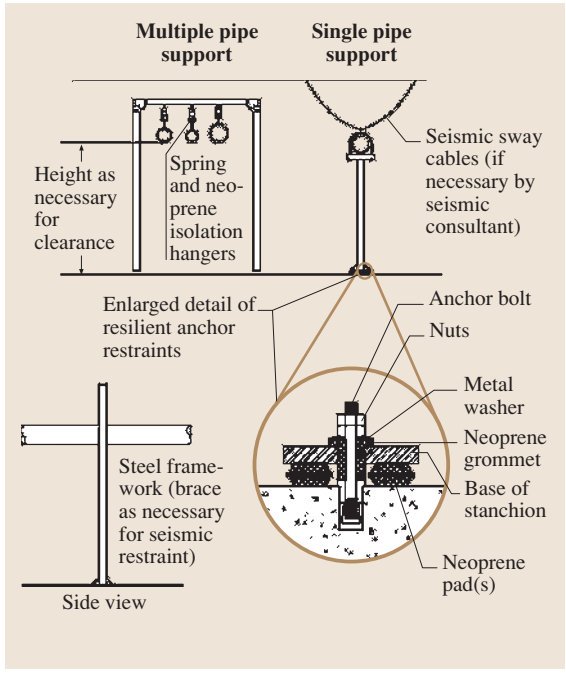


Fig. 11.58 Pipe floor supports with resilient anchors

tems we are dealing with liquid flow rather than air flow and pipe noise and vibration isolation rather than duct noise and vibration isolation. As for ducted systems, pipes can be wrapped with lagging materials to reduce breakout noise radiated from the pipe walls. Figures 11.61 and 11.62 show the general design principles to bear in mind for pipe penetrations through common

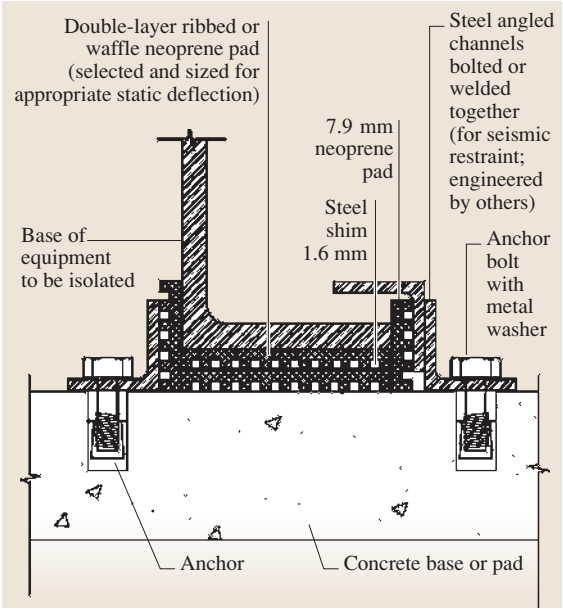


Fig. 11.59 Cross section of a lateral restraint for an equipment base on neoprene pads

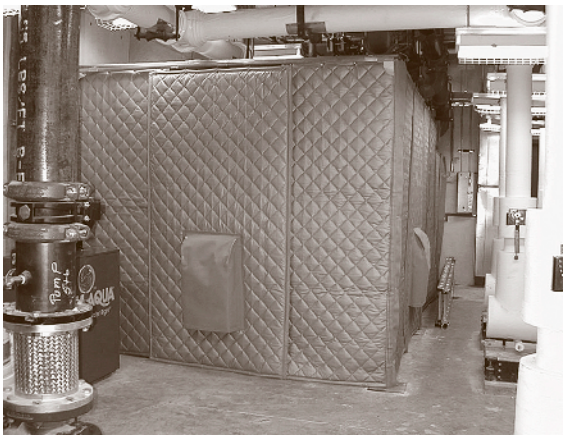


Fig. 11.60 A heavy curtain enclosure system for floor-based equipment

walls and Fig. 11.63 shows these principles in practice in an installation.

11.5.4 Electrical Systems

Electrical system noise control issues typically deal with transformer noise control and the vibration channel set up by rigid conduit between noise and vibration sources and noise-sensitive spaces. When transformers cause

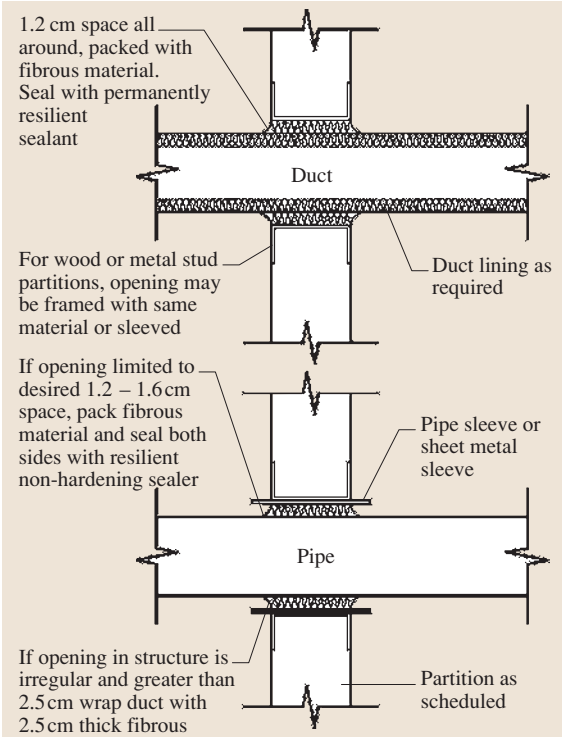


Fig. 11.61 Cross sections of sealed wall penetrations with a duct and a pipe

noise issues, they are most commonly addressed by enclosing the transformers. Conduit issues can be addressed by using a flexible conduit that is not stretched. Figure 11.64 shows an example of stretched electrical conduit, which provides a clear vibration channel from the source to remote locations.

An air space between an electrical box and other equipment will break any vibration channel that would otherwise have been established, as is shown in Fig. 11.65.

11.5.5 Exterior Sources

Environmental noise sources, such as transportation or industrial sources, can only be controlled in buildings by using exterior-wall designs that effectively reduce noise in the frequency ranges of those sources. Exterior noise sources, such as rooftop mechanical equipment or mechanical equipment in lots beside buildings (associated with the operations of the building in question) can be controlled by appropriately designed enclosure systems.

Amazon EX1014 Page 431

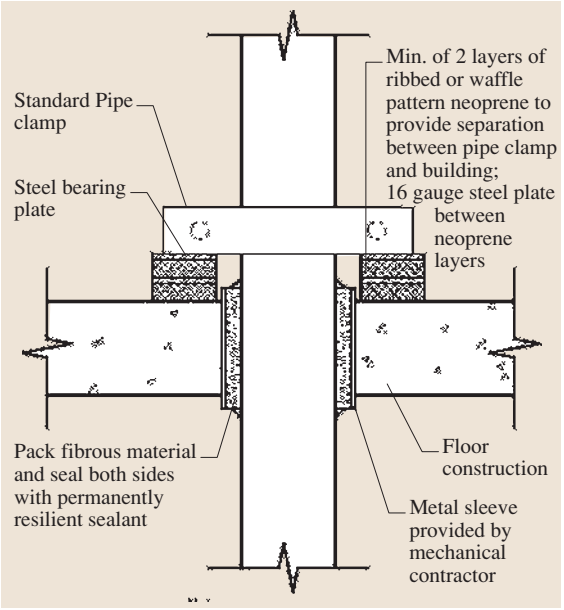


Fig. 11.62 General considerations for sealing pipe penetrations



Fig. 11.63 Examples of sealed pipe and duct penetrations

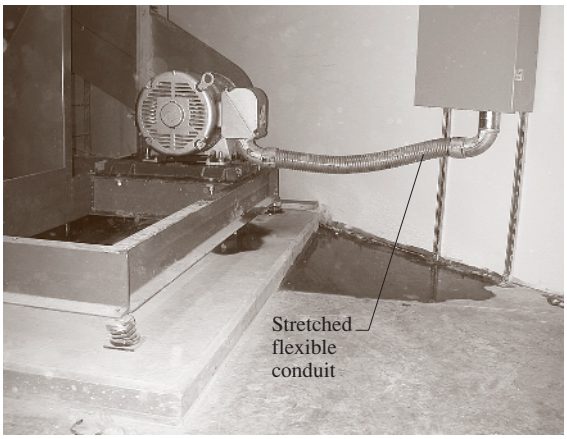


Fig. 11.64 An example of stretched flexible conduit



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 EX1014 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 EX1014 Page 433

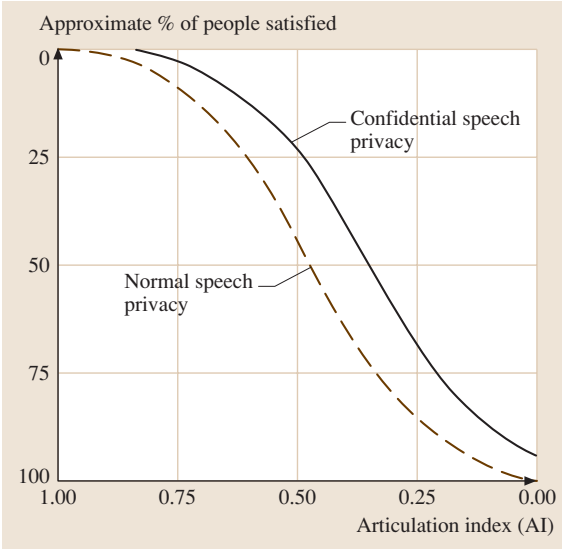


Fig. 11.69 Correlation between articulation index and speech privacy (adapted from [11.1])

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.

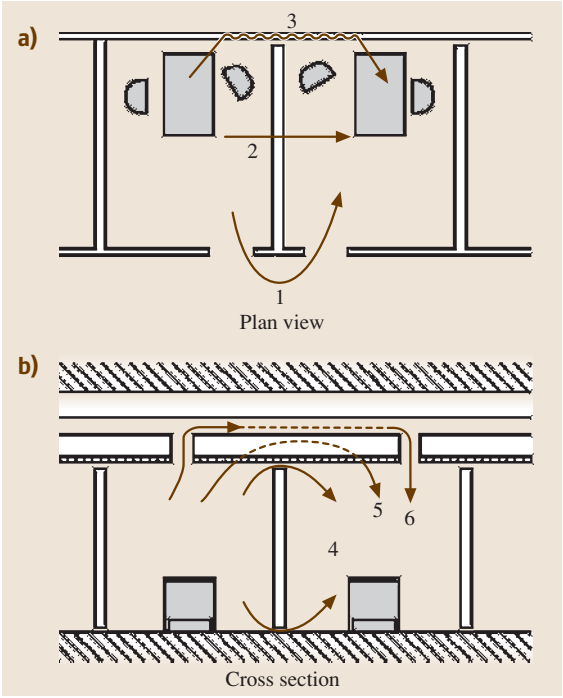


Fig. 11.70a,b The most common paths for sound leakage between closed offices. **(a)** Plan view; **(b)** cross section (1-diffraction between openings, 2-transmission through partitions, 3-structure-borne transmission, 4-transmission through air gaps in wall seals, 5-diffraction through plenum when partition 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 EX1014 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 EX1014 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

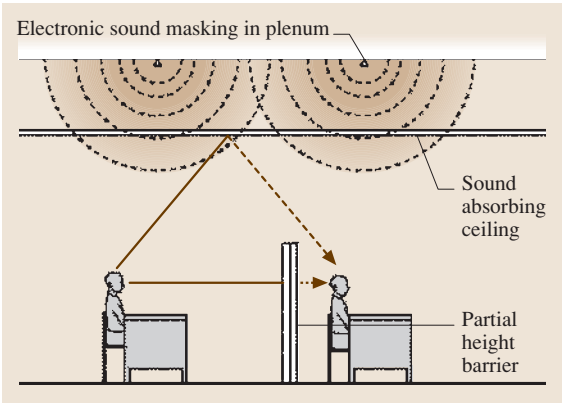


Fig. 11.71 Acoustic considerations for open-plan office design in cross section

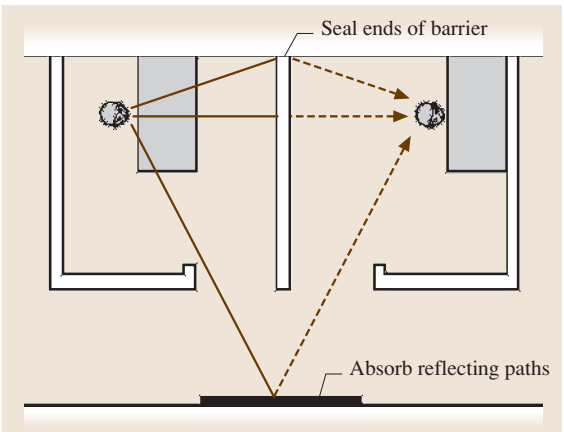


Fig. 11.72 Acoustic considerations for open-plan office design in plan view

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.

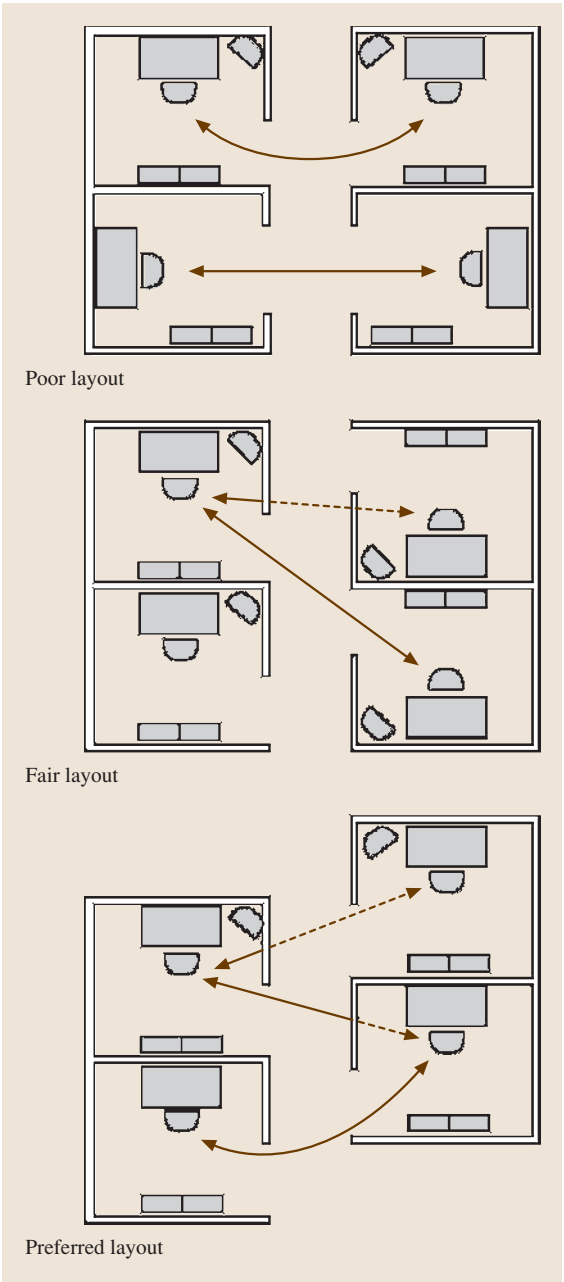


Fig. 11.73 Acoustical considerations for open-plan office layout

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 EX1014 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)

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 EX1014 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 EX1014 Page 441

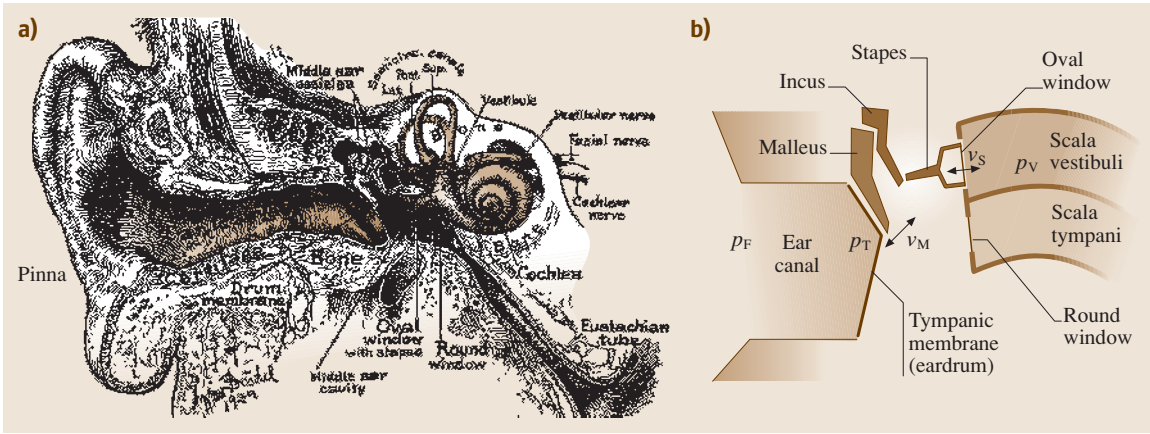


Fig. 12.1 **(a)** Drawing of a cross section of the human ear, showing the pinna, the ear canal, eardrum (drum membrane), middle-ear bones (ossicles), and the labyrinth. The latter contains the vestibular organs of balance and the cochlea, or inner ear. (After [12.8]) **(b)** Schematic of the mammalian 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-

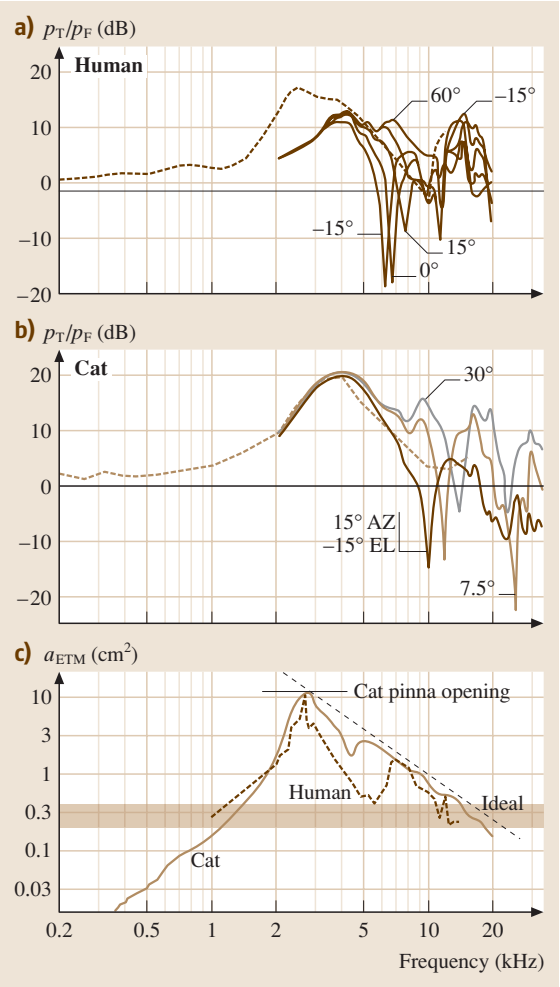


Fig. 12.2a–c Transformation of sound pressure by the external ear, considered as the ratio p_T/p_F for the human ear **(a)** and the cat ear **(b)**. p_T is the pressure measured by a probe microphone in the ear canal, within 2 mm of the eardrum for all cases except the *solid curves* in **(a)**, which 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 EX1014 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

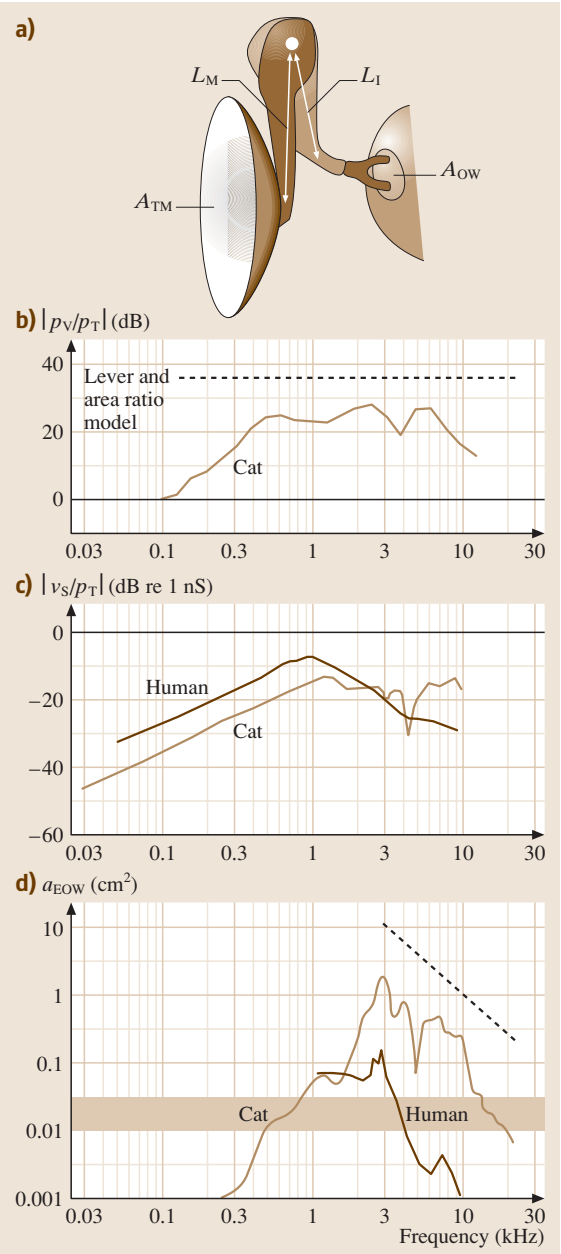


Fig. 12.3 **(a)** Schematic drawing of the mammalian middle ear from the eardrum (TM on the *left*) to the oval window (OW on the *right*). The malleus, incus, and stapes are shown in their approximate anatomical arrangement. The areas of the TM and OW are shown along with the lever arms of the malleus (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])

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].

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

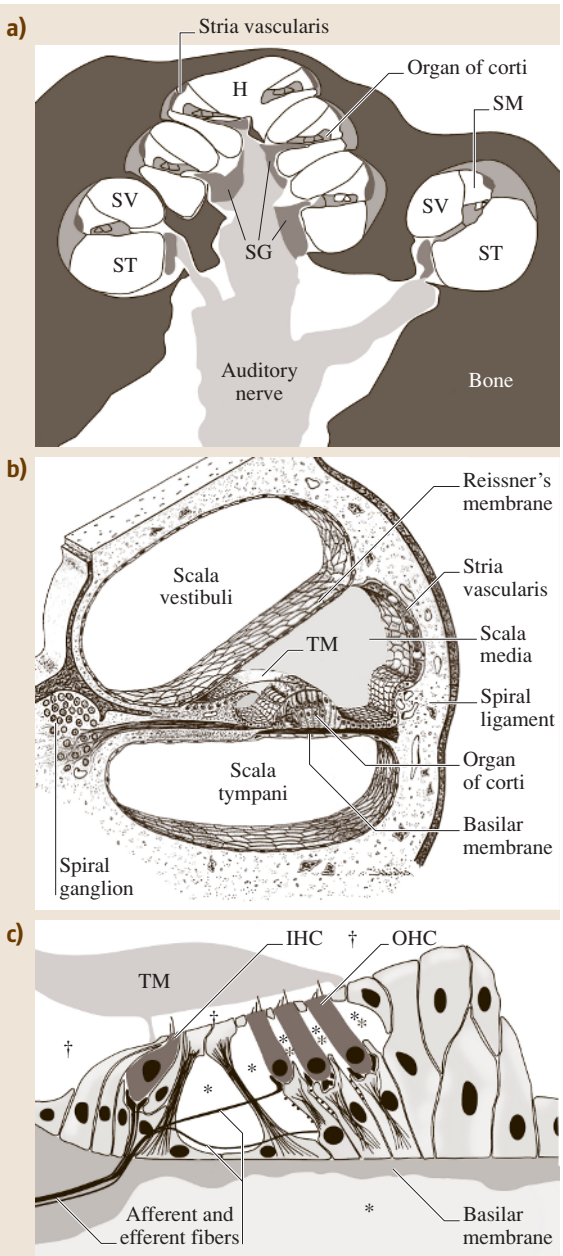


Fig. 12.4 **(a)** Sketch of a cross section of the cochlea, cut through the center of the coil. The cochlear spirals are cut in cross section, shown at higher resolution in **(b)**. The cochlea consists of three fluid-filled chambers that spiral together: ST – scala tympani, SV – scala vestibuli, and 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

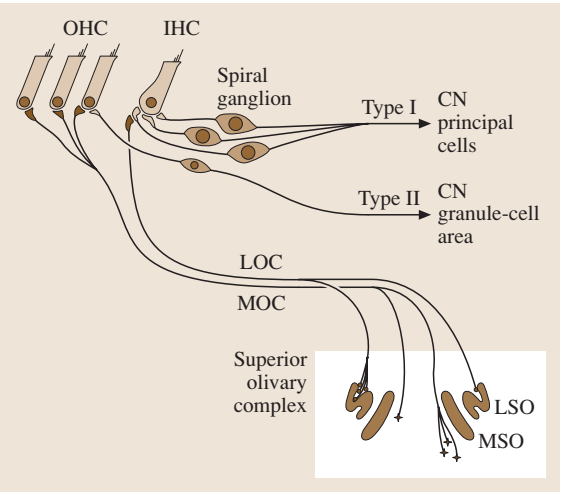


Fig. 12.5 A schematic of the wiring diagram of the mammalian cochlea. The hair cells are on the left. Afferent terminals on hair cells, those in which the hair cell excites an AN fiber, are shown *lightly shaded*; efferent terminals, in which an axon from a cell body in the central nervous system contacts 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

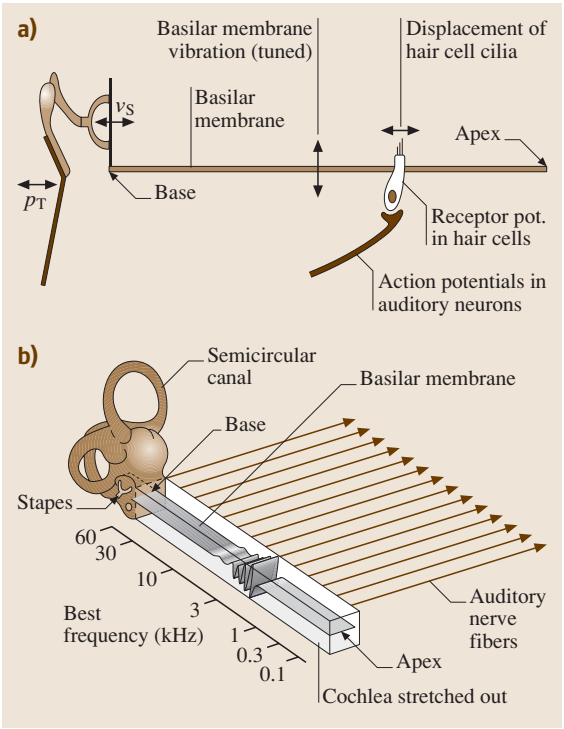


Fig. 12.6 **(a)** Summary of the steps in direct cochlear transduction. **(b)** Schematic diagram showing the basilar membrane in an unrolled cochlea and the nature of cochlear frequency analysis. A snapshot of basilar-membrane displacement (greatly magnified) by a tone with frequency near 3 kHz is shown. 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

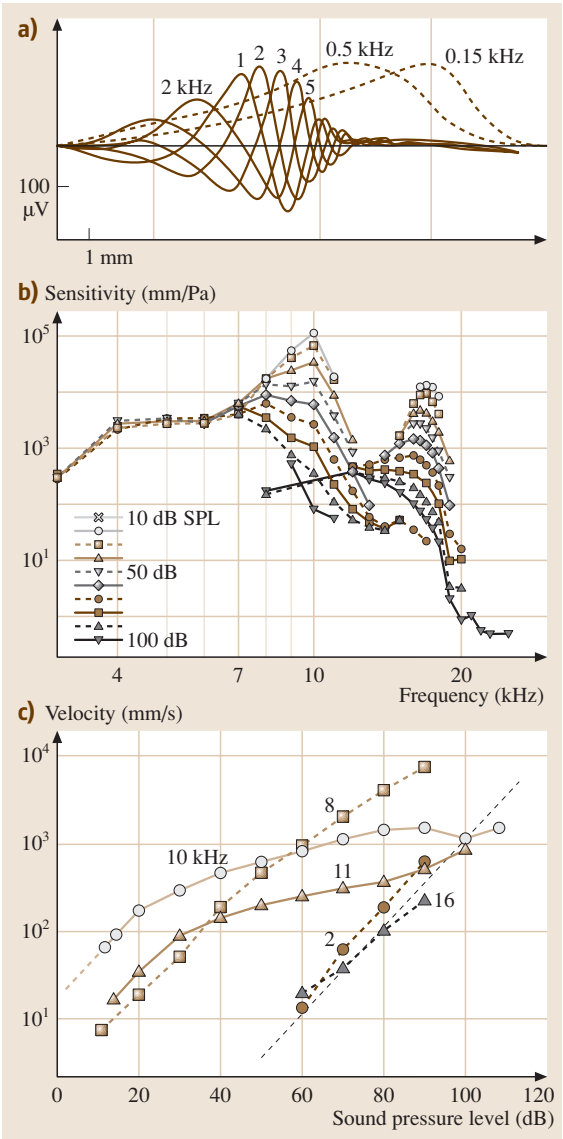


Fig. 12.7 **(a)** An estimate of the instantaneous displacement of the basilar membrane in the guinea-pig cochlea for a 2 kHz tone (*solid lines*) and estimates of the envelope of the displacement for two lower-frequency tones (*dashed lines*). Distance along the basilar membrane runs from left (*base*) 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 EX1014 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-

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

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 EX1014 Page 453

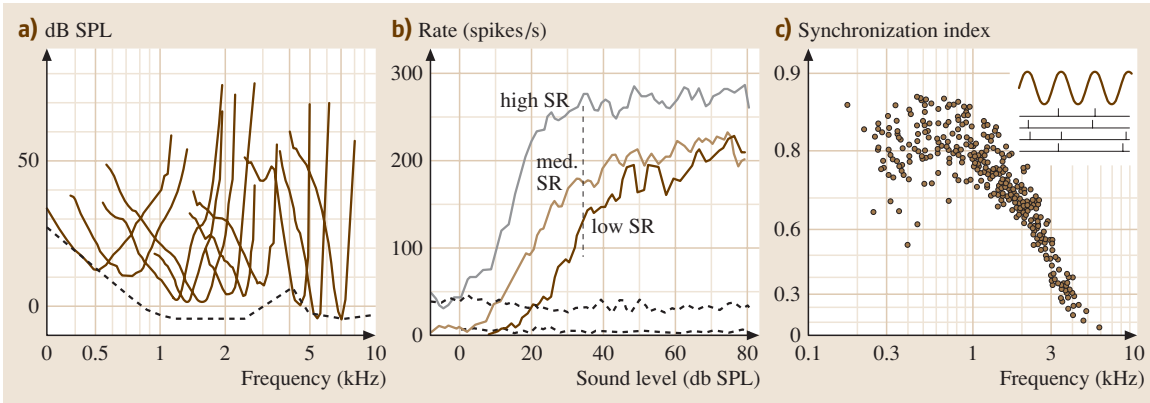


Fig. 12.8a–c Basic properties of responses to sound in AN fibers. **(a)** Tuning curves, showing the threshold sound level for a response plotted versus frequency. the *dashed line* shows the lowest thresholds across a population of animals. **(b)** Rate versus sound-level plots for responses to BF tones 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

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

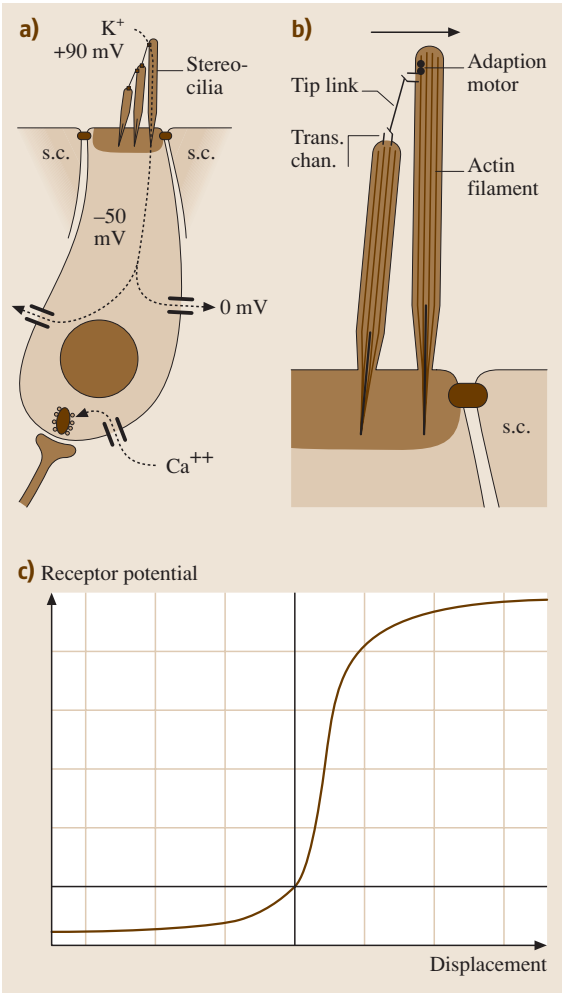


Fig. 12.9a–c Transduction in IHCs. **(a)** Schematic of an IHC showing the components important for transduction. The stereocilia protrude into the endolymphatic space (*top*), containing mainly K^+; the extracellular potential here is about +90 mV (the endolymphatic potential). The transduction apparatus 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

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.

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.

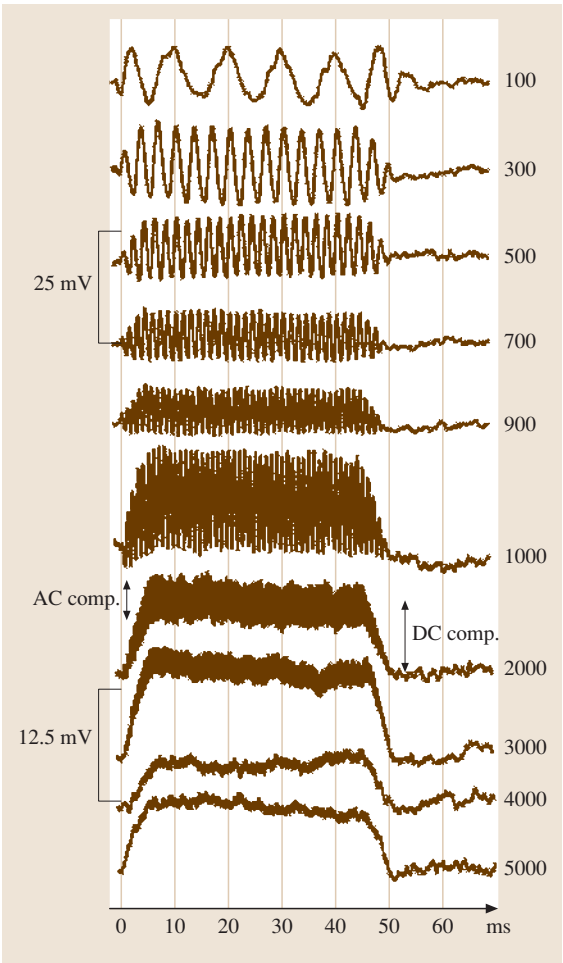


Fig. 12.10 Receptor potentials in an IHC in the basal turn of a guinea-pig cochlea. The stimulus level was set at 80 dB SPL so that the cell would respond across a wide range of frequencies. Stimulus frequency is given at right. Notice the change in ordinate scale between the 900 Hz and 1000 Hz traces.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

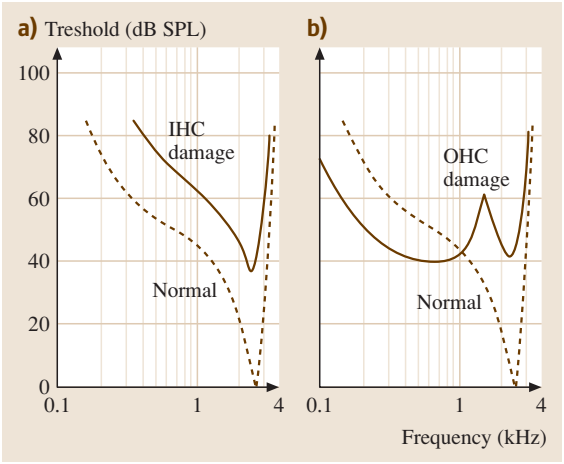


Fig. 12.11a,b Schematic summary of AN tuning curves in normal ears and in ears with damaged cochleae. **(a)** In regions of the cochlea with remaining OHCs but damaged IHCs tuning curves mainly show a threshold elevation. **(b)** In regions with missing OHCs but intact IHCs there is an elevation of threshold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

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

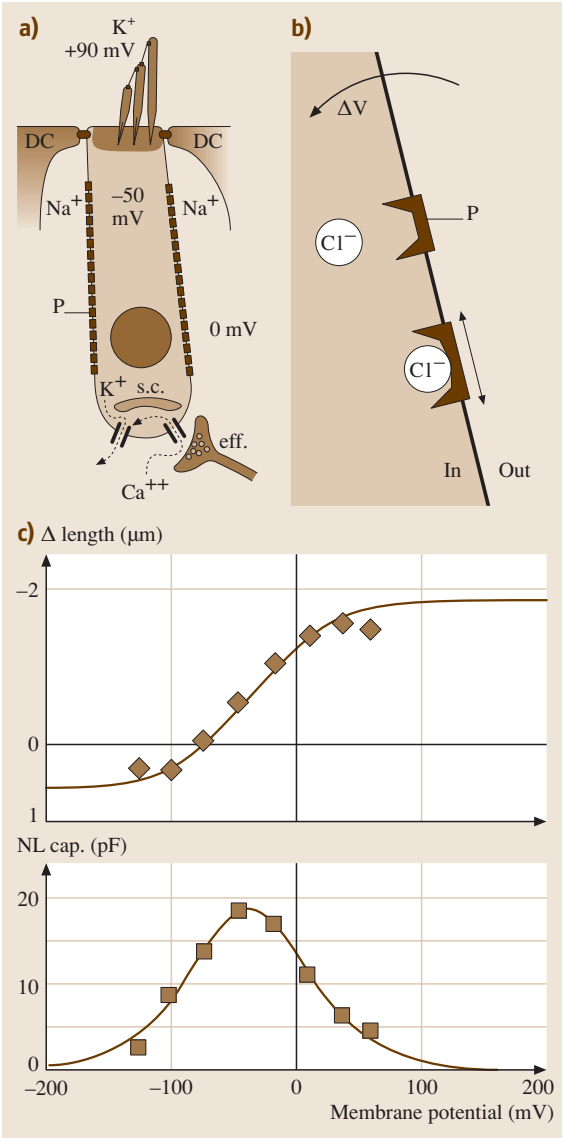


Fig. 12.12 **(a)** Schematic picture of an OHC showing elements important for OHC function. The diagram is labeled as for Fig. 12.9a except for the following: DC – Deiter's cell extensions that are linked to OHCs by tight junctions to form the top surface of the organ of Corti; Na^+ – the extracellular 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)

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-

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

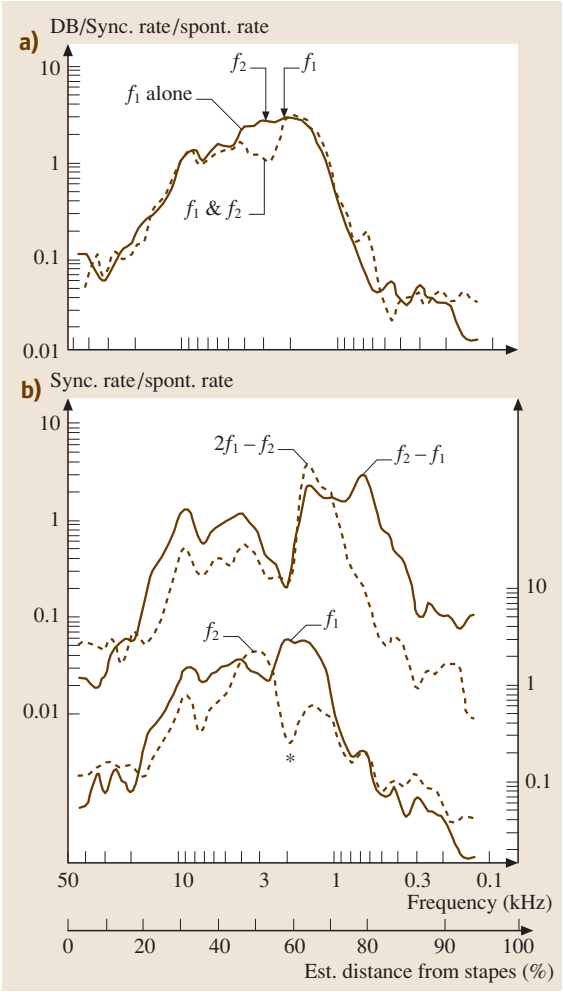


Fig. 12.13a,b Responses of a population of AN fibers to a two-tone complex consisting of f_1=2.17 kHz and f_2=2.79 kHz at 65 dB SPL. The *ordinates* show the strength of the phase-locked response as the synchronized rate divided by the spontaneous rate. Synchronized rate is the Fourier transform of the 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-

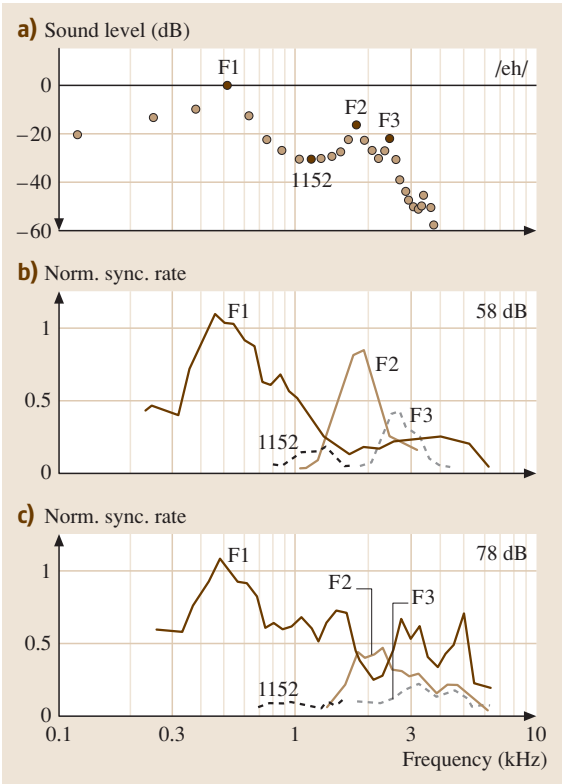


Fig. 12.14a–c Responses to a stimulus similar to the vowel /eh/ presented at two sound levels in a population of AN fibers. **(a)** The magnitude spectrum of the stimulus. It is periodic and consists of harmonics of a 125 Hz fundamental. There are peaks of energy at the formant frequencies of a typical /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.

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.

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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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.

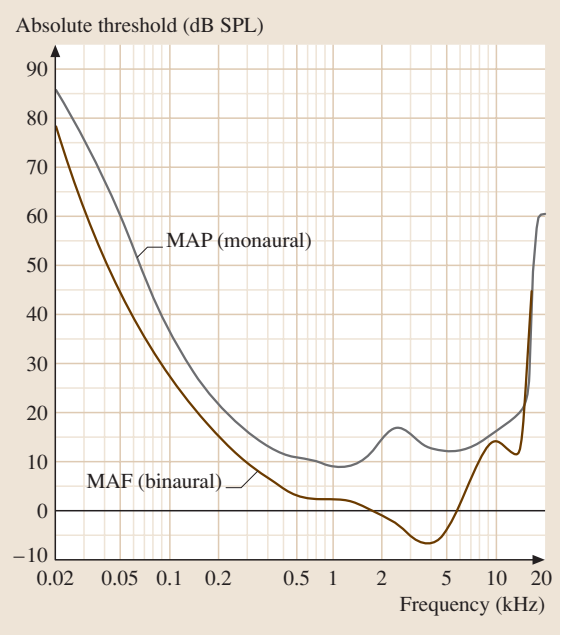


Fig. 13.1 The minimum audible sound level as a function of frequency. The *solid curve* shows the minimum audible field (MAF) for binaural listening published in an International Standards Organization (ISO) standard [13.1]. The *dashed curve* shows the minimum audible pressure (MAP) for monaural listening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

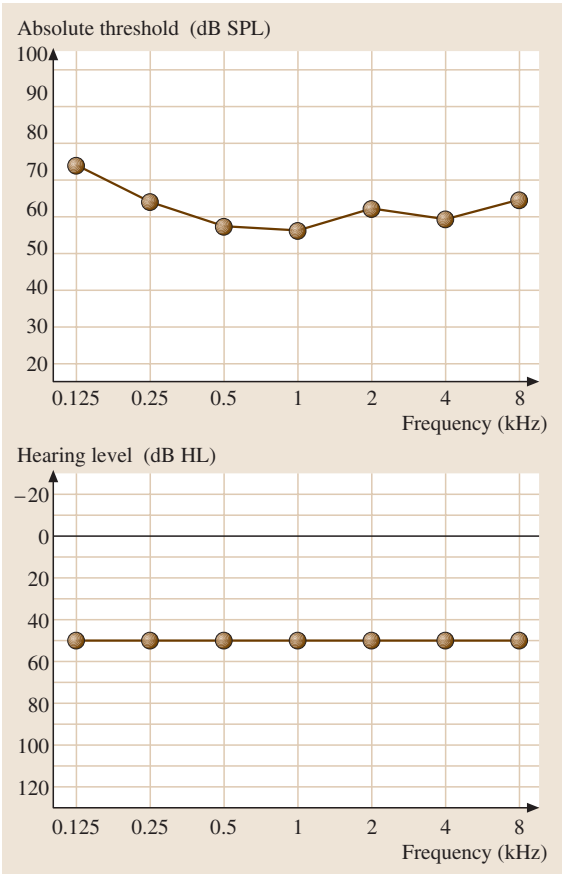


Fig. 13.2 Comparison of a clinical audiogram for a 50 dB hearing loss at all frequencies *(bottom)* and the absolute threshold curve for the same hearing loss plotted in terms of the MAP *(top)*

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).

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.

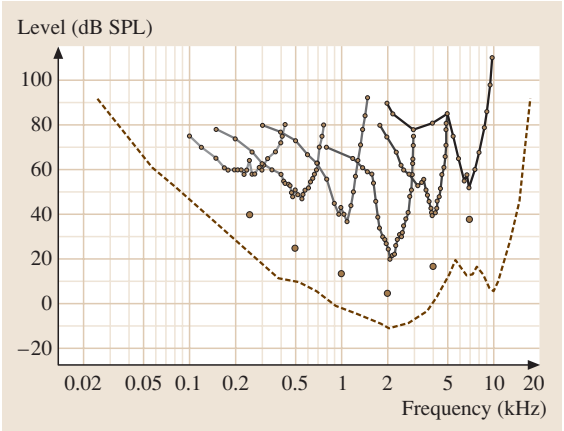


Fig. 13.3 Psychophysical tuning curves (PTCs) determined in simultaneous masking, using sinusoidal signals at 10 dB SL. For each curve, the *circle* below it indicates the frequency and level of the signal. The masker was a sinusoid which had a fixed starting phase relationship to the 50 ms signal. The 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

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

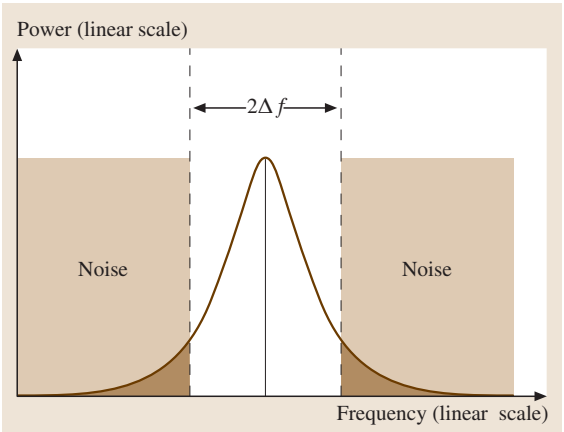


Fig. 13.4 Schematic illustration of the technique used by *Patterson*. [13.19] to determine the shape of the auditory filter. The threshold of the sinusoidal signal (indicated by the *bold vertical line*) is measured as a function of the width of a spectral notch in the noise masker. The amount of noise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

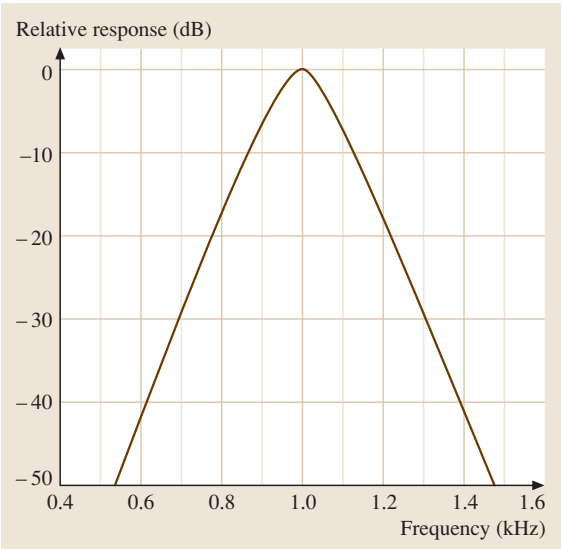


Fig. 13.5 A typical auditory filter shape determined using Patterson's method. The filter is centered at 1 kHz. The relative response of the filter (in dB) is plotted as a function of frequency

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

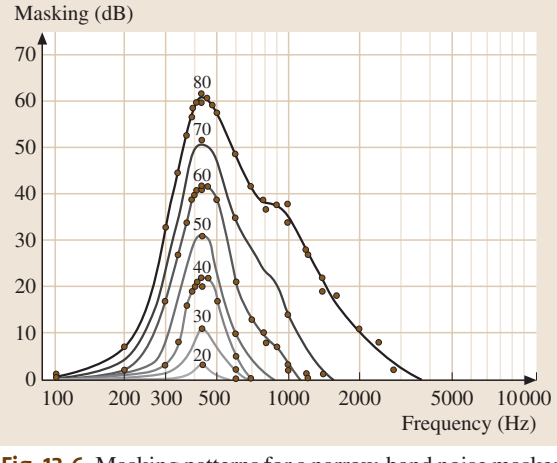


Fig. 13.6 Masking patterns for a narrow-band noise masker centered at 410 Hz. Each curve shows the elevation in threshold of a pure-tone signal as a function of signal frequency. The overall noise level for each curve is indicated in the figure (After *Egan* and *Hake* [13.21])

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.

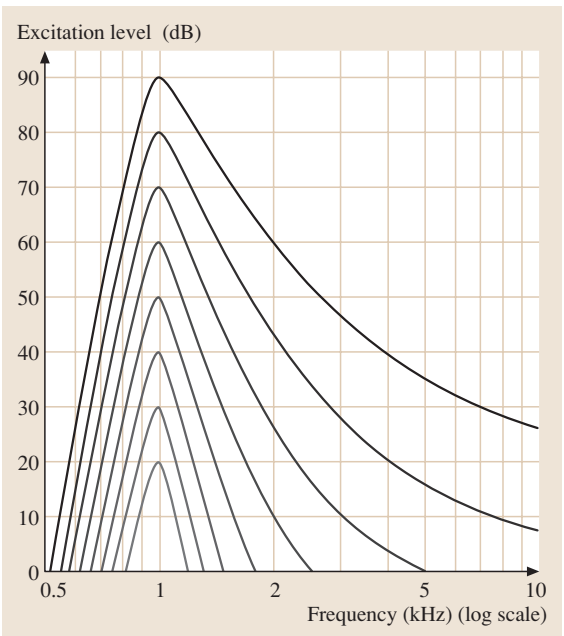


Fig. 13.7 Calculated psychoacoustical excitation patterns for a 1 kHz sinusoid at levels ranging from 20 to 90 dB SPL in 10 dB steps

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-

Amazon EX1014 Page 476

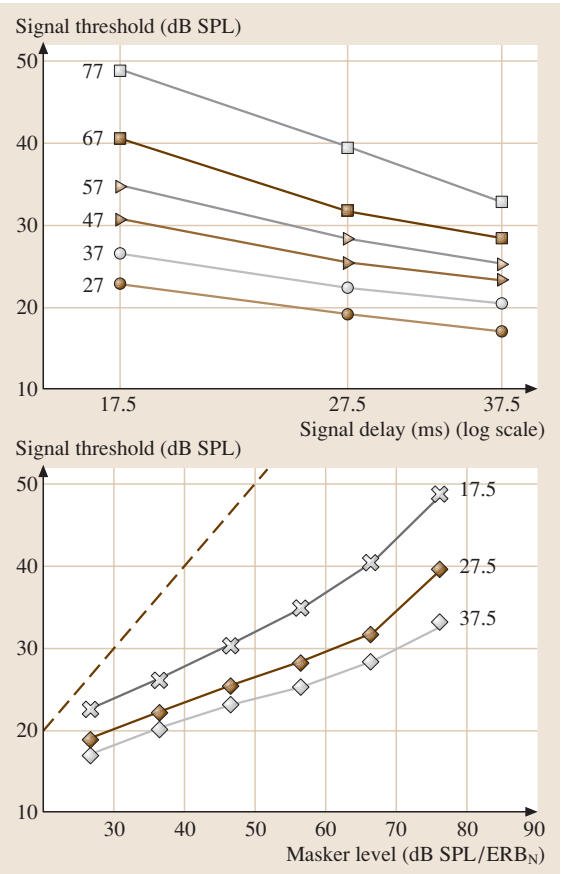


Fig. 13.8 The top panel shows the amount of forward masking of a brief 4 kHz signal, plotted as a function of the time delay of the signal after the end of the noise masker. Each curve shows results for a different noise level, specified as the level in a one-ERB$_N$-wide band centered at 4 kHz. The 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 EX1014 Page 477

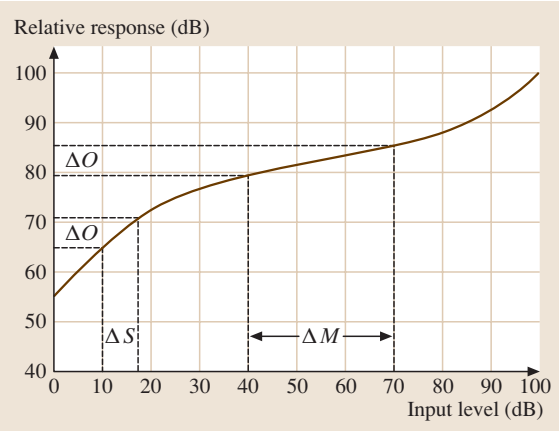


Fig. 13.9 Illustration of why growth of masking functions in forward masking usually have shallow slopes. The *solid curve* shows a schematic input–output function on the basilar membrane. The relative response is plotted on a dB scale with an arbitrary origin. When the masker is increased in level by 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-

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.

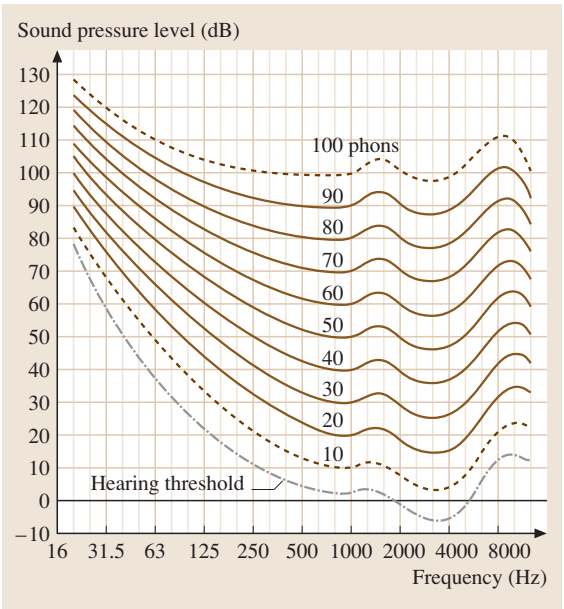


Fig. 13.10 Equal-loudness contours for loudness levels from 10 to 100 phons for sounds presented binaurally from the frontal direction. The absolute threshold curve (the MAF) is also shown. The curves for loudness levels of 10 and 100 phons are *dashed*, as they are based on interpolation and extrapolation,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

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.

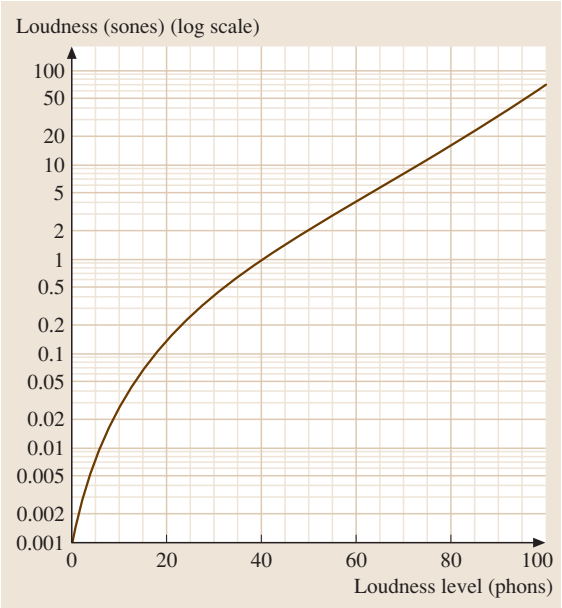


Fig. 13.11 The relationship between loudness in sones and loudness level in phons for a 1000 Hz sinusoid. The curve is based on the loudness model of *Moore* et al.[13.3]

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

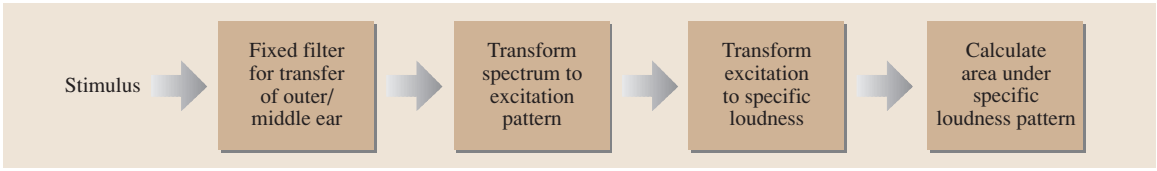


Fig. 13.12 Block diagram of a typical loudness model

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 EX1014 Page 481

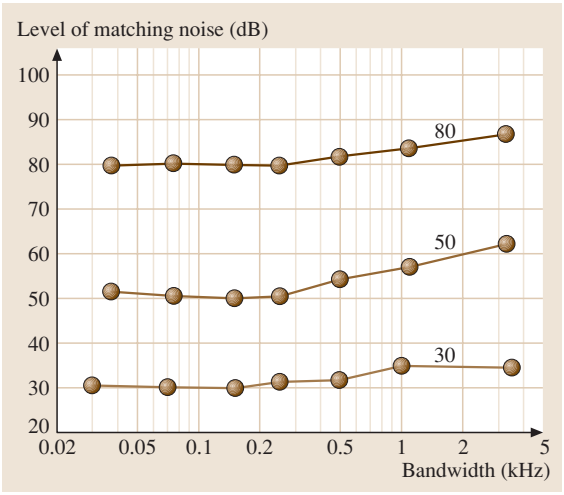


Fig. 13.13 The level of a 210 Hz-wide noise required to match the loudness of a noise of variable bandwidth is plotted as a function of the variable bandwidth; the bands of noise were geometrically centered at 1420 Hz. The number by each curve is the overall noise level in dB SPL. (After *Zwicker* et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

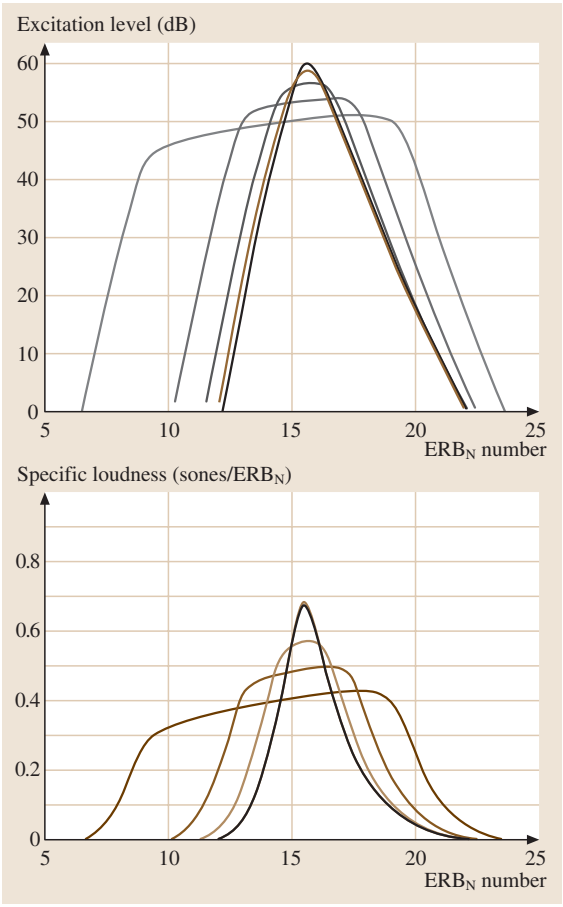


Fig. 13.14 The upper panel shows excitation patterns for a 1 kHz sinusoid with a level of 60 dB SPL (the narrowest pattern with the highest peak) and for noise bands of various widths, all centered at 1 kHz and with an overall level of 60 dB SPL. The frequency scale has been transformed to an ERB_N number 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

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 EX1014 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.

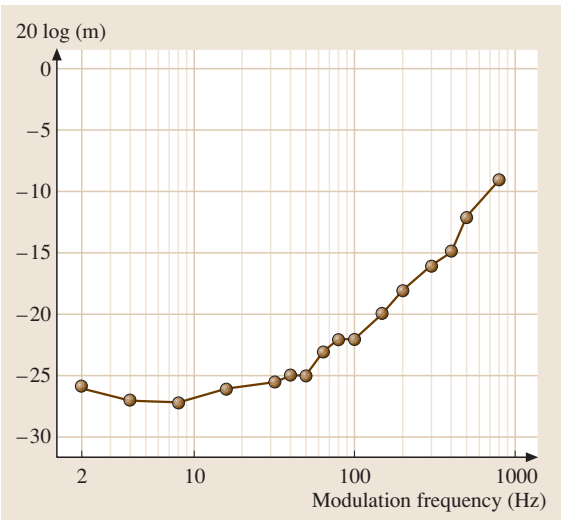


Fig. 13.15 A temporal modulation transfer function (TMTF). A broadband white noise was sinusoidally amplitude modulated, and the threshold amount of modulation required for detection is plotted as a function of modulation rate. The amount of modulation is specified as $20\log(m)$, where m is the modulation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

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.

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 EX1014 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.

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 EX1014 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.

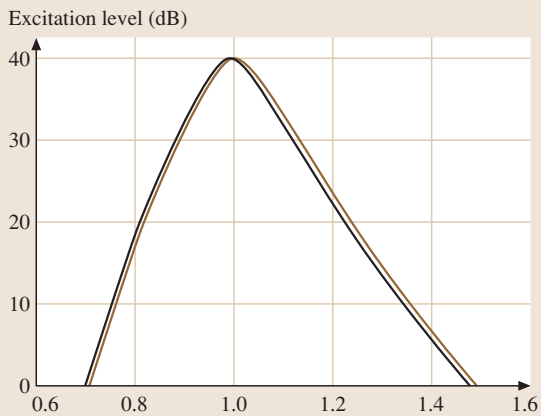


Fig. 13.16 Schematic illustration of an excitation-pattern model for frequency discrimination. Excitation patterns are shown for two sinusoidal tones differing slightly in frequency; the two tones have frequencies of 995 Hz and 1005 Hz. It is assumed that the difference in frequency, ΔF, can be detected 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 ERB_N, rather than in terms of linear frequency. The slope is about 18 dB per ERB_N. To achieve a change in excitation level of 1 dB, the frequency has to be changed by one eighteenth of an 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 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 EX1014 Page 489

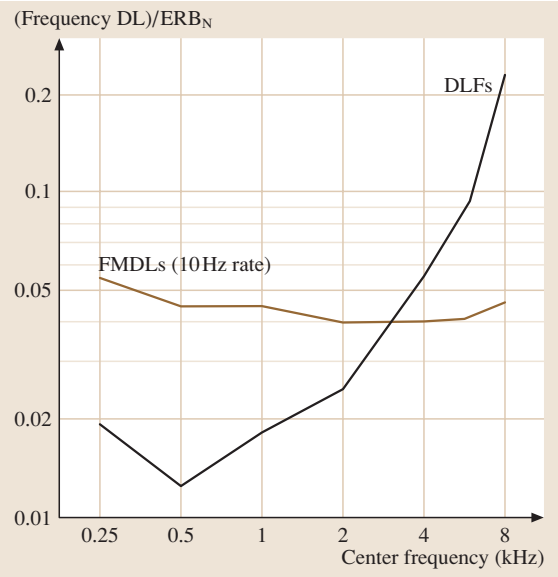


Fig. 13.17 Thresholds for detecting differences in frequency between steady pulsed tones (DLFs) and for detecting frequency modulation (FMDLs), plotted relative to the ERB$_N$ of the auditory filter at each center frequency. (After *Sek* and *Moore* [13.103])

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 EX1014 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 EX1014 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.

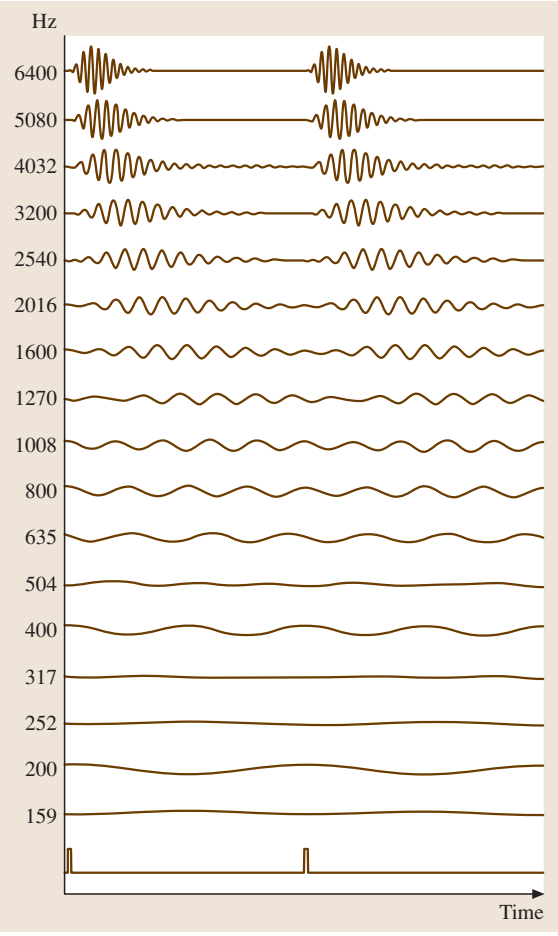


Fig. 13.18 A simulation of the responses on the basilar membrane to periodic impulses of rate 200 pulses per second. The input waveform is shown at the bottom; impulses occur every 5 ms. Each number on the left represents the frequency which would maximally excite a given point on the basilar membrane. 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 EX1014 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 EX1014 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

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

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 EX1014 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-

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.

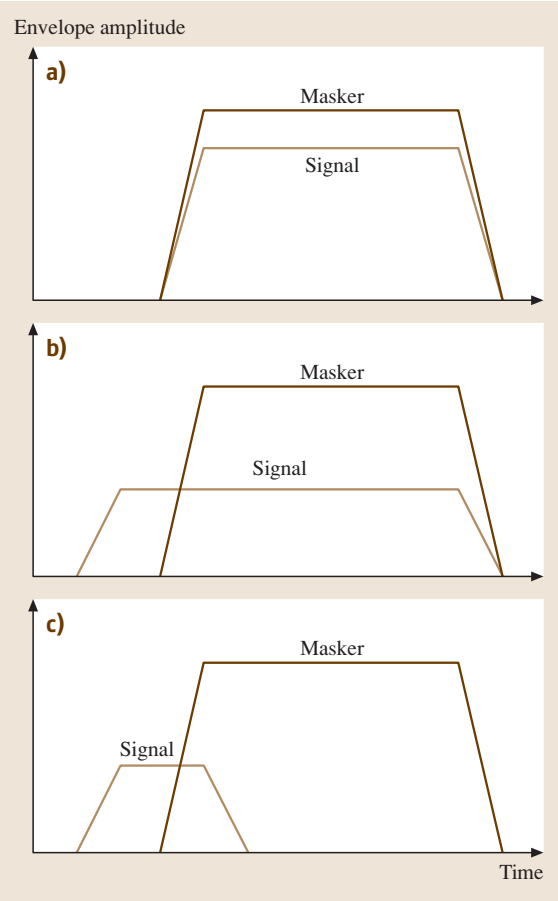


Fig. 13.19a–c Schematic illustration of the stimuli used by *Rasch* [13.172]. Both the signal and the masker were periodic complex tones, with the signal having the higher fundamental frequency. When the signal and masker were gated on and off synchronously **(a)**, the threshold for the signal was relatively 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 EX1014 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-

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 EX1014 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

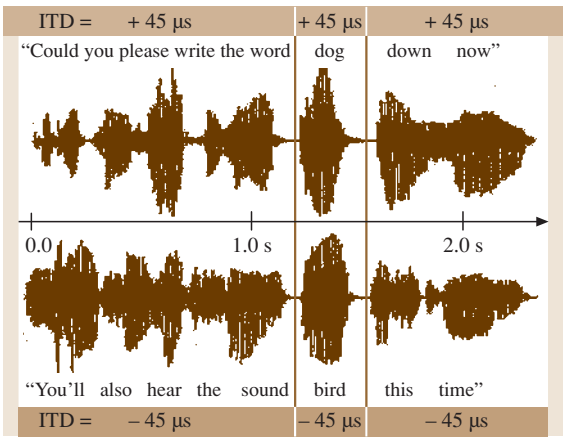


Fig. 13.20 Example of the stimuli used by *Darwin* and *Hukin* [13.198] (After Fig. 1 of [13.198])

Amazon EX1014 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 EX1014 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

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.

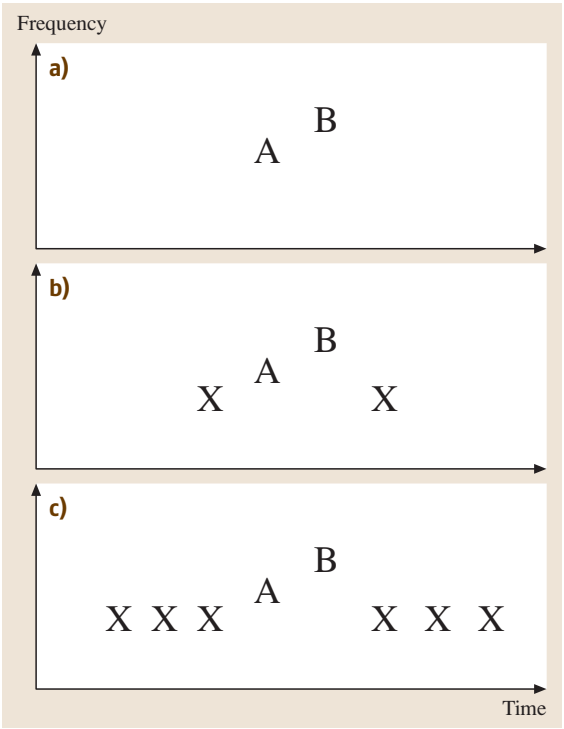


Fig. 13.21 Schematic illustration of the stimuli used by *Bregman* and *Rudnicky* [13.225]. When the tones A and B are presented alone *(panel a)*, it is easy to tell their order. When the tones A and B are presented as part of a four-tone complex XABX *(panel b)*, it is more difficult to tell their 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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,

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

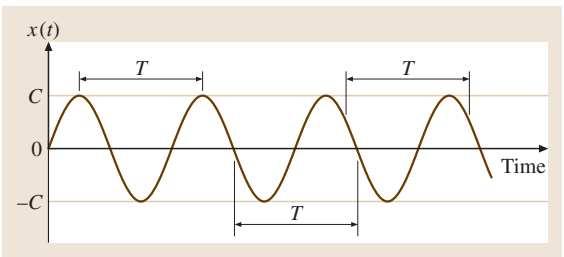


Fig. 14.1 A sine wave with amplitude C and period T. A little more than three and a half cycles are shown. The starting phase is $\varphi = 0$

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 EX1014 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 EX1014 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

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

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

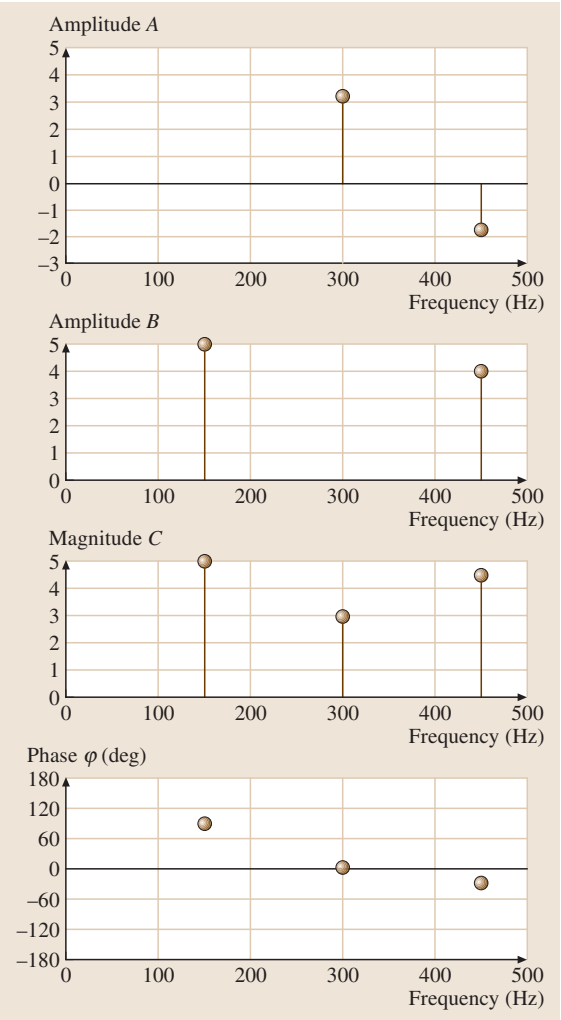


Fig. 14.2 The amplitudes A and B for the signal in (14.22) are shown in the *top* two plots. The corresponding magnitude and phases are shown in the *bottom* two

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 EX1014 Page 516

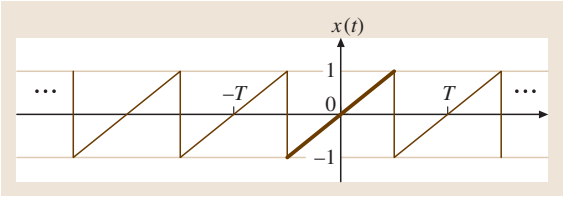


Fig. 14.3 The Fourier series of an odd function like this sawtooth consists of sine terms only. The Fourier coefficients can be computed by an integral over a single period from $-T/2$ to $T/2$

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 EX1014 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

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

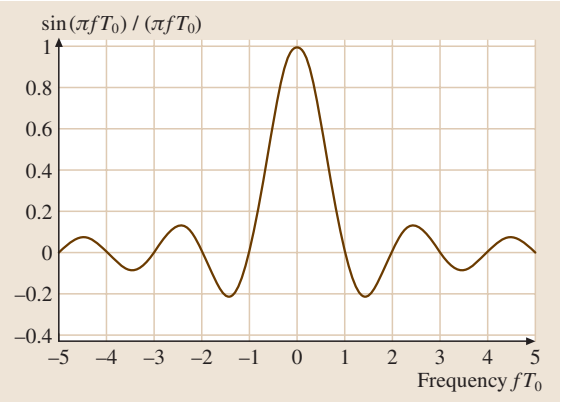


Fig. 14.4 The Fourier transform of a single pulse with duration T_0 as a function of frequency f expressed in dimensionless form fT_0

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 EX1014 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 EX1014 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 EX1014 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 EX1014 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.

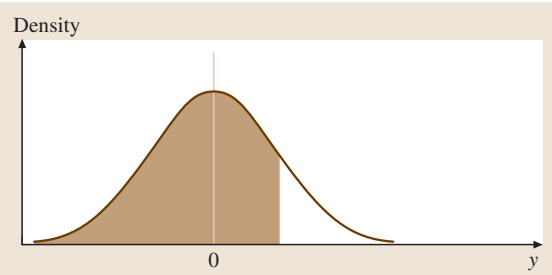


Fig. 14.5 The *area* under the normal density is the cumulative normal. Here the area is the function $C(y)$

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 EX1014 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

$$\begin{aligned} &\mathcal{H}\left[\sum_n C_n \sin(\omega_n t + \varphi_n)\right] \\ &= -\sum_n C_n \cos(\omega_n t + \varphi_n) \\ &= \sum_n C_n \sin(\omega_n t + \varphi_n - \pi/2) . \end{aligned} \tag{14.91}$$

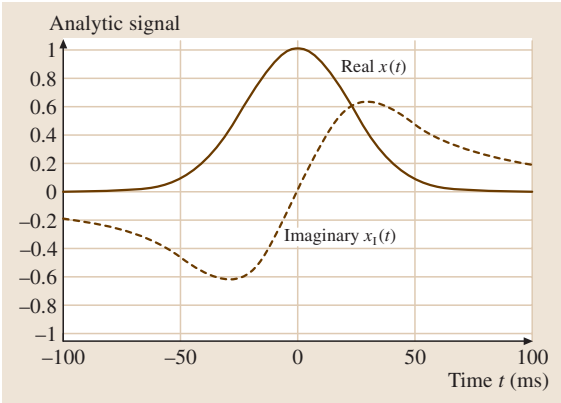


Fig. 14.6 A Gaussian pulse $x(t)$ and its Hilbert transform $x_\mathrm{I}(t)$ are the real and imaginary parts of the analytic signal corresponding to the Gaussian pulse

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

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 EX1014 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

$$H(\omega) = \frac{\omega_o^2}{\omega_o^2 - \omega^2 + j\omega\omega_o/Q}\,. \tag{14.106}$$

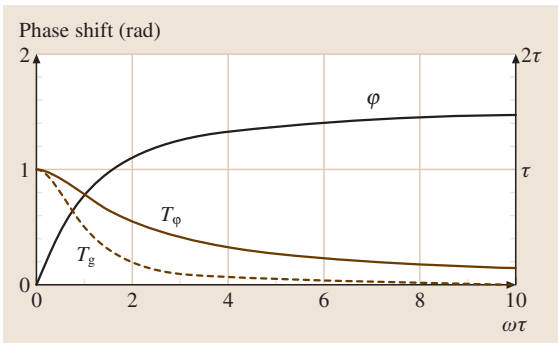


Fig. 14.7 The phase shift Φ for a one-pole low-pass filter can be read on the *left ordinate*. The phase and group delays can be read on the *right ordinate*

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

$$\begin{aligned} h(t) = {} & \frac{\omega_o}{\sqrt{1-[1/(2Q)]^2}}\,\mathrm{e}^{-\frac{\omega_o}{2Q}t} \\ & \times \sin\left\{\omega_o t\sqrt{1-[1/(2Q)]^2}\right\}\,, \quad t \geq 0\,, \\ h(t) = {} & 0\,, \quad t < 0\,. \end{aligned} \tag{14.110}$$

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.

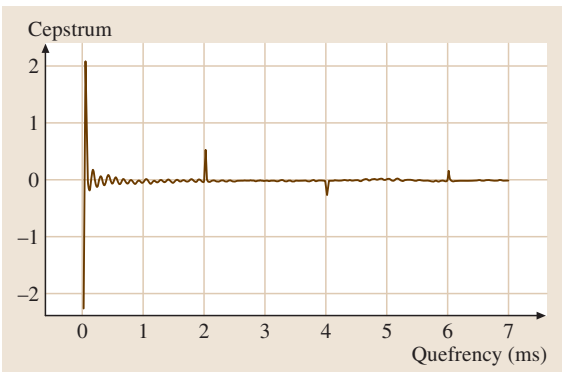


Fig. 14.8 The cepstrum of an original signal to which is added a delayed version of the same signal, with a delay of 2 ms ($a = 1$). The original signal is the sum of two female talkers

Amazon EX1014 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 EX1014 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 EX1014 Page 529

14.9.5 Equal-Amplitude Random-Phase Noise

Equal-amplitude random-phase (EARP) noise is of the form

$$x(t) = C \sum_{n=1}^{N} \cos(\omega_n t + \varphi_n) \,, \qquad (14.127)$$

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

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.

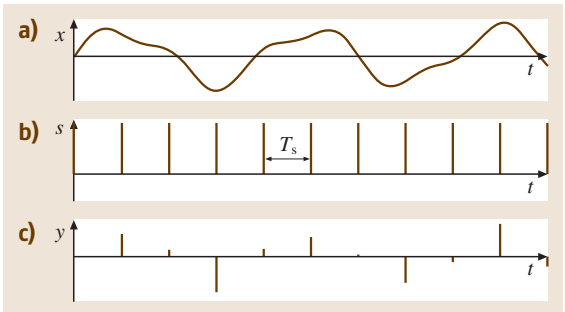


Fig. 14.9 An analog signal $x(t)$ is multiplied by a train of delta functions $s(t)$ to produce a sampled signal $y(t)$

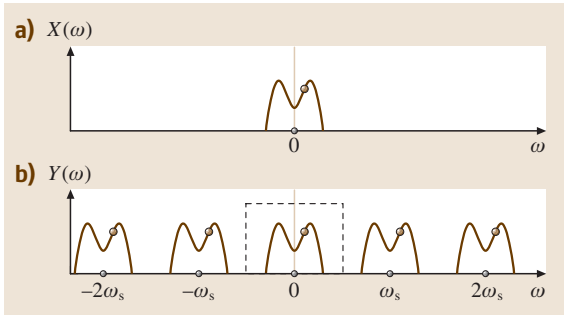


Fig. 14.10 **(a)** The spectrum of the analog signal $X(\omega)$ is bounded in frequency. **(b)** The spectrum of the sampled signal, $Y(\omega)$, is the convolution of $X(\omega)$ and the Fourier transform of the sampling train of delta functions. Consequently, multiple images of $X(\omega)$ appear. 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 EX1014 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 EX1014 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

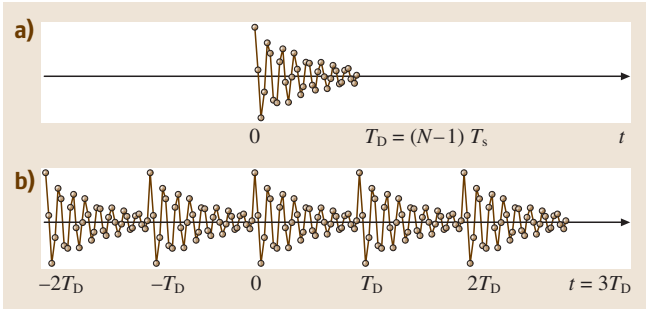


Fig. 14.11a,b A decaying function in part **(a)** is periodically repeated in part **(b)** to create a periodic signal with period T_D

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

limits,

$$X(\omega) = \frac{1}{T_\mathrm{D}} \int_{-T_\mathrm{D}/2}^{T_\mathrm{D}/2} \mathrm{d}t' \, \mathrm{e}^{-\mathrm{i}\omega t'} \times \sum_{n=-(N-1)/2}^{(N-1)/2} X(n\omega_o) \, \mathrm{e}^{\mathrm{i}n\omega_o t'} \, \mathrm{e}^{-\mathrm{i}\omega T_\mathrm{D}/2} \, \mathrm{e}^{\mathrm{i}n\omega_o T_\mathrm{D}/2} , \tag{14.138}$$

which reduces to

$$X(\omega) = \sum_{n=-(N-1)/2}^{(N-1)/2} X(n\omega_o) \frac{\sin[(\omega - n\omega_o)T_\mathrm{D}/2]}{(\omega - n\omega_o)T_\mathrm{D}/2} \times \mathrm{e}^{-\mathrm{i}\omega T_\mathrm{D}/2} \, \mathrm{e}^{\mathrm{i}\pi n} . \tag{14.139}$$

Part D | 14.12

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 $t = t_k = kT_\mathrm{s}$, 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,

$$z = A\,\mathrm{e}^{\mathrm{i}\varphi} = A\,\mathrm{e}^{\mathrm{i}\omega T_\mathrm{s}} , \tag{14.141}$$

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,

$$\omega = \omega_\mathrm{R} + \mathrm{i}\omega_\mathrm{I} , \tag{14.142}$$

so that

$$z = \mathrm{e}^{-\omega_\mathrm{I} T_\mathrm{s}} \, \mathrm{e}^{\mathrm{i}\omega_\mathrm{R} T_\mathrm{s}} . \tag{14.143}$$

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
$a^k u_k$	$z/(z-a)$	$\lvert z\rvert > a$
$ka^k u_k$	$az/(z-a)^2$	$\lvert z\rvert > a$
$a^k \cos(\omega_o T_\mathrm{s} k) u_k$	$\frac{z^2 - az\cos(\omega_o T_\mathrm{s})}{z^2 - 2az\cos(\omega_o T_\mathrm{s}) + a^2}$	$\lvert z\rvert > a$
$a^k \sin(\omega_o T_\mathrm{s} k) u_k$	$\frac{az\sin(\omega_o T_\mathrm{s})}{z^2 - 2az\cos(\omega_o T_\mathrm{s}) + a^2}$	$\lvert z\rvert > a$

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

$$\begin{aligned} x_k &= 2^k \qquad \text{for} \quad k \geq 0 , \\ x_k &= 0 \qquad \text{for} \quad k < 0 . \end{aligned} \tag{14.144}$$

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 EX1014 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	$x_k y_k$	$\frac{1}{2\pi \mathrm{i}} \oint_C \mathrm{d}z'/z'\ X(z')Y(z/z')$

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

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

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 14.6 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 EX1014 Page 536

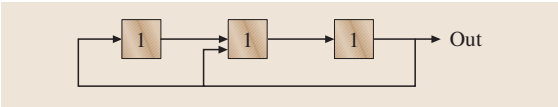


Fig. 14.12 A three-stage shift register [3: 1, 2] in which the output is fed back into the first and second stages

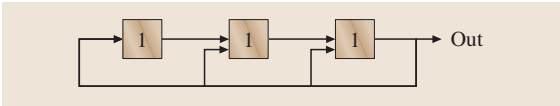


Fig. 14.13 A three-stage shift with feedback into all the stages does not produce an MLS

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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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 EX1014 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

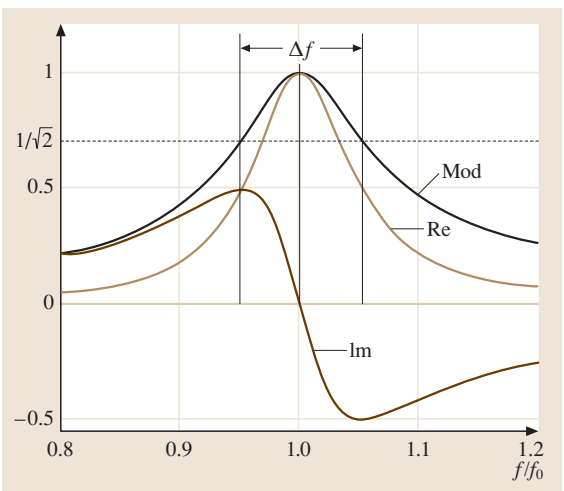


Fig. 15.1 Normalised real (Re) and imaginary (Im) components and the modulus (Mod) of the induced velocity of a simple harmonic resonator driven by a constant amplitude sinusoidal force for a Q-factor of 10

point p, for an applied sinusoidal force $F\,\mathrm{e}^{\mathrm{i}\omega t}$, is then given by

$$\frac{\partial \xi_n}{\partial t} = \mathrm{i}\omega\xi_n = \frac{F}{m_n}\frac{\mathrm{i}\omega}{\left(\omega_n^2 - \omega^2 + \mathrm{i}\omega\omega_n/Q_n\right)}. \qquad (15.2)$$

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 $\mathrm{i}\omega F\xi_n$ 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

$$A_{pp} = \sum_n \frac{1}{m_n}\frac{\mathrm{i}\omega}{\omega_n^2 - \omega^2 + \mathrm{i}\omega\omega_n/Q_n}, \qquad (15.3)$$

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

$$A_{pq} = \sum_n \frac{1}{m_{n,p}}\frac{\mathrm{i}\omega}{\left(\omega_n^2 - \omega^2 + \mathrm{i}\omega\omega_n/Q_n\right)}\frac{\xi_{n,p}\xi_{n,q}}{\xi_{n,p}^2}, \qquad (15.4)$$

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 EX1014 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.

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.

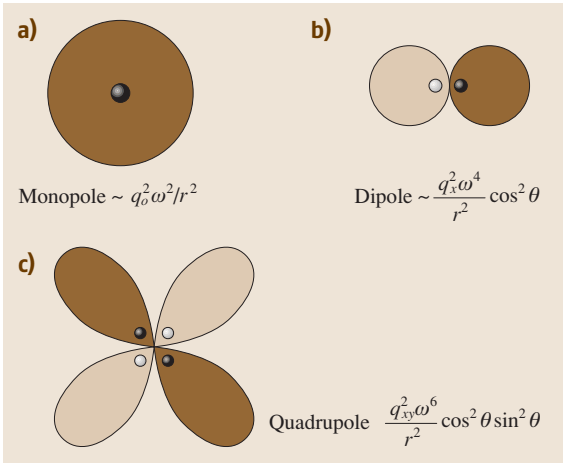


Fig. 15.2 Typical radiation patterns and intensities for monopole, dipole and quadrupole sources. The *two colours* represent monopole sources and sound pressures of opposite signs

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 EX1014 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

xy-plane with displacements in the z-direction varying as $\sin(k_x x)\mathrm{e}^{\mathrm{i}\omega t}$. We look for propagating sound waves solutions radiating from the surface of the form $\sin(k_x x)\mathrm{e}^{\mathrm{i}(\omega t - k_z z)}$, which must satisfy the wave equation and hence the relationship,

$$k_z^2 = \frac{\omega^2}{c_0^2} - k_x^2 = \omega^2\left(\frac{1}{c_0^2} - \frac{1}{c_\mathrm{m}^2}\right), \qquad (15.18)$$

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 $c_\mathrm{m} > c_0$. 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 $\mathrm{e}^{-z/\delta}$ 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,

$$a\cos(2\pi f t + \phi) = a\cos(\omega t + \phi) = \mathrm{Re}(\boldsymbol{a}\,\mathrm{e}^{\mathrm{i}\omega t}), \qquad (15.19)$$

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 EX1014 Page 549

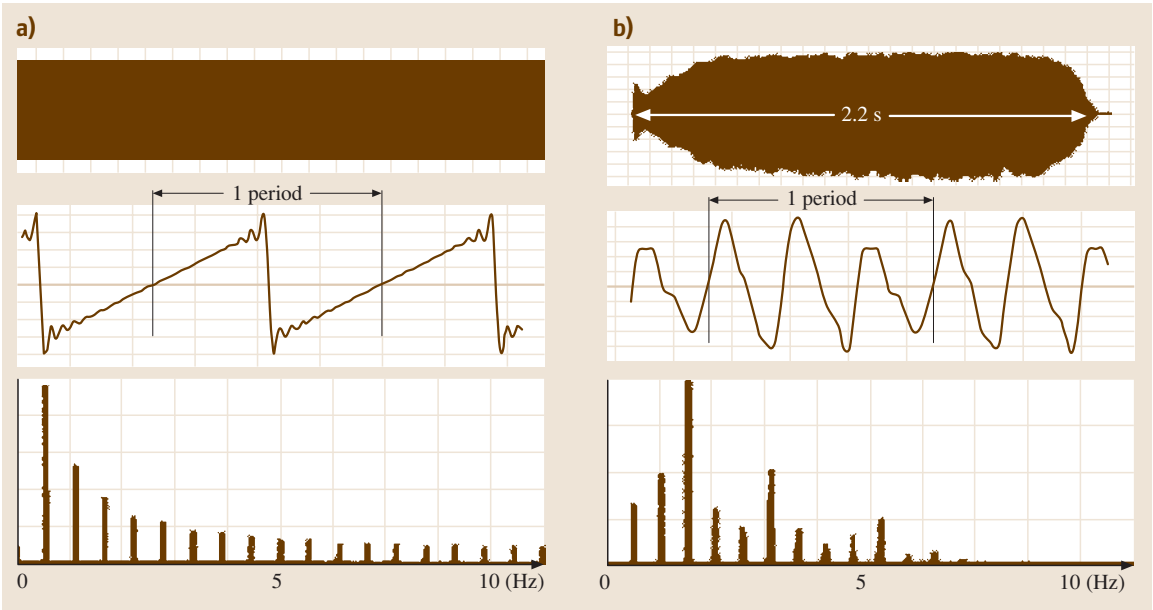


Fig. 15.3a,b Comparison of the envelope, repetitive waveform and spectrum of **(a)** a synthesised sawtooth and **(b)** a note played on an oboe

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

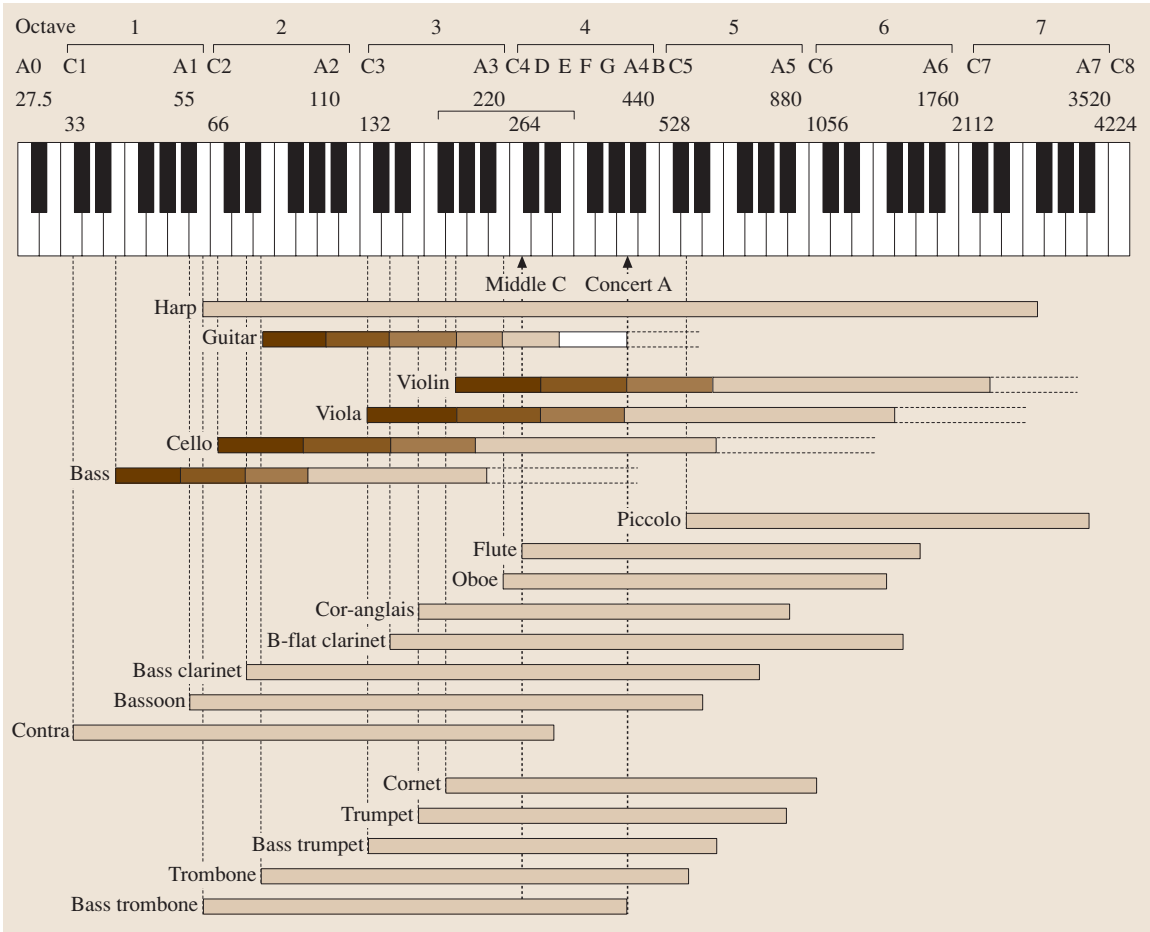


Fig. 15.4 Notation used for notes of the musical scale and the playing range of classical western musical instruments. Subdivisions for stringed instruments represent the tuning of the open strings

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 EX1014 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

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

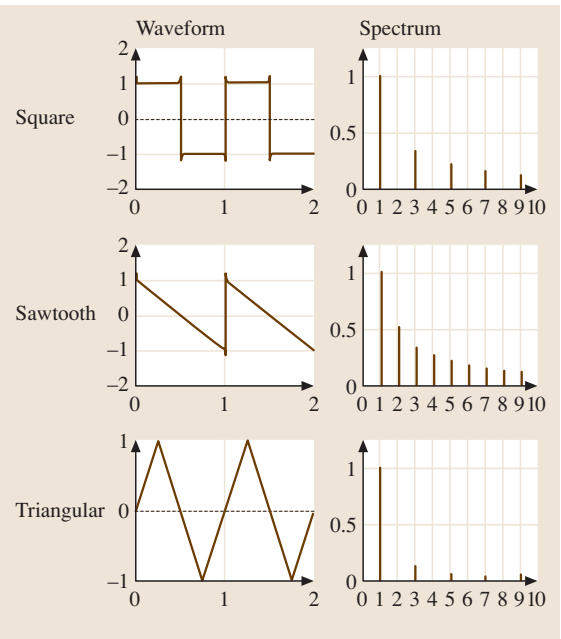


Fig. 15.6 Typical repetitive waveforms (synthesised from the first 100 components of a Fourier series) and the amplitudes of the first few partials normalised to the amplitude of the fundamental

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	$\frac{2}{\pi}\left(\sin\omega_1 t - \frac{1}{3^2}\sin 3\omega_1 t + \frac{1}{5^2}\sin 5\omega_1 t \ldots\right)$

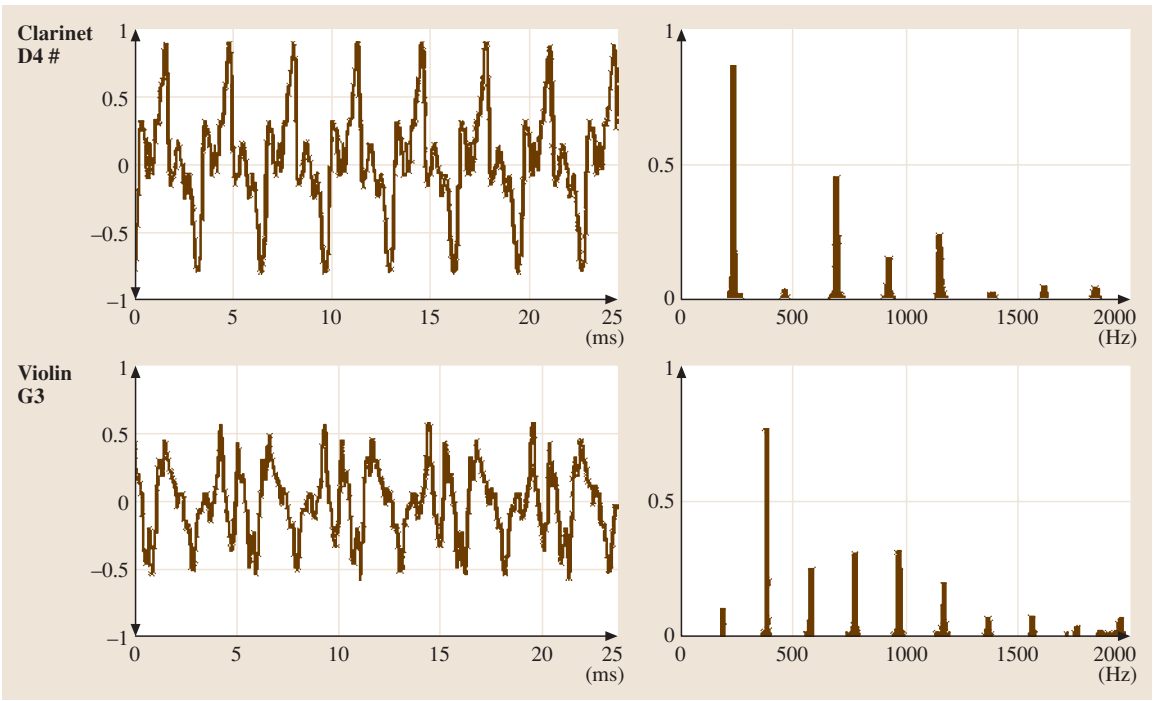


Fig. 15.7 Short-period samples of clarinet (D#4) and bowed violin (G3) tones and the corresponding Fourier spectra. The *vertical scales* are linear

Amazon EX1014 Page 554