

FUNDAMENTALS OF ACOUSTICS

THIRD EDITION

LAWRENCE E. KINSLER

Late Professor Emeritus
Naval Postgraduate School

AUSTIN R. FREY

Professor Emeritus
Naval Postgraduate School

ALAN B. COPPENS

Associate Professor of Physics
Naval Postgraduate School

JAMES V. SANDERS

Associate Professor of Physics
Naval Postgraduate School



John Wiley & Sons

New York Chichester Brisbane Toronto Singapore

LGE EXHIBIT 1004

Amazon v. Jawbone
U.S. Patent 7,246,058
Amazon Ex. 1004

F

P5L
QC
243
F86
1982

It is
co-a
revis
Fuñ
acou
writi
Nav
me.
the
Frey
by E
date

Copyright © 1950, 1962, 1982, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of
this work beyond that permitted by Sections
107 and 108 of the 1976 United States Copyright
Act without the permission of the copyright
owner is unlawful. Requests for permission
or further information should be addressed to
the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:
Fundamentals of acoustics.

Includes index.

1. Sound-waves. 2. Sound--Apparatus. 3. Archi-
tectural acoustics. 4. Underwater acoustics.

I. Kinsler, Lawrence E.

QC243.F86 1982 534 81-7463

ISBN 0-471-02933-5 AACR2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

FOREWORD

It is with great regret that I report the untimely death of my good friend and co-author, Larry Kinsler. It is Dr. Kinsler who was primarily responsible for the revisions and alterations of the original text that appeared in the second edition of *Fundamentals of Acoustics*.

For the past thirty years my own interests have been in fields other than acoustics, and as a consequence I have not felt competent to undertake the task of writing a third edition. I have therefore asked two of my younger associates at the Naval Postgraduate School, Dr. A. B. Coppens and Dr. J. V. Sanders to do this for me. Both are highly competent acousticians, who have done important research in the field and have used both the original and the second editions of "Kinsler and Frey" in their teaching at the Postgraduate School.

I hope that the reader will agree with me that the changes and additions made by Drs. Coppens and Sanders have improved this text and have brought it up to date in this rapidly advancing field of physics and engineering.

Austin R. Frey

Monterey, California
January 1980.

CONTENTS

CHAPTER I FUNDAMENTALS OF VIBRATION

1.1	Introduction	1
1.2	The Simple Oscillator	2
1.3	Initial Conditions	3
1.4	Energy of Vibration	4
1.5	Complex Exponential Method of Solution	5
1.6	Damped Oscillations	7
1.7	Forced Oscillations	10
1.8	Transient Response of an Oscillator	12
1.9	Power Relations	14
1.10	Mechanical Resonance	15
1.11	Mechanical Resonance and Frequency	17
*1.12	Equivalent Electrical Circuit for a Simple Oscillator	19
1.13	Linear Combinations of Simple Harmonic Vibrations	20
1.14	Analysis of Complex Vibrations by Fourier's Theorem	23
*1.15	Transients and Fourier Analysis	25

CHAPTER 2 TRANSVERSE MOTION— THE VIBRATING STRING

2.1	Vibration of Extended Systems	30
2.2	Transverse Waves on a String	30
2.3	The One-Dimensional Wave Equation	30
2.4	General Solution of the Wave Equation	32
2.5	Wave Nature of the General Solution	33
2.6	Initial Values and Boundary Conditions	33
2.7	Reflection at a Boundary	34
2.8	Forced Vibrations of an Infinite String	36
2.9	Forced Vibrations of a String of Finite Length	39
(a)	<i>The Forced, Fixed String</i>	39
(b)	<i>The Forced, Mass-Loaded String</i>	42
*(c)	<i>The Forced, Resistance-Loaded String</i>	44

x CONTENTS

2.10	Normal Modes of the Fixed, Fixed String	45
*2.11	Effects of More Realistic Boundary Conditions on the Freely Vibrating String	48
	(a) <i>The Fixed, Mass-Loaded String</i>	49
	(b) <i>The Fixed, Resistance-Loaded String</i>	49
2.12	Energy of Vibration of a String	52
2.13	Overtones and Harmonics	53

CHAPTER 3 VIBRATIONS OF BARS

3.1	Longitudinal Vibrations of a Bar	57
3.2	Longitudinal Strain	57
3.3	Longitudinal Wave Equation	58
3.4	Simple Boundary Conditions	60
3.5	The Free, Mass-Loaded Bar	62
*3.6	The Freely Vibrating Bar: General Boundary Conditions	64
3.7	Forced Vibrations of a Bar: Resonance and Antiresonance Revisited	65
*3.8	Transverse Vibrations of a Bar	68
*3.9	Transverse Wave Equation	70
*3.10	Boundary Conditions	72
	(a) <i>Clamped End</i>	72
	(b) <i>Free End</i>	72
*3.11	Bar Clamped at One End	72
*3.12	Bar Free at Both Ends	75

CHAPTER 4 THE TWO-DIMENSIONAL WAVE EQUATION: VIBRATIONS OF MEMBRANES AND PLATES

4.1	Vibration of a Plane Surface	78
4.2	The Wave Equation for a Stretched Membrane	78
4.3	Free Vibrations of a Rectangular Membrane with Fixed Rim	80
4.4	Free Vibrations of a Circular Membrane Fixed at the Rim	82
4.5	Symmetric Vibrations of a Circular Membrane Fixed at the Rim	85
*4.6	The Kettledrum	87
4.7	Forced Vibrations of a Membrane	89
*4.8	Vibration of Thin Plates	92

CHAPTER 5 THE ACOUSTIC WAVE EQUATION AND SIMPLE SOLUTIONS

5.1	Introduction	98
5.2	The Equation of State	100
5.3	The Equation of Continuity	101

45	5.4 The Simple Force Equation: Euler's Equation	102
	5.5 The Linearized Wave Equation	104
48	5.6 Speed of Sound in Fluids	105
49	5.7 Harmonic Plane Waves	107
49	5.8 Energy Density	108
52	5.9 Acoustic Intensity	110
53	5.10 Specific Acoustic Impedance	111
	5.11 Spherical Waves	112
	5.12 Decibel Scales	115
	*5.13 Rays and Waves	117
	*5.14 The Inhomogeneous Wave Equation	120
	*5.15 The Point Source	121

CHAPTER 6 TRANSMISSION PHENOMENA

64	6.1 Changes in Media	124
65	6.2 Transmission from One Fluid to Another: Normal Incidence	125
68	6.3 Transmission through a Layer: Normal Incidence	127
70	6.4 Transmission from One Fluid to Another: Oblique Incidence	131
72	6.5 Reflection at the Surface of a Solid: Normal Incidence	133
72	*6.6 Reflection at the Surface of a Solid: Oblique Incidence	137

CHAPTER 7 ABSORPTION AND ATTENUATION OF SOUND WAVES IN FLUIDS

78	7.1 Introduction	141
78	7.2 A Phenomenological Approach to Absorption	142
80	7.3 Complex Sound Speed	145
82	7.4 The Classical Absorption Coefficient	146
85	7.5 Molecular Thermal Relaxation	149
87	7.6 Absorption Phenomena in Liquids	157
89	*7.7 Attenuation in Inhomogeneous Fluids	160

CHAPTER 8 RADIATION AND RECEPTION OF ACOUSTIC WAVES

98	8.1 Radiation from a Pulsating Sphere	163
100	8.2 Source Strength	164
101	8.3 Acoustic Reciprocity	165
	8.4 Simple Sources	167
	8.5 Dipole Radiation	169
	8.6 The Continuous-Line Source	172
	8.7 Directional Factor and Beam Patterns	175

xii CONTENTS

8.8	Radiation from a Plane Circular Piston	176
(a)	<i>Axial Response</i>	176
(b)	<i>Far Field</i>	178
8.9	Beam Width, Source Level, Directivity, and Directivity Index	182
(a)	<i>Beam Width</i>	182
(b)	<i>Source Level</i>	182
(c)	<i>Directivity</i>	183
(d)	<i>Directivity Index</i>	184
*8.10	Directional Factors of Reversible Transducers	185
8.11	Rough Estimates of Radiation Patterns for Sources of Simple Geometry	187
(a)	<i>Extent of the Near Field</i>	187
(b)	<i>Behavior of the Far Field</i>	188
(c)	<i>Estimation of Directivity</i>	189
8.12	Radiation Impedance	191
*8.13	The Simple Line Array	193
*8.14	The Product Theorem	197

CHAPTER 9 PIPES, CAVITIES, AND WAVEGUIDES

9.1	Introduction	200
9.2	Resonance in Pipes	200
9.3	Power Radiation from Open-Ended Pipes	203
9.4	Standing Wave Patterns	204
9.5	Absorption of Sound in Pipes	206
9.6	Behavior of the Combined Driver-Pipe System	210
9.7	The Rectangular Cavity	214
*9.8	The Waveguide of Constant Cross Section	216

CHAPTER 10 RESONATORS, DUCTS, AND FILTERS

10.1	The Long-Wavelength Limit	225
10.2	The Helmholtz Resonator	225
*10.3	The Resonant Bubble	228
10.4	Acoustic Impedance	231
(a)	<i>Lumped Acoustic Impedance</i>	231
(b)	<i>Distributed Acoustic Impedance</i>	232
10.5	Waves in a Pipe	234
10.6	Acoustic Filters	237
(a)	<i>Low-Pass Filters</i>	237
(b)	<i>High-Pass Filters</i>	239
(c)	<i>Band-Pass Filters</i>	241
*10.7	Filter Networks	242

CHAPTER 11 NOISE, SIGNAL DETECTION, HEARING, AND SPEECH

11.1	Introduction	246
11.2	Noise, Spectrum Level, and Band Level	246
11.3	Combining Band Levels and Tones	250
*11.4	Detecting Signal in Noise	252
*11.5	Detection Threshold	255
11.6	The Ear	257
11.7	Some Fundamental Properties of the Human Ear	262
(a)	<i>Thresholds</i>	262
(b)	<i>Equal Loudness Level Contours</i>	264
(c)	<i>Critical Bandwidth</i>	264
(d)	<i>Masking</i>	266
(e)	<i>Cochlear Nonlinear Effects: Beats, Combination Tones, and Aural Harmonics</i>	267
(f)	<i>Nonlinear Processing Effects: Consonance and the Restored Fundamental</i>	268
11.8	Loudness Level and Loudness	270
11.9	Pitch and Frequency	273
*11.10	The Voice	274

CHAPTER 12 ENVIRONMENTAL ACOUSTICS

12.1	Introduction	279
12.2	Weighted Sound Levels	280
12.3	Speech Interference	283
12.4	Privacy	285
12.5	Noise Rating Curves	285
12.6	The Statistical Description of Community Noise	287
12.7	Criteria for Community Noise	291
*12.8	Highway Noise	293
*12.9	Aircraft Noise Rating	295
*12.10	Community Response to Noise	296
12.11	Noise-Induced Hearing Loss	298
12.12	Noise and Architectural Design	302
12.13	Specification and Measurement of Sound Isolation	303
12.14	Recommended Isolation	306
12.15	Design of Partitions	306
(a)	<i>Single-Leaf Partitions</i>	307
(b)	<i>Double-Leaf Partitions</i>	308
(c)	<i>Doors and Windows</i>	311

CHAPTER 13 ARCHITECTURAL ACOUSTICS

13.1	Sound in Enclosures	313
13.2	A Simple Model for the Growth of Sound in a Room	314
13.3	Reverberation Time—Sabine	317
13.4	Reverberation Time—Other Formulas	319
13.5	Sound Absorption Materials	321
13.6	Measurement of the Acoustic Output of Sound Sources in Live Rooms	324
13.7	Direct and Reverberant Sound	325
13.8	Acoustic Factors in Architectural Design	326
	(a) <i>The Direct Arrivals</i>	326
	(b) <i>Reverberation at 500 Hz</i>	326
	(c) <i>Warmth</i>	331
	(d) <i>Intimacy</i>	331
	(e) <i>Diffusion, Blend, and Ensemble</i>	332
*13.9	Standing Waves and Normal Modes in Enclosures	333
	(a) <i>The Rectangular Enclosure</i>	333
	(b) <i>Damped Normal Modes</i>	333
	(c) <i>The Driven Standing Wave</i>	334
	(d) <i>Growth and Decay of Sound from a Source</i>	335
	(e) <i>Frequency Distribution of Enclosure Resonances</i>	335
	(f) <i>Reverberation of the Normal Modes</i>	338

CHAPTER 14 TRANSDUCTION

14.1	Introduction	344
14.2	The Transducer as an Electrical Network	344
	(a) <i>Reciprocal Transducers</i>	346
	(b) <i>Antireciprocal Transducers</i>	346
14.3	Canonical Equations for Two Simple Transducers	349
	(a) <i>The Electrostatic Transducer (Reciprocal)</i>	349
	(b) <i>The Moving-Coil Transducer (Antireciprocal)</i>	351
14.4	Transmitters	353
	(a) <i>Reciprocal</i>	354
	(b) <i>Antireciprocal</i>	359
14.5	Moving-Coil Loudspeaker	363
*14.6	Loudspeaker Cabinets	369
	(a) <i>The Enclosed Cabinet</i>	370
	(b) <i>The Open Cabinet</i>	371
	(c) <i>Bass-Reflex Cabinet</i>	373
*14.7	Horn Loudspeakers	375
14.8	Receivers	376
	(a) <i>Microphone Directivity</i>	376

	(b) <i>Microphone Sensitivities</i>	376
	(c) <i>Reciprocal Receiver</i>	377
	(d) <i>Antireciprocal Receiver</i>	378
313	14.9 Condenser Microphone	378
314	14.10 Moving-Coil Electrodynanic Microphone	380
317	14.11 Pressure-Gradient Microphones	383
319	*14.12 Other Microphones	386
321	(a) <i>The Carbon Microphone</i>	386
324	(b) <i>The Piezoelectric Microphone</i>	388
325	*14.13 Calibration of Receivers	389
326		
326		
326		
331		

CHAPTER 15 UNDERWATER ACOUSTICS

	15.1 Introduction	396
	15.2 Speed of Sound in Seawater	396
	15.3 Sound Transmission Loss	397
	15.4 Refraction	399
	15.5 Mixed Layer	402
	15.6 Deep Sound Channel	406
	15.7 Surface Reflection	408
	15.8 The Sonar Equations	409
	(a) <i>Passive Sonar</i>	410
	(b) <i>Active Sonar</i>	411
	15.9 Noise and Bandwidth Considerations	412
	(a) <i>Ambient Noise</i>	412
	(b) <i>Self Noise</i>	413
	(c) <i>Doppler Shift</i>	415
	(d) <i>Bandwidth Considerations</i>	416
	15.10 Passive Sonar	417
	(a) <i>The Source: Radiated Noise</i>	417
	(b) <i>An example</i>	418
	15.11 Active Sonar	419
	(a) <i>Target Strength</i>	419
	(b) <i>Reverberation</i>	421
	(c) <i>An Example</i>	426
	*15.12 Isospeed Shallow-Water Channel	427
	(a) <i>Perfectly Rigid Bottom</i>	429
	(b) <i>Slow Bottom</i>	429
	(c) <i>Fast Bottom</i>	430
	*15.13 Solution of the Point Source in Cylindrical Coordinates	433
	*15.14 Isospeed Channel with Rigid Bottom	434
	*15.15 Isospeed Channel with Fluid Bottom	434
	(a) <i>Boundary Conditions</i>	437
	(b) <i>Fast Bottom</i>	

APPENDIX

A1. Conversion factors between MKS and CGS.	446
A2. Complex numbers.	447
A3. Functions of complex argument.	448
(a) <i>Relations to hyperbolic functions.</i>	448
(b) <i>Circular functions in terms of real and imaginary parts.</i>	448
(c) <i>Useful relationships.</i>	448
A4. Bessel functions of integer order.	449
(a) <i>Relationships between solutions.</i>	449
(b) <i>Series expansions for J_0 and J_1.</i>	449
(c) <i>Approximations for small argument, $x < 1$.</i>	449
(d) <i>Approximations for large argument, $x > 2\pi$.</i>	450
(e) <i>Recursion relations.</i>	450
(f) <i>Modified Bessel functions.</i>	450
A5. Tables of Bessel functions, zeros, and inflection points.	451
(a) <i>Bessel functions of order 0, 1, and 2.</i>	451
(b) <i>Zeros of J_m : $J_m(j_{mn}) = 0$.</i>	452
(c) <i>Inflection points of J_m : $(dJ_m/dx)'_{j_{mn}} = 0$.</i>	452
A6. Tables of directivities and impedance functions for a piston.	453
A7. Vector operators.	455
(a) <i>Cartesian coordinates.</i>	455
(b) <i>Cylindrical coordinates.</i>	455
(c) <i>Spherical coordinates.</i>	455
A8. Gauss' law and Green's theorem.	456
A9. A little thermodynamics and the perfect gas.	457
A10. Tables of physical properties of matter.	460
(a) <i>Solids.</i>	461
(b) <i>Liquids.</i>	462
(c) <i>Gases.</i>	463

GLOSSARY OF SYMBOLS

464

ANSWERS TO ODD- NUMBERED PROBLEMS

469

INDEX

474

CHAPTER II

NOISE, SIGNAL DETECTION, HEARING, AND SPEECH

11.1 INTRODUCTION. The *physical characteristics* of speech, music, and noise can be measured with considerable precision by standard acoustic instrumentation including microphones, filters, spectrum analyzers, and oscilloscopes, and the results of such measurements may be expressed in terms of precise *physical* parameters. By contrast, the *interpretative characteristics* of hearing are expressible in terms of *subjective* parameters; they are determined by experiments that lead to statistical predictions of the subjective judgments of an average listener under assumed, or known conditions. For example, in attempting to measure the subjective sensation of *loudness*, one finds that different listeners will agree with surprising accuracy as to the relative loudnesses of two sounds of different frequencies; curves can be obtained relating the *subjective* parameter of loudness to the *physical* parameters of intensity and frequency. Similar investigations have determined typical values for the onset of pain, the *just noticeable differences* (jnd) for frequency and intensity, and so on. Such experiments often suffer from uncertainty as to whether or not all pertinent independent variables, including any mental bias or attitude on the part of the subject toward the experiment, are held constant. Therefore, in reading the material on hearing and in interpreting the curves that are presented, bear in mind that the data were obtained by particular experimenters using particular stimuli, presented to particular listeners under particular conditions. Other experimenters attempting to repeat a given experiment may obtain different results unless great care is taken to duplicate all factors involved in the initial experiment. For a wealth of information beyond that given here, the reader is advised to consult the references given below^{1, 2} and in subsequent sections.

11.2 NOISE, SPECTRUM LEVEL, AND BAND LEVEL. Until now, we have considered only monofrequency acoustic signals (tones)—a serious restriction. Whether listening to a string quartet or trying to ignore a jet aircraft, we are presented with sounds containing a distribution of frequencies. The acoustic intensity for most sounds is nonuniformly distributed over frequency, and it is convenient

¹ Fletcher, *Sound and Hearing in Communication*, Van Nostrand (1953).

² Harris, *Handbook of Noise Control*, McGraw-Hill (1979).

to describe the distribution by the *spectral density*

$$\mathcal{J} = \frac{\Delta I}{\Delta f}$$

where ΔI is the intensity within the frequency interval $\Delta f = 1$ Hz. The total intensity I contained within a band with upper and lower frequencies f_2 and f_1 is

$$I = \int_{f_1}^{f_2} \mathcal{J} df \quad (11.1)$$

The interval $w = f_2 - f_1$ is the bandwidth.

For almost all noise encountered in real situations, the *instantaneous* spectral density $\mathcal{J}(t)$ is a time-varying quantity and $\mathcal{J}$ in (11.1) is a time average over some suitable interval τ , $\mathcal{J} = \langle \mathcal{J}(t) \rangle_\tau$. If $\mathcal{J}$ is constant for each frequency regardless of when the average is performed, the noise is termed *stationary*.

Many conventional acoustic filters and meters have both *fast* (1/8 s) and *slow* (1 s) integration times. The slow integration time is particularly useful in smoothing fluctuations. In general, the fluctuations in I will decrease with increasing integration time and with increasing bandwidth.

The decibel measure of $\mathcal{J}$ is the *intensity spectrum level (ISL)*,

$$ISL = 10 \log \frac{\mathcal{J} \cdot 1 \text{ Hz}}{I_{ref}} \quad (11.2)$$

where for air $I_{ref} = 10^{-12} \text{ W/m}^2$.

Obtaining the *intensity level* $IL = 10 \log I/I_{ref}$ over a desired bandwidth is straightforward if *ISL* is essentially constant over the interval, but becomes involved if it depends strongly on frequency. Let us examine the simpler case first.

If the *ISL* is constant over the bandwidth w , the total intensity calculated from (11.1) has the simple form $I = \mathcal{J}w$ and the *IL* for the band is

$$IL = 10 \log \frac{\mathcal{J} \cdot 1 \text{ Hz}}{I_{ref}} + 10 \log \frac{w}{1 \text{ Hz}}$$

The first term on the right is *ISL*

$$IL = ISL + 10 \log w \quad (11.3)$$

(The 1 Hz divisor of w is suppressed for convenience.)

If the intensity spectrum level *ISL* is a function of frequency, the problem can be handled by subdividing the total bandwidth into small intervals within each of which the *ISL* is relatively constant. The intensity level *IL* in each is calculated from (11.3) and converted to intensity $\hat{I}$. These intensities are summed and then converted to a total intensity level,

$$IL = 10 \log \frac{\sum I_i}{I_{ref}}$$

This calculation will be simplified in the next section.

If the reference intensity and reference pressure are compatible,

$$I_{ref} = P_{ref}^2 / (\rho_0 c^2)$$

then the intensity level IL and the sound pressure level SPL are equal for progressive waves and (11.3) is equivalent to

$$SPL = PSL + 10 \log w \quad (11.4)$$

where PSL , the *pressure spectrum level*, is equivalent to the intensity spectrum level ISL . Recall $P_{ref} = 20 \mu\text{Pa}$ for air. The physical significance of the pressure spectrum level for sounds other than tones may give the reader momentary pause. However, this is no more than the acoustic analog of the total effective voltage generated across a resistance by several series voltage sources, each independent with its own frequency. Each source V_i acting alone dissipates power V_i^2/R in the resistance. The total dissipated power is the sum $\sum V_i^2/R$ and this yields a total effective voltage $V = (\sum V_i^2)^{1/2}$, which appears across the resistance and dissipates the same power.

The sound pressure level SPL and the intensity level IL of a frequency band are each referred to as the *band level*.

Instruments used to analyze noise may be either *constant bandwidth* or *proportional bandwidth* devices.

The constant bandwidth instrument is essentially a tunable narrow-band filter with (constant) bandwidth $w = f_u - f_l$ where f_u and f_l are the upper and lower half-power frequencies. The center frequency of the filter, defined in general as

$$f_c = \sqrt{f_u f_l} \quad (11.5)$$

is usually variable so that the filter can be swept over the desired frequency range. Bandwidths range from a few tens of a hertz to less than a few hundredths of a hertz.

The proportional-bandwidth instrument consists of a series of relatively broad-band filters with upper and lower half-power frequencies satisfying the relationship $f_u/f_l = \text{constant}$. Each bandwidth, being *proportional* to the center frequency, increases with increasing center frequency. Most of these filters have fixed center frequencies with contiguous bands. Common instruments of this type are the *octave-band* filter with $f_u/f_l = 2$, the *1/3-octave-band* filter with $f_u/f_l = 2^{1/3}$, and the *1/10-octave-band* filter with $f_u/f_l = 2^{1/10}$. The *preferred* center frequencies and the corresponding upper and lower frequencies for modern octave-band and 1/3-octave-band filter are shown in Table 11.1. Notice that three adjacent 1/3-octave bands span an octave.

Octave-band and 1/3-octave-band filters are used most often to analyze relatively smoothly varying spectra. If tonals are present, a 1/10-octave or a narrow-band filter should be used.

In theory, the PSL of a noise can be measured by detecting the sound with a calibrated microphone and passing the resulting voltage through a filter of 1-Hz bandwidth tunable over the desired frequency range. In practice, it is conventional

Table 11.1. Preferred octave-band and 1/3-octave-band center frequencies

Center Frequency (Hz)		10 log(Bandwidth)	
Octave	1/3-Octave	Octave	1/3-Octave
16	10	10.5	3.6
	12.5		4.6
	16		5.7
	20		6.6
31.5	25	13.4	7.6
	31.5		8.6
	40		9.7
63	50	16.5	10.6
	63		11.6
	80		12.7
125	100	19.5	13.6
	125		14.6
	160		15.7
250	200	22.5	16.7
	250		17.6
	315		18.6
500	400	25.5	19.7
	500		20.6
	630		21.6
1000	800	28.5	22.7
	1000		23.6
	1250		24.6
2000	1600	31.5	25.7
	2000		26.7
	2500		27.6
4000	3150	34.5	28.6
	4000		29.7
	5000		30.6
8000	6300	37.5	31.6
	8000		32.7

to use a filter of wider bandwidth w , determine the SPL of the band, and then invert (11.4) to obtain a "smoothed" spectrum level

$$\langle PSL \rangle_w = SPL - 10 \log w \quad (11.6)$$

If w is not too great and if the spectrum does not contain strong tones, then $\langle PSL \rangle_w$ will be a relatively good estimate of PSL . If there is significant structure, however, $\langle PSL \rangle_w$ will vary more smoothly with frequency than PSL . If tones are important,

then (11.6) may yield misleading results. A representative spectrum for noise containing both *broadband noise* and *tones* is given in Fig. 11.1. Although the $\langle \text{PSL} \rangle_w$ allows construction of a bar graph from which the overall band level can conveniently be calculated, the important tones are obscured by this means of presentation.

If a broadband noise has a pressure spectrum level that is independent of frequency, the noise is termed *white noise*. This noise is characterized as being rather

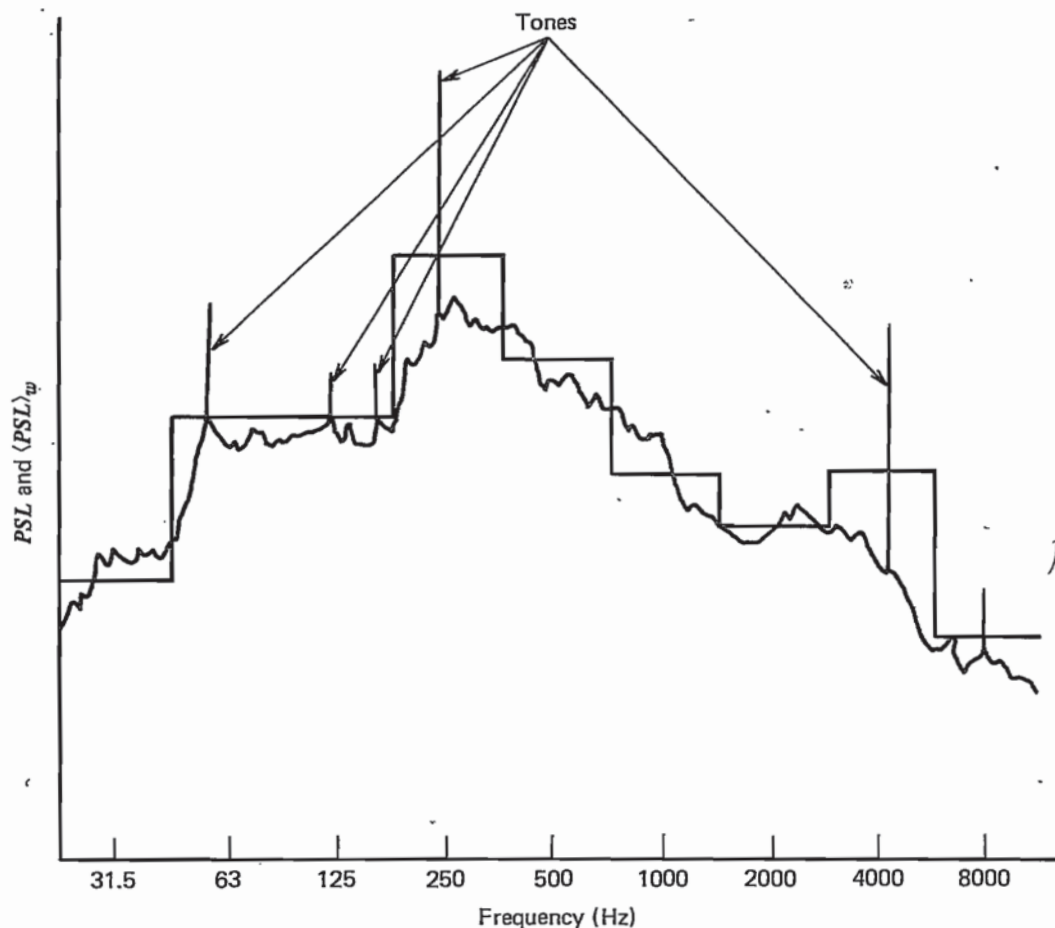


Fig. 11.1. Representative spectrum for noise with the $\langle \text{PSL} \rangle$ determined over octave bands.

sharp and “hissing.” Another type of noise is *pink noise*, for which equal fractions of an octave contain the same power; the spectrum level decreases uniformly with increasing frequency with a slope of -3 dB/octave. This noise is characterized by a “hushing” sound that is less irritating than that of white noise.

11.3 COMBINING BAND LEVELS AND TONES. Calculation of the overall band level can be simplified with the help of a nomogram (Fig. 11.2) that avoids the necessity of taking antilogs and logs as discussed in the previous section. As an example of its use, Table 11.2 shows the calculation of the overall level for a noise

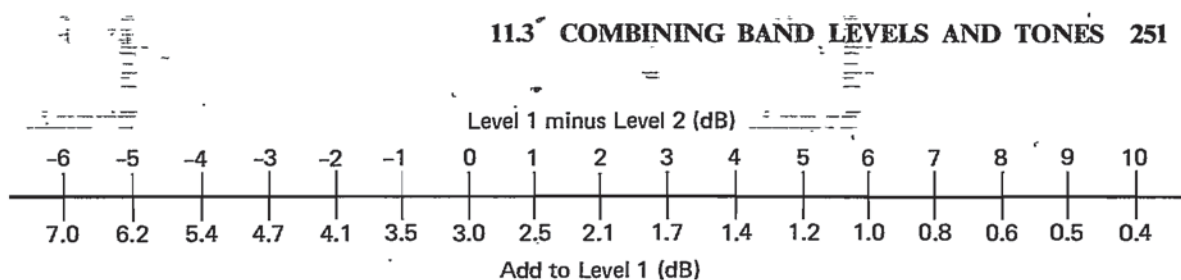


Fig. 11.2. Nomogram for combining levels.

Table 11.2. Calculation of an overall band level from octave-band values

Octave-Band Center Frequency	Octave-Band Level dB re 20 μ Pa	Calculation of Overall Band Level
31.5	70	76.2
63	75	
125	75	79.6
250	73	
500	76	85.1
1000	78	
2000	80	83.7
4000	75	
8000	65	65
		85.1 + ≈ 85 dB re 20 μ Pa

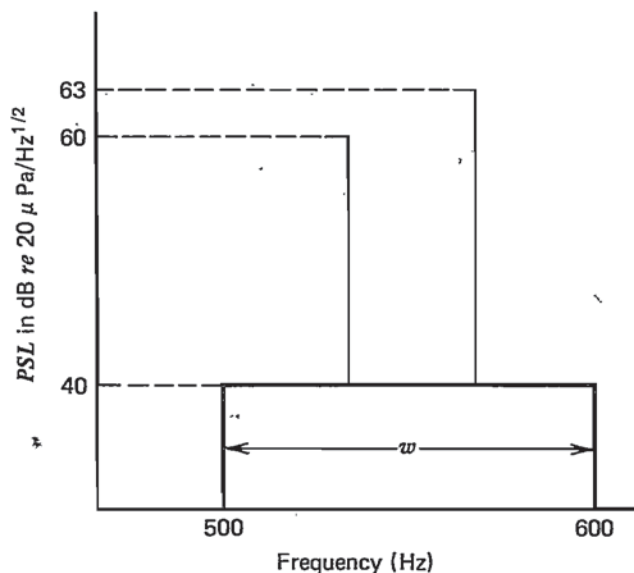


Fig. 11.3. Example of a spectrum containing both noise and tones

analyzed with an octave-band filter. If a more exact calculation is desired, a convenient formula for use on hand-held or desk-top calculators is

$$IL = 10 \log \left(\sum_i 10^{IL_i/10} \right)$$

where IL_i are the individual band levels for the i bands and IL is the overall band level.

Because PSL is defined over a 1-Hz interval, the SPL of a tone is identical with its PSL . If a spectrum contains background noise and tones, the band level of the continuous spectrum is combined with the PSL 's of the tones. For example, consider the spectrum of Fig. 11.3 where, for calculated convenience, the background noise is constant over the bandwidth. From (11.4), the band level of the background alone is 60 dB. When this is combined with the tones of levels 60 dB and 63 dB, the overall band level is 66 dB.

11.4 DETECTING SIGNAL IN NOISE. The detection of a *signal* in the presence of *noise* ultimately reduces to a statistical problem.³ Whether trying to focus on one voice at a noisy party or to locate a submarine underwater, a listener is attempting to isolate desired information, *signal*, from undesired information, *noise*. In humans, the complicated chain of bioacoustical, neurological, and psychoacoustic processes leading from the mechanical stimulation of the eardrum to the perception of sound in the brain constitutes a signal-processing system of stunning sophistication contained within an impressively small volume. It is designed to detect and classify sounds in the presence of noise, and in this role many of its functions and attributes can be well-described quantitatively with the aid of detection theory.

Consider a signal that is *detected*, *filtered* over bandwidth w , *processed* for a time τ , with an *output* presented. This output may be the time average over some selected interval of the rms voltage from a microphone, the subjective loudness of a sample of noise which may contain a tone, the instantaneous voltage developed across an electrical load by a thermosensitive bimetallic strip, or the perceived pitches of two successive samples of frequency-limited noise. Whatever the detection system, the output, from which some decision is made, will be designated by A . If the output A_i of the processor for each time $t_i = t_{i-1} + \tau$ is recorded for a long period of time, the frequency of occurrence of each value can be determined and a histogram constructed giving the probability of any particular value A being observed (Fig. 11.4). Figure 11.4a represents the *probability density function* ρ_N if *noise alone* is present, and Fig. 11.4b the *probability density function* $\rho_{S,N}$ for *signal with noise*. Each has its own mean A_N and $A_{S,N}$ and standard deviation σ_N and $\sigma_{S,N}$. The probability of any value of A_i occurring must be unity, so that

$$\int_0^\infty \rho_N dA = \int_0^\infty \rho_{S,N} dA = 1$$

For any one time interval τ , the only way to decide if a signal is present is to select a *threshold criterion* A_T and assume that if $A_i > A_T$ there is a signal present and if $A_i < A_T$ there is no signal. This yields an independent decision for each time interval, with each decision having a certain probability of being correct or incorrect. From Fig. 11.4c it is seen that the area under

³ For a more complete introduction to detection theory, the reader is referred to Green and Swets, *Signal Detection Theory and Psychophysics*, Wiley (1966).

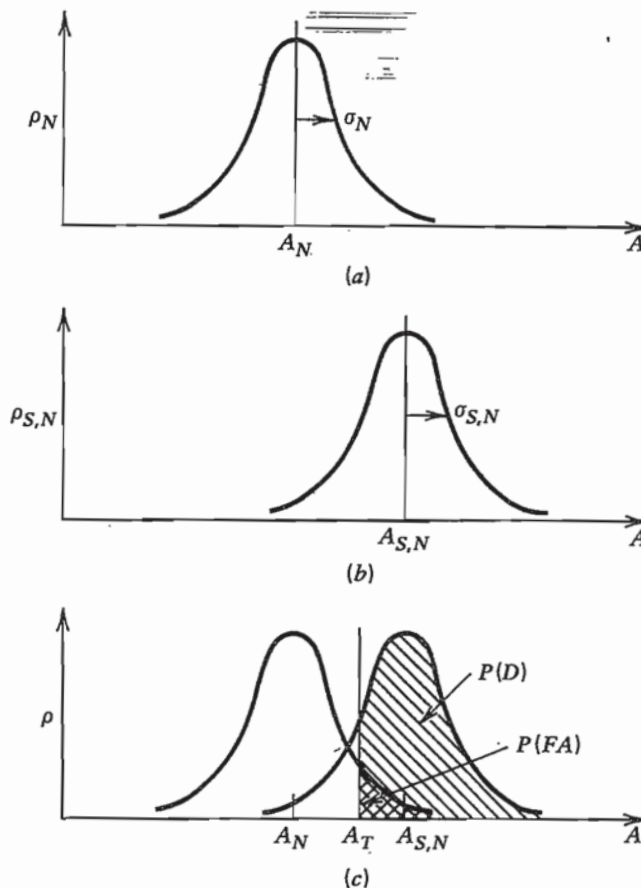


Fig. 11.4. Probability density functions for (a) noise and (b) signal with noise. (c) The probability of detection and the probability of a false alarm for a given threshold criterion.

the signal-with-noise curve to the right of A_T gives the probability of a true detection

$$P(D) = \int_{A_T}^{\infty} \rho_{S,N} dA \quad (11.7a)$$

and the area under the noise curve to the right of A_T gives the probability of a false alarm

$$P(FA) = \int_{A_T}^{\infty} \rho_N dA \quad (11.7b)$$

Thus, the decision as to whether a signal is present during each τ is always statistical in nature, and choosing a particular criterion A_T amounts to finding a suitable balance between the probability of a detection and the probability of a false alarm.

As A_T increases from zero to an arbitrarily large value, $P(D)$ and $P(FA)$ decrease from unity to zero (Fig. 11.5a). Which particular one of the receiving operator characteristic (ROC) curves for a detection system will be followed depends on the separation between the two probability functions. For example, if the value of A_T is held constant and the signal is allowed to increase, the probability density curve for signal with noise will move to the right and $P(D)$ will increase. This shifts the operating point of the system vertically upward to a

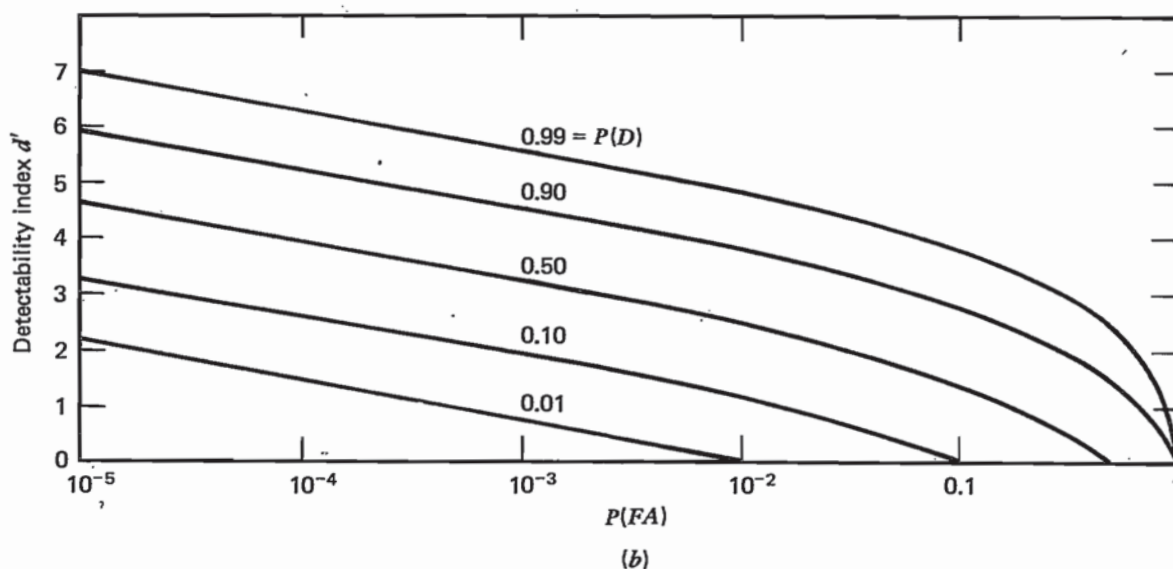
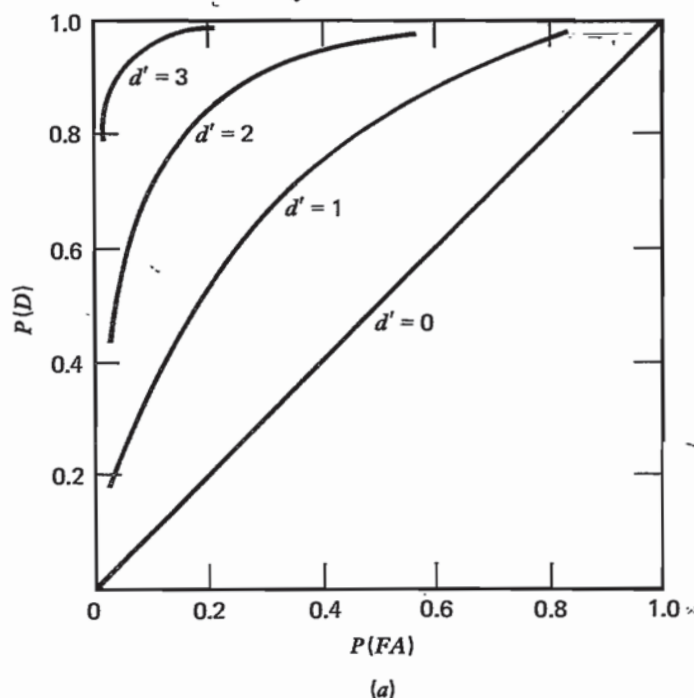


Fig. 11.5. Receiving operator characteristics (ROC) curves for Gaussian distributions with equal standard deviations.

higher curve in Fig. 11.5a. Each individual detection system—whether it be a hungry spider, a submarine detection system with its operator, an egg inspector, a smoke detector, or a piano tuner—has its own set of ROC curves. Each ROC curve of a set is labeled by a detectability index that has larger values for the higher curves.

A very important special case is that for which ρ_N and $\rho_{S,N}$ are both Gaussian with the

same standard deviation $\sigma_{S,N} = \sigma_N = \sigma$. Under these circumstances, the *detectability index* is given by

$$d' = \frac{A_{S,N} - A_N}{\sigma}$$

and measures the number of standard deviations separating the means.

While any particular detection problem may have its own specific probability density functions, for many processes the assumption of Gaussian distributions with equal σ 's is adequate or at least reasonably good. In a real situation, deviation of an experimentally determined ROC curve from the shapes shown in Fig. 11.5a demonstrates that this assumption is not exactly satisfied.

Determining the ROC curves for some detection process is an extremely laborious task. [It takes hundreds of trials to obtain good estimates of $P(D)$ and $P(FA)$ for a single point on an ROC curve and at least several points to determine each of the family of curves.] We summarize three methodologies that lend themselves to analysis by detection theory.

1. **Yes-No Task.** The subject is presented during each trial with either noise or signal with noise. Usually the signal strength is kept constant (d' fixed) so that the results will fall on a single ROC curve. The particular criterion is specified by the instructions to the observer; for example, "say yes only when you are sure the signal is present" sets a higher value of A_T than does "say yes when there is even a hint of the signal". By forcing the subject to change criterion between sets of trials, several points on a single ROC curve can be obtained. The procedure can then be repeated for different signal strengths to obtain curves for different values of d' . The decision process used as the basis of the discussion in this section, and to which the ROC curves of Fig. 11.5a apply, is a yes-no task.

2. **Rating Task.** The subject is presented with either noise or signal with noise and is asked to rate (for example, on a scale from 1 to 4) the chance that the signal is present. This amounts to applying four different values of d' to each task. If the observer remains reasonably consistent, ROC curves should result that are the same as those obtained from the yes-no task.

3. **n -Alternative Forced-Choice (n AFC) Task.** The subject is presented with n samples in each trial, only one of which has signal with noise. The observer must designate which sample is most likely to contain the signal. If the distributions of noise and signal with noise are Gaussian with $\sigma_N = \sigma_{S,N}$, then the ROC curves for 2AFC will be the same as for the yes-no task, but the value of d' for each curve will be $\sqrt{2}$ times that indicated in Fig. 11.5a.

Until the 1950s, there was little recognition throughout the field of psychoacoustics of the importance of both $P(D)$ and $P(FA)$, and the latter was rarely determined. This meant that the criterion used for establishing the threshold was not completely specified, making it difficult to compare the results of different experiments. Subsequent to the 1950s, the relevance and importance of these concepts of detection theory to an understanding of hearing have met with increasing acceptance. Detection theory, by yielding rigorously quantitative specifications of the detection process, has proven its worth.

11.5 DETECTION THRESHOLD. Up to now we have dealt with the detection process at the *output* of the detection system. A most important task is to relate the above considerations to the *input* of the detection system. For example, the ear receives acoustic signals plus acoustic noise. There is then a succession of mechanical, neurological, and psychophysical transmutations of the combined signal and noise that results in a decision as to whether or not the signal is present.

The receiver (ear) has as its input the *signal power* S and the *noise power* N within the bandwidth of the system. We can ask, What is the minimum ratio S/N of (average) signal power to noise power that, after all the processing (mental analysis) of the received stimulus (signal plus noise) is completed, will guarantee a given $P(D)$ for a specified $P(FA)$? The decibel measure of this ratio is the *detection threshold*⁴

$$DT = 10 \log \frac{S}{N} \quad (11.8)$$

Relating this to the properties observed at the output (perception) is nontrivial. The calculations lie beyond the scope of this book, but we can quote two results⁵ for Gaussian probability densities with equal σ 's that correspond to two extremes.

1. If the signal is known exactly, then the optimum processing system is based on *cross correlation*, which searches for a facsimile of the known signal in the received signal with noise. Assume the signal $s(t)$ has duration τ and the noise is given by $n(t)$. The signal with noise is $r(t) = s(t - T') + n(t)$ where T' is some unspecified delay time. (It could be the time of flight of the signal from the source to the input of the detection system.) Cross correlation amounts to generating and studying the function

$$F(T) = \int_{\tau} s(t)r(t + T) dt \quad (11.9a)$$

where T is an adjustable time in the correlator. Now,

$$F(T) = \int_{\tau} s(t)s(t - T' + T) dt + \int_{\tau} s(t)n(t + T) dt \quad (11.9b)$$

Without going into detail, it is clear that the first integral will be maximized when $T = T'$ and the second integral (whose integrand is the product of two independently varying quantities) will, on the average, contribute very little for any value of T . Thus, $F(T)$, which peaks when $T = T'$ and is smaller for other T , plays the role of the detection system output A . If $n(t)$ and $r(t)$ are both Gaussian with equal standard deviations, the detectability index is

$$d' = \sqrt{2w\tau \cdot S/N} \quad (\text{correlation}) \quad (11.10a)$$

It may well be that such correlation plays a role in the ability of the trained listener to "focus" on some particular instrument of an orchestra and follow its performance even during loud *tutti* passages.

2. If the properties of the signal are completely known, *square-law* processing is best. (A detection system receives acoustic pressure, squares it to determine the acoustic energy in each observation interval, and presents that as the output A .) In the process of hearing, there is evidence that this may occur for the detection of a pure tone in noise. In the absence of the transients, associated with musical instruments as one example, or stored information on overtone prominence, the hearing process may fall back on pure energy detection. Under the restrictions $S/N \ll 1$ and $w\tau \gg 1$,

$$d' = \sqrt{w\tau} S/N \quad (\text{square law}) \quad (11.10b)$$

⁴ The detection threshold is often used in the form $DT_1 = 10 \log[S/(N/w)]$ where N/w is the noise power in a 1-Hz band. See, for example, Urlick, *Principles of Underwater Sound*, 2nd ed., McGraw-Hill (1975).

⁵ Peterson, Birdsall, and Fox, *Trans. IRE, PGIT4*, 171 (1954).

The detection thresholds for these two processing schemes are

$$DT = 10 \log[(d')^2/(2w\tau)] \quad (\text{correlation}) \quad (11.11a)$$

$$DT = 5 \log[(d')^2/(w\tau)] \quad (\text{square law}) \quad (11.11b)$$

For example, assume that the system is to detect an unknown signal with $P(D) = 0.5$ and $P(FA) = 0.02$. From Fig. 11.5 we see that $d' = 2$. (1) If the bandwidth of the system is 100 Hz and the processing time is $\tau = 0.5$ s, then with square-law processing $DT = -5.5$ dB and the required ratio of signal power to noise power is 0.28. If the signal were a tone, its *SPL* would be 14 dB *above* the *PSL* of the noise (assuming the noise has constant *PSL* over the bandwidth). (2) If the signal is known exactly, then with correlation processing $DT = -14$ dB and the required $S/N = 0.04$. The *SPL* of the pure tone can be 14 dB *less* than the *PSL* of the noise and still satisfy the required $P(D)$ and $P(FA)$.

In real detection systems (perhaps including some aspects of hearing), there is often a *postdetection filter* which averages the output A over a time interval T_s to reduce output fluctuations. It can be shown that the effect on the detection threshold is to increase it by $|5 \log(\tau/T_s)|$. When such a device is present, the resultant detection threshold DT' of the *combined* system of processor and postdetection filter is then increased above the DT of the processor alone,

$$DT' = DT + |5 \log(\tau/T_s)| \quad (11.11c)$$

Note that this filter can *never decrease* DT , but if $\tau \approx T_s$ it has little deleterious effect.

11.6 THE EAR. The properties of the human ear are phenomenal. The ear is capable of responding over a frequency range from approximately 20 Hz to 20 kHz, and, at 1 kHz, sounds that displace the eardrum only one-tenth of the diameter of the hydrogen molecule can be detected. However, it is much more than a sensitive, broadband receiver. In conjunction with the nervous system, it acts as a frequency analyzer of enviable selectivity. In this book we can give only a limited introduction to the ear and hearing. For more information, consult the references.^{1,6,7}

The human ear (Fig. 11.6) is one of the most intricate and delicate mechanical structures in the human body. It consists of three main parts: the *outer*, *middle*, and *inner* ears.

The *pinna* of the outer ear serves as a horn collecting sound into the *auditory canal*. In a human the pinna is a relatively ineffective device, but in some animals it supplies an appreciable gain over certain frequency ranges. The auditory canal is an approximately straight tube, about 0.7 cm in diameter and 2.5 cm long, closed at its inner end by the *tympanic membrane* (eardrum). At the lowest resonance of this tube (about 3 kHz) the *SPL* at the drum is about 10 dB higher than at the entrance of the canal. Since the resonance curve is quite broad, appreciable gains are observed over a frequency range extending from about 2 to 6 kHz. If the effects of the diffraction of sound waves by the head are also taken into account, the *SPL* at the eardrum exceeds the free-field *SPL* by 15 or 20 dB at some frequencies.

The eardrum forms a flattened cone lying obliquely across the auditory canal with its apex facing inward. It is quite flexible in the center and attached around its

⁶ Gelard, *The Human Senses*, 2nd ed., Wiley (1972).

⁷ Jerger, *Modern Developments in Audiology*, 2nd ed., Academic Press (1973).
Gelfand, *Hearing*, Dekker (1981).

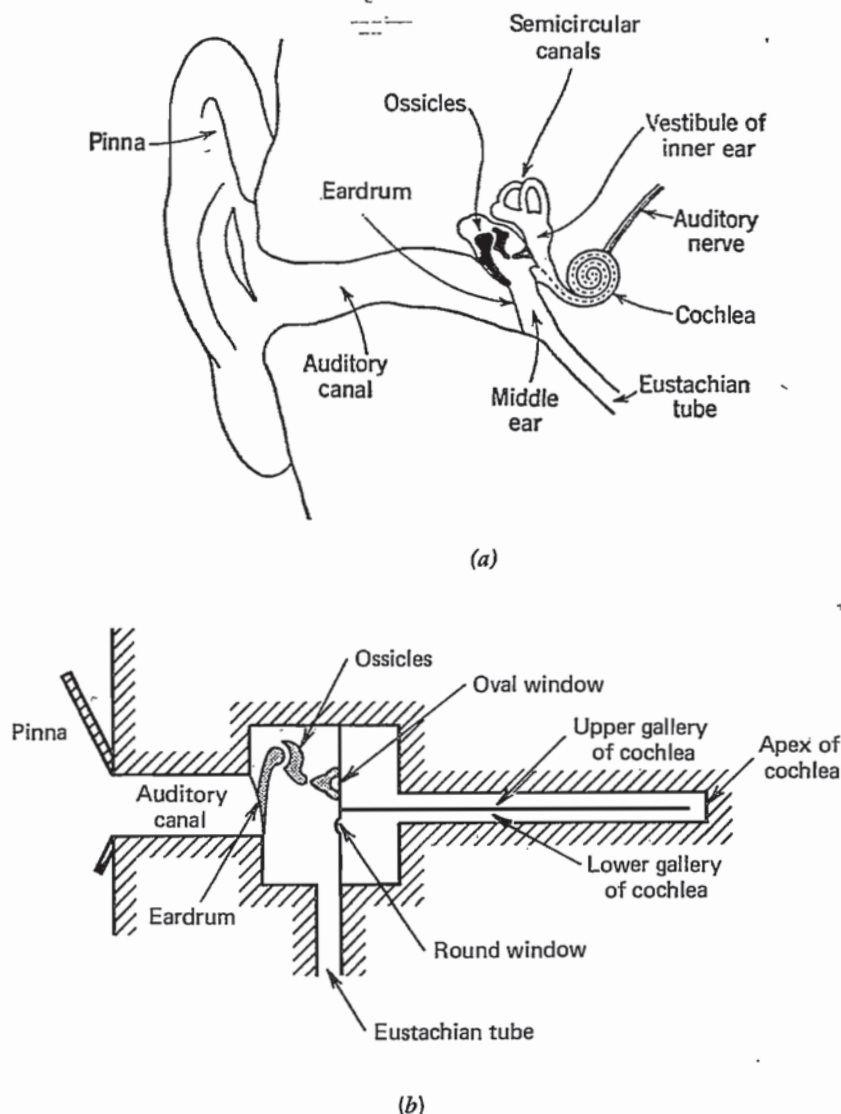


Fig. 11.6. Sketch of the ear.

edges to the end of the canal. This membrane is the entrance to the middle ear, an air-filled cavity of about 2 cm^3 volume which contains three *ossicles* (bones): the *malleus* (hammer), *incus* (anvil), and *stapes* (stirrup), together with their supporting muscles and ligaments. This cavity is connected to the throat through the *Eustachian tube* (which is normally closed, but opens during swallowing or yawning to equalize the pressure on each side of the eardrum). Vibrations of the eardrum are transmitted to the malleus, through the incus, and then to the stapes, which covers the *oval window*.

This linkage of bones, in combination with the area ratio of about 30 to 1 between the eardrum and the oval window, supplies an approximate impedance match between the air in the auditory canal and the liquid in the inner ear. This impedance match varies with the intensity of the received sound. For high intensities, the muscles controlling the motion of the ossicles change their tension to

reduce the amplitude of motion of the stapes, thereby protecting the delicate mechanism of the inner ear from damage. This process is known as the *acoustic reflex*. Since it takes about 0.5 ms after perceiving a loud sound for the acoustic reflex to become effective, it offers no protection from sudden impulsive sounds such as gunshots, explosions, and so forth.

The inner ear (*labyrinth*) has three parts: the *vestibule* (entrance chamber), the *semicircular canals*, and the *cochlea*. The vestibule connects with the middle ear through two openings, the *oval window* and the *round window*. Both of these windows are sealed to prevent the escape of the liquid filling the inner ear; the former by the stapes and its support, and the latter by a thin membrane. With these two exceptions, the entire inner ear is surrounded by bone. (The semicircular canals play no part in the process of hearing; they provide us with a sense of balance.) The cochlea is a tube of roughly circular cross section, wound in the shape of a snail shell. This tube makes about 2.5 turns and has a total length of about 3.5 cm. Its cross-sectional area decreases in a somewhat irregular manner from its base to its apex. Its volume is about 0.05 cm^3 .

The tube of the cochlea is divided by the *cochlea partition* into two longitudinal channels: the *upper gallery* (*scala vestibuli*) and the *lower gallery* (*scala tympani*). The only communication between the two galleries is through the *helicotrema*, a small opening at the apex of the cochlea. The other ends of the upper and lower galleries connect with the oval and round windows, respectively. A cross section of one of the turns of the cochlea is shown in Fig. 11.7. The *bony ledge* projects from the central portion of the shell-like structure into the liquid-filled tube and carries the *auditory nerve*. At the termination of the bony ledge the nerve fibers enter the *basilar membrane*, which continues across the tube to the farther side where it is attached to the *spiral ligament*. Lying above the basilar membrane is the *tectorial membrane*, attached along one edge to the bony ledge with its opposite edge projecting into the cochlear liquid. Running diagonally across the cochlear canal from the bony ledge to the opposite wall is *Reissner's membrane*, just two cells thick. This membrane, with the basilar membrane, isolates the majority of the cochlea (the upper and lower galleries) from the pie-shaped cross section named the *cochlea duct*. This duct is filled with *endolymphatic fluid*; the galleries are filled with *perilympatic fluid*. (The endolymph is closely related to the intercellular fluid found throughout the body, and the perilymph is similar to spinal fluid.)

Attached to the top of the basilar membrane is the *Organ of Corti* that contains four rows of *hair cells* (about 3×10^4 cells in all) spanning the entire length of the cochlea. The 3500 hair cells in the inner row are less vulnerable to damage than are the cells in the outer three rows. Several dozen small hairs extend from each hair cell to the under surface of the tectorial membrane.

When the ear is exposed to a pure tone, the motion of the eardrum is transmitted by the bones of the middle ear to the oval window, creating a fluid disturbance that travels down the upper gallery toward the apex of the cochlea, through the helicotrema, into the lower gallery, and back to the round window, which acts as a pressure-release termination. The detailed properties of this disturbance and its role in the mechanism of hearing were elucidated in a Nobel-Prize winning series of

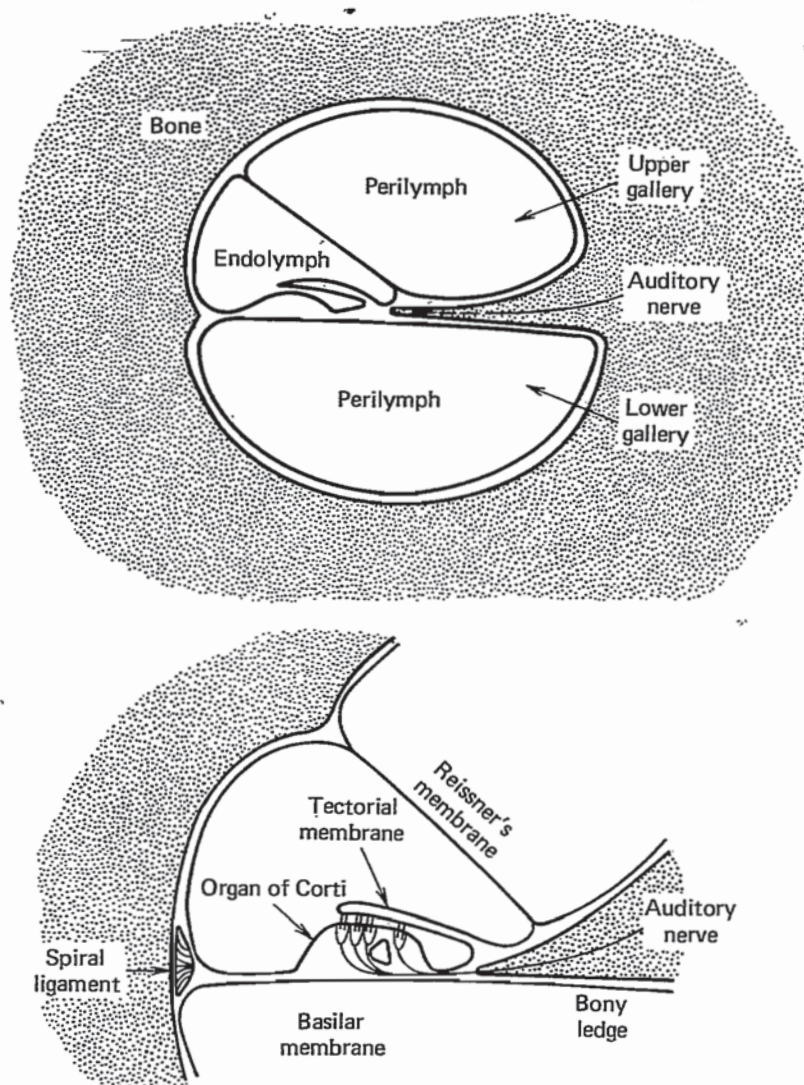


Fig. 11.7. Cross section of the cochlea duct.

investigations by Bekesy. His experiments demonstrated that the basilar membrane is driven into highly damped motion with a peak amplitude that increases slowly with distance away from the stapes, reaches a maximum, and then diminishes rapidly toward the apex. The peak amplitude maximizes closer to the apex for lower frequencies. A few examples are sketched in Fig. 11.8 where z is the distance in centimeters from the stapes. The relation of the position of the maximum peak amplitude to the frequency can be well-approximated by

$$f = 2.5 \times 10^4 - 0.72z \quad (11.12)$$

for frequencies above about 200 Hz.¹ For lower frequencies, the distance from the apex is underestimated by (11.12). These motions of the basilar membrane occur whether the mechanical excitation is produced by airborne sounds (through the eardrum, ossicles, etc.) or by sounds conducted through the skull.

Since the Organ of Corti is attached to the basilar membrane while the tectorial

membrane is attached to the bony ledge, the relative motions generated between them flex the hairs, thereby exciting the nerve endings attached to the hair cells into producing electrical impulses. These nerves do not necessarily fire at the frequency with which they are excited, but quasirandomly when they are stressed beyond certain limits, usually firing more often when highly stressed. These pulses form the information communicated from the cochlea to the brain. From each ear, information routes go to a number of interlinked processing centers within the brain.

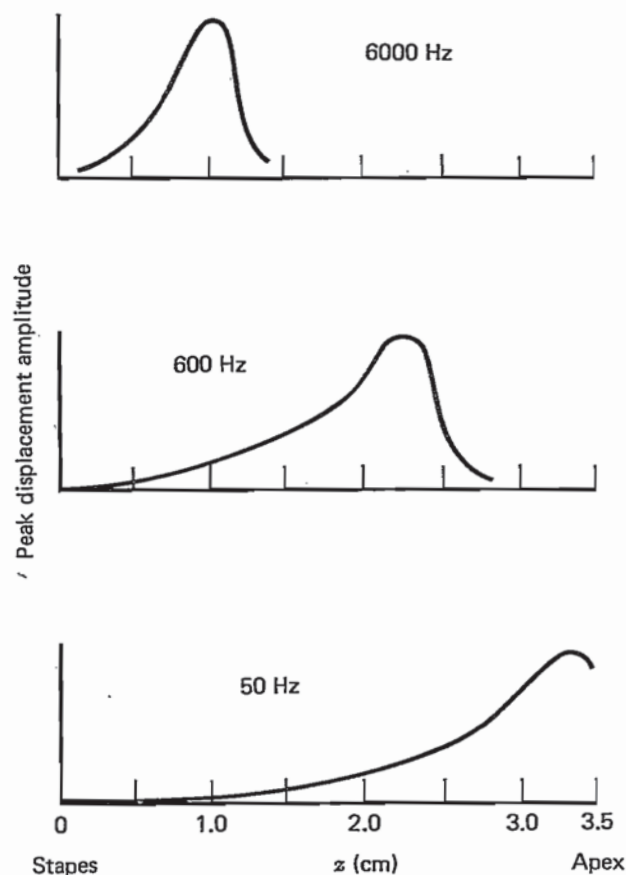


Fig. 11.8. Peak displacement amplitude of the basilar membrane for a pure tone.

There is considerable communication between processors on different sides of the brain so that the signals from both ears are processed and mixed. There are, in addition, sets of nerves that send information from the brain to the ears. (For example, the left ear transmits information to the left *superior olivary complex* which in turn has nerves going directly to the right ear.) This highly complicated, subtle, and sophisticated communication and processing network is just beginning to be understood. For further information, the reference below⁸ may be a suitable start.

⁸ Carteratte and Friedman, *Handbook of Perception*, Vol. 2, Academic Press (1974).

11.7 SOME FUNDAMENTAL PROPERTIES OF THE HUMAN EAR.

(a) **Thresholds.** The threshold of audibility is the minimum perceptible free-field intensity level of a tone that can be detected at each frequency over the entire range of the ear. The symbol L_I is recommended by the International Standardization Organization (ISO) for IL re 10^{-12} W/m². The tone presented to both ears should have a duration of about 1 s. For tones shorter than 0.3 s the apparent loudness increases with increasing tone duration, and for tones longer than 3 s "fatigue" sets in and the apparent loudness diminishes with time.

The average threshold of audibility for the normal (undamaged) ear is shown as the lowest curve on Fig. 11.9. The frequency of maximum sensitivity is near 4 kHz. Below this, the threshold rises with decreasing frequency, the minimum power required to produce an audible sound at 30 Hz being nearly a million times as great

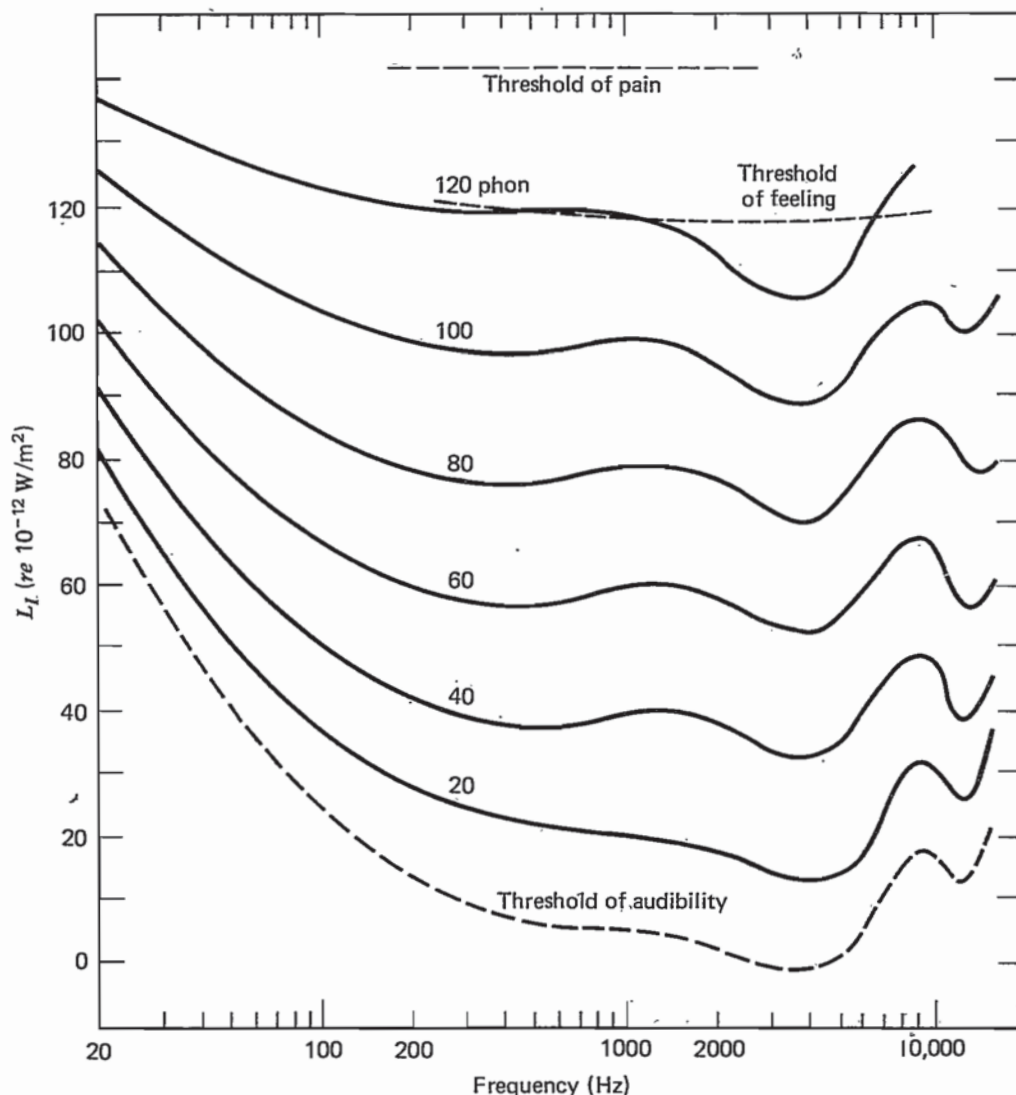


Fig. 11.9. Thresholds and free-field, equal-loudness-level contours⁹ for pure tones with subject facing the source.

as at 4 kHz. For high frequencies, the threshold rises rapidly to a cutoff. It is in this frequency region that the greatest variability is observed between different listeners, particularly if they are over 30 years of age. The cutoff frequency for a young person may be as high as 20 kHz or even 25 kHz, but people who are 40 or 50 years of age can seldom hear frequencies in excess of 15 kHz, and in some the cutoff is below 10 kHz. In the range below 1 kHz, the threshold is usually independent of the age of the listener.

As the intensity of the incident acoustic wave is increased, the sound grows louder and eventually produces a tickling sensation. This level, less dependent on frequency than the threshold of audibility, has a value of approximately 120 dB and is called the *threshold of feeling*. As with the lower threshold, it varies somewhat from individual to individual, but not to so great an extent. As the intensity is still further increased, the tickling sensation gives way to one of *pain* at about 140 dB. It must be emphasized, however, that even for short exposures permanent damage to the ear can occur at levels below 100 dB.

Since the ear responds to loud sounds by reducing the sensitivity (the acoustic reflex mentioned earlier), the threshold of audibility shifts upward under exposure, the amount of shift depending on the intensity and duration of the sound. After the sound is removed, the threshold of hearing will begin to reduce and, if the ear fully recovers its original threshold, we have observed *temporary threshold shift TTS*. (The communication between right and left ears evinces itself here in that strong stimulation of one ear results in a small threshold shift in the other ear.) The amount of time required for a complete recovery increases with increasing intensity and duration of the sound. If the exposure is long enough or the intensity high enough, the recovery of the ear is not complete, and the threshold never returns to its original value; *permanent threshold shift PTS* has occurred. It is important to realize that the damage leading to *PTS* occurs in the inner ear; the hair cells are damaged with no possibility of recovery or reconstruction.

Also of importance are *differential thresholds*, one of which is the differential threshold for *intensity determination*. If two tones of almost identical frequency are sounded together, one tone much weaker than the other, the resultant signal is indistinguishable from a single frequency whose amplitude fluctuates slightly and sinusoidally (the "beat" phenomenon discussed in Sect. 1.13). The amount of fluctuation that the ear can just barely detect, when converted into the difference in intensity between the stronger and weaker portions, determines the differential threshold. As might be expected, values depend on frequency, number of beats per second, and intensity level. Generally, the greatest sensitivity to intensity changes is found for a beat frequency of about 3 beats per second. Sensitivity decreases at the frequency extremes, particularly for low frequencies, but this effect diminishes with increasing sound level. For sounds more than about 40 dB above threshold, the ear is sensitive to intensity level fluctuations of less than 2 dB at the frequency extremes; and less than about 1 dB between 100 and 1000 Hz.

Other differential thresholds involve the ability to discriminate between two sequential signals of nearly the same frequency. The frequency difference required to make the discrimination is termed the *difference limen*. Older methods of measuring

the difference limen consist of exposing the ear to a frequency-modulated tone where the amount and rate of modulation can be controlled. The difference limen determined by this method depends on the intensity of the signal, its center frequency, and its rate of modulation. Greatest sensitivity seems to occur when the modulation rate is about 3 Hz. At moderate levels, the difference limen is about 1 Hz for tones with frequencies up to about 100 Hz, rising smoothly to 5 Hz for 1 kHz tones, and up to 70 Hz for 10 kHz tones. Thus, for this method, at the midrange and higher frequencies, the ear can detect changes of frequency of about 0.7 percent. More recent experiments use carefully shaped tone bursts of each frequency and yield difference limens up to about five-times smaller than those measured by the modulation method.

(b) Equal Loudness Level Contours.⁹ Experiments can be performed to gauge when two tones of dissimilar frequencies, sounded alternately, appear to be equally loud. As seen in Fig. 11.9, high- and low-frequency tones require greater values of L_I to sound as loud as those in the midfrequency range. The curves resulting from such comparisons are labeled by the L_I they have at 1 kHz. Each curve is called an *equal loudness level contour* and expresses the *loudness level* L_N in *phon* which is assigned to all tones whose L_I fall on the contour. Therefore, $L_N = L_I$ for a 1-kHz tone, regardless of its level. On the other hand, a pure tone of 40 Hz with $L_I = 90$ dB (*re* 10^{-12} W/m²) has a loudness level $L_N = 70$ phon, as does a 4-kHz tone with $L_I = 61$ dB. Notice, however, that for higher loudness levels the curves become straighter, so that L_N and L_I become more nearly equal at all frequencies.

The flattening of the equal loudness level contours for higher loudness levels explains the well-known fact that turning up the "volume" on a hi-fi system causes the frequency extremes to appear louder—there are more bass and treble. Conversely, turning the volume down removes both "body" and "brilliance," and the sound takes on a "thin" and "tinny" quality. This also reveals a problem in selecting loudspeakers of roughly equivalent quality; the one which is more efficient, and therefore louder than the other for equal signal input, may well sound more impressive merely because it is louder.

(c) Critical Bandwidth. If a subject listens to a sample of noise with a tone present, the tone cannot be detected until its L_I exceeds a value that depends on the amount of noise present. In a pivotal set of experiments, Fletcher and Munson¹⁰ found that the masking of a tone by a broadband noise is independent of the noise bandwidth until the bandwidth became smaller than some critical value that depends on the frequency of the tone. The ear acts like a collection of *parallel filters*, each with its own bandwidth, and the detection of a tone requires that its level exceed the noise level in its particular band by some detection threshold.

In early experiments it was assumed that the signal must *equal* the noise for detection to occur ($DT = 0$). On this basis, and assuming that the sensitivity of the

⁹ Robinson and Dadson, *Brit. J. of Appl. Physics*, 7, 166 (1956) and International Standardization Organization ISO R226-1961.

¹⁰ Fletcher and Munson, *J. Acous. Soc. Am.*, 9, 1 (1937).

ear is constant across each bandwidth w_{cr} , the critical value of the bandwidth could be found from $w_{cr} = S/N_1$, where S is the signal power and N_1 the noise power per hertz. For reasons that will be clear in a moment, the critical values measured this way are now termed the *critical ratios* (see Fig. 11.10).

Later experiments based on the perceived loudness of noise have yielded *critical bandwidths* w_{cb} larger than the critical ratios. In some of these experiments, the loudness of a band of noise is observed as a function of bandwidth while the *overall noise level is held constant*. For noise bandwidths less than critical, the loudness will

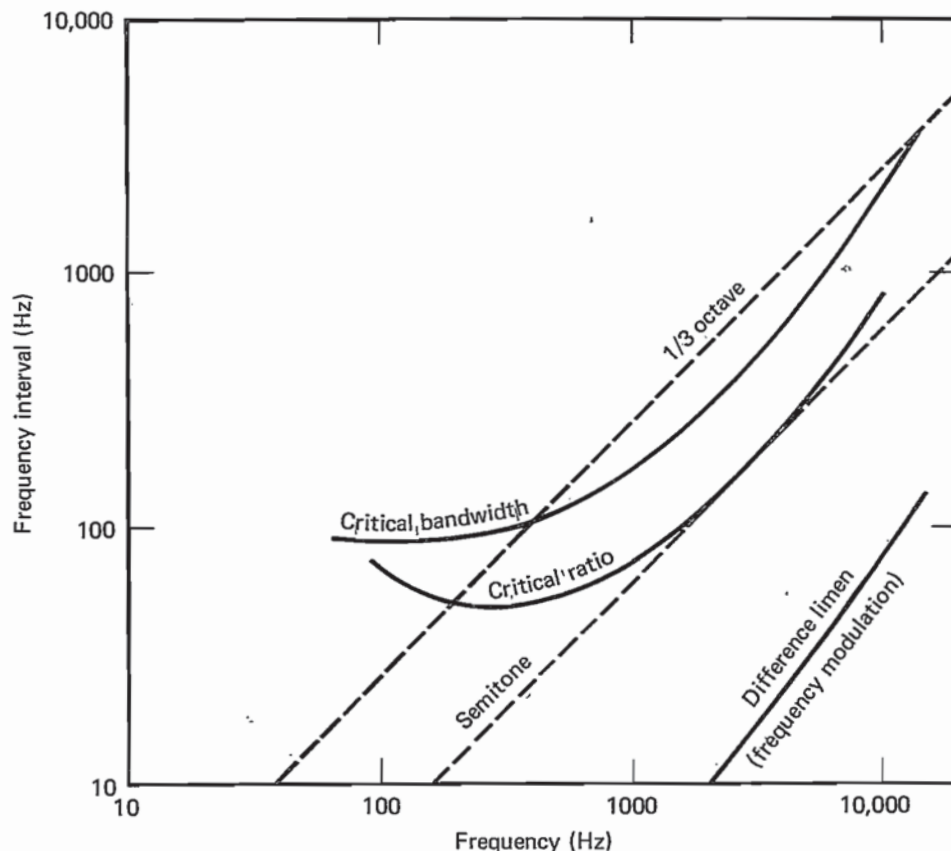


Fig. 11.10. Critical bandwidths of the ear.

be constant; when the bandwidth exceeds the critical bandwidth, the loudness will increase (see Sect. 11.8).

A representative curve¹¹ of this critical bandwidth is shown in Fig. 11.10. It is interesting to note that the critical bandwidth is nearly 1/3-octave for frequencies above about 400 Hz. Over this range, the critical bandwidth is about 2.5 times the critical ratio; one cause of the discrepancy¹² may be that the strong amplitude modulations of narrow-band noise used in the critical-ratio experiments allowed the tone to “peek through” and be detected during relatively quiet periods. It should

¹¹ Zwicker, Flottop, and Stevens, *J. Acous. Soc. Am.*, **29**, 548 (1957).

¹² de Boer, *J. Acous. Soc. Am.*, **34**, 985 (1962).

also be noted that the earlier experiments gave no specification of $P(D)$ and $P(FA)$ and so were incompletely documented.

A complete curve of the difference limen obtained by frequency modulation would show that the ear can perceive frequency changes of about $1/30$ of the critical bandwidth. The mechanism for this frequency selectivity within intervals of a critical bandwidth cannot be explained on the basis of the simple filtering action mentioned above. One suggestion is that the signal processing occurring within the brain culminates in a "sharpening" of the nerve excitations transmitted to the brain so that there is an effective narrowing of the region of response and hence an enhancement of frequency discrimination. Another suggestion, supported by experimental evidence, is that the ear may recover frequency information from the long-term combined rate at which the nerves fire. Even though each nerve usually fires with a repetition rate far below the detected frequency, if there are many nerves firing at lower rates, the collection of all the nerve firings will yield "volleys" generated with a repetition rate equal to the frequency.

(d) Masking. Everyone is familiar with the difficulty of hearing and understanding speech in the presence of noise. The amount of *masking* is the amount the threshold of audibility of the signal is raised in the presence of noise.

It is simplest to begin by considering the masking of one pure tone by another. The subject is exposed to a single tone of fixed frequency and L_I and then asked to detect another tone of different frequency and level. Analysis yields the *threshold shift*: the increase in L_I of the masked tone above its value for the threshold of audibility before it can be detected. Figure 11.11 gives representative results for masking frequencies of 400 and 2000 Hz. The frequency range over which there is appreciable masking increases with the L_I of the masker, the increase being greater for frequencies above that of the masker. This is to be expected because the region of the basilar membrane, excited into appreciable motion at moderate values of L_I ,

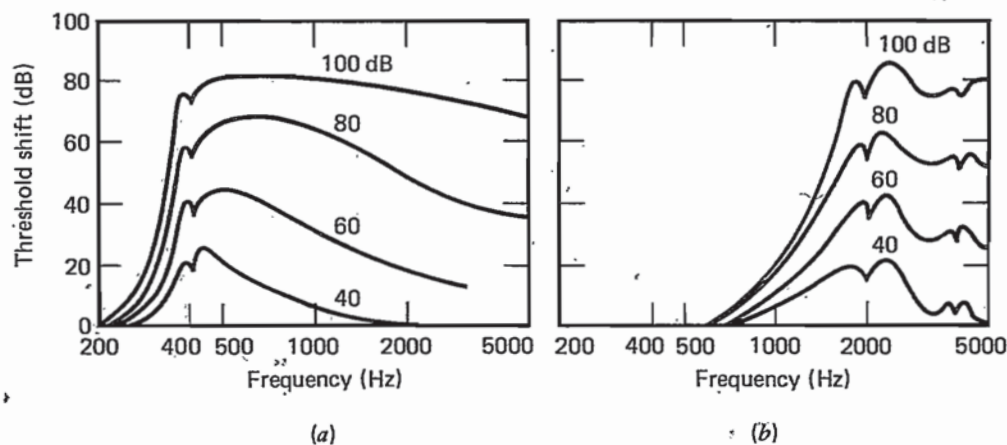


Fig. 11.11. Masking of one pure tone by another. The abscissa is the frequency of the masked tone, and the curves are labeled with the L_I of the masking tone. The frequency of the masking tone is 400 Hz in (a) and 2000 Hz in (b).

extends from the maximum further toward the stapes than the apex. For stronger excitation of the membrane both regions grow, the region toward the stapes more significantly; it is this region that covers the frequencies higher than that of the masker: low frequencies mask high frequencies. (The notches in the curves will be explained in the discussion that follows.)

Masking of pure tones by a band of noise narrower than w_{cb} is essentially the same as that of an equally intense pure tone having the same frequency as that at the center of the band. Consequently, when the spectrum level is relatively constant, the intensity of a narrow band of noise is directly proportional to the bandwidth w and, correspondingly, the masking expressed in decibels increases as $10 \log w$. Ultimately, the bandwidth will equal the critical bandwidth, beyond which any further increase in the width of the pass band of the noise has little influence on the amount of masking of a pure tone at the center of the band.

(e) Cochlear Nonlinear Effects: Beats, Combination Tones, and Aural Harmonics. Let two tones of similar frequency f_1 and f_2 and of equal L_I be presented to one (or both) ear(s). When the two frequencies are very close together, the ear perceives a tone of single frequency $f_c = (f_1 + f_2)/2$ fluctuating in intensity at the *beat frequency* $f_B = |f_1 - f_2|$ (see Sect. 1.13). As the frequency interval between the two tones increases, the sensation of *beating* changes to *throbbing* and then to *roughness*. As the frequency interval increases further, the roughness gradually diminishes and the sound becomes smoother, finally resolving into two separate tones. For frequencies falling in the midrange of hearing, the transition from beats to throbbing occurs at about 5 to 10 beats per second and this turns into roughness at about 15 to 30 beats per second. These transitions occur for higher beat frequencies as the frequencies of the primary tones are increased. Transition to separate tones occurs when the frequency interval has increased to about the critical bandwidth. None of this occurs if each of the tones is presented to a different ear. When each ear is exposed to a separate tone, the combined sound perceived by the brain is the smooth sounding of two independent frequencies; *there are no beats*. This suggests that the beats arise because the two tones generate overlapping regions of excitation on the basilar membrane, and it is not until these regions become separated by the distance corresponding to the critical bandwidth that they can be separately sensed by the ear. (When the tones are presented one in each ear, of course, each basilar membrane is separately excited, and there is no possibility of mechanical interaction.)

If the two tones (both presented together in one or both ears) are separated far enough and are of sufficient loudness, a trained observer can detect *combination tones*. These combination tones are not present in the original sound, but are *manufactured by the ear*. There is a collection of possible combination tones whose frequencies are various sums and differences of the original frequencies f_1 and f_2 ,

$$f(\text{combination}) = |mf_2 \pm nf_1| \quad n, m = 1, 2, 3, \dots \quad (11.13)$$

Only a few of these frequencies will be sensed. One of the easiest to detect is the *difference frequency* $|f_2 - f_1|$.

A *linear* combination of two frequencies does not result in any combination tones. On the other hand, if two frequencies are supplied to a *nonlinear* system, the response will contain not only the original frequencies, but also various sum and difference frequencies. Nonlinear electrical circuits are common in radio receivers and transmitters for generating such combination tones. The cochlea is the origin of the nonlinearity creating combination tones. (The middle ear remains remarkably linear up to the threshold of pain.) One means of studying this nonlinearity is through measurement of *cochlear potentials*. When the basilar membrane is set into motion and the hair cells stressed, slight potential differences are generated that can be detected by means of an electrical probe inserted into the cochlea. The amplitudes and waveforms of the cochlear potentials appear to represent quite accurately those of the sounds heard. For two pure tones of low intensity, the cochlear potential has a waveform identical with that of the received sound. As the incident intensity is increased, however, the wave becomes distorted, indicating the generation of combination tones. In general, difference tones appear at somewhat lower intensity levels than do summation tones, but both are observed.

The fact that there are nonlinear mechanisms operating in the cochlea leads to the formation of various combination tones when an ear is exposed to two pure tones. Furthermore, there is nonlinear distortion when only one tone is received. Indeed, if m is set equal to zero in (11.13), which is equivalent to removing the tone of frequency f_2 , there are still nonlinearly generated tones with frequencies nf_1 . These are called *aural harmonics*. Tones having frequencies above 500 Hz and loudness levels below 40 phon do not generate aural harmonics of appreciable magnitude. For frequencies around 100 Hz, the loudness level at which distortion first appears is about 20 phon. With increasing loudness the aural harmonics appear in order; initially the second harmonic appears, then the third, and so on. In general, the loudnesses of the aural harmonics are less than that of the fundamental, and they decrease progressively with increasing order. (However, in an intense, low-pitched sound, such as a 60-Hz tone at an intensity of 100 dB, the second harmonic may be louder than the fundamental and a number of the higher harmonics may also approach the fundamental in loudness.) The strongest evidence supporting their existence is from cochlear potentials. Another method introduces a second tone whose frequency is brought close to a frequency of an aural harmonic so that beats are generated between the second tone and the harmonic. (This is less convincing because the vibratory motion of the basilar membrane is perturbed by the introduction of the second tone.) This second method, however, demonstrates the effectiveness of beats in enhancing detection. The notches in the masking curves of Fig. 11.11 show that these beats lower the threshold at the frequencies of the aural harmonics.

(f) Nonlinear Processing Effects: Consonance and the Restored Fundamental. In addition to the effects described above, there are other nonlinear effects that, while somewhat more subtle, have great importance for music.

One effect involves the perception of mistuned intervals. If two tones *almost* an octave apart are sounded simultaneously and tuned toward the octave, there will be a strong impression of beating that slows and dies away as the interval of the octave

is approached. This beating is not like that for the amplitude modulation previously discussed, but is apparently a response to the slowly time-varying nature of the waveform (see Fig. 11.12). An important feature of this beating is that *it is not necessary that both tones be presented in the same ear*. If one tone is presented to one ear and the second tone to the other ear, the beating is still sensed. Thus, this interference is not the result of nonlinearities of the cochlea, but arises out of

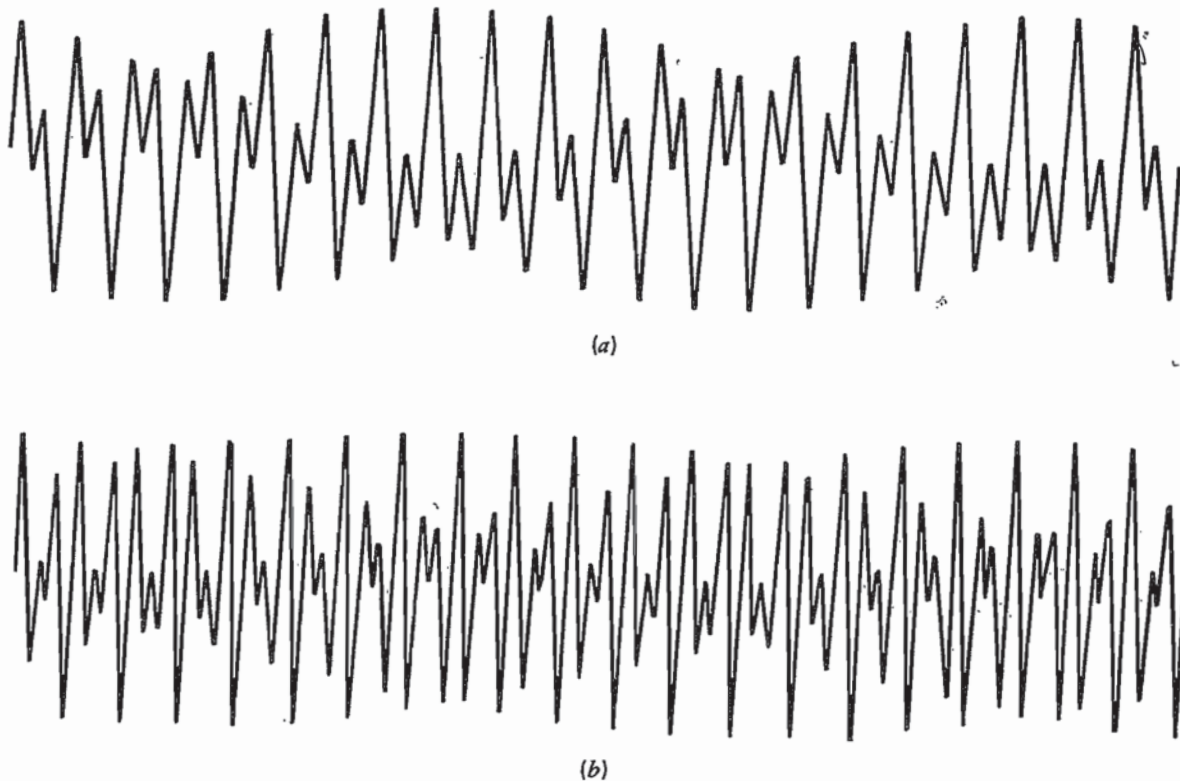


Fig. 11.12. Waveforms for mistuned consonances. (a) Mistuned octave. (b) Mistuned fifth. (The mistuning has been greatly exaggerated for visual clarity.)

nonlinearities in the processing of the information in the relevant centers of the brain. Besides the octave, these beats are also observed when the ratio of the two frequencies is nearly that of two integer numbers: $f_2/f_1 = 2/1, 3/2, 4/3, \dots$, with the beating becoming more subtle as the integers get greater. When the frequencies are exactly aligned to these ratios, the beating vanishes and the sensation is of *consonance*. It is no accident that these ratios form the foundation of our musical scales.^{13,14}

A second effect is the generation of the fundamental when two or more consonances are presented. For example, if tones of 800 and 1000 Hz are presented to the

¹³ Roederer, *Introduction to the Physics and Psychophysics of Music*, Springer-Verlag (1973).

¹⁴ Strong and Plitnik, *Music, Speech, and High Fidelity*, Brigham Young University Publications (1977).

ears, a frequency of 200 Hz can be detected. This tone is the fundamental of which the other two are adjacent harmonics. Unlike the difference tone discussed earlier which can be masked by noise, this fundamental can be perceived even when it should be masked and therefore undetectable. Furthermore, it is perceived when the signals are too weak to generate a detectable difference frequency. Finally, as in the first effect, it occurs even if the tones are fed into different ears.

The "generation of the missing fundamental" has practical application in the design of inexpensive, small radios. To eliminate the cost of expensive networks to filter out the 60-Hz line frequency and the 120-Hz harmonic generated in the rectifier, manufacturers deliberately limit the low-frequency response of such radios to eliminate frequencies below about 150 Hz. (Higher harmonics of the line frequency are filtered out quite inexpensively.) Thus, while there is no significant output below about 150 Hz, the nonlinear processing in the brain restores the fundamentals of the bass notes from the higher harmonics which are still present.

11.8 LOUDNESS LEVEL AND LOUDNESS. Although two sounds having the same loudness level are judged to be *equally loud*, this does not mean that the subjective sensation called *loudness* N is proportional to *loudness level* L_N . A tone of $L_N = 60$ phon will not sound twice as loud as one of 30 phon. The unit of loudness is the *sone*, and a loudness $N = 1$ sone is equal by definition to a loudness level of 40 phon, independent of frequency. A loudness of 16 sone is twice as loud as one of 8 sone and four times as loud as one of 4 sone.

Loudness is not easy to measure, and its determination requires ingenuity. The results of a great many experiments at different frequencies have been summarized by Fletcher,¹ and a graph representing the relationship between loudness N and loudness level L_N is given in Fig. 11.13. While the lower portion of the graph is noticeably curved, the portion in excess of 1 sone is relatively straight. In the linear portion (corresponding to sounds from comfortably audible to unpleasantly loud), an increase in the loudness level of 9 phon is approximately equivalent to doubling the loudness. An empirical formula relating loudness and loudness level over the linear portion is¹⁵

$$N = 0.046 \times 10^{L_N/30} \quad (11.14)$$

Notice there is no dependence on frequency.

Loudness and intensity can now be related. For a frequency of 1 kHz, where by definition the loudness level L_N is equal to the intensity level L_I , substitution of this equality into (11.14) and use of $L_I = 10 \log(I/10^{-12})$ results in

$$N (1 \text{ kHz}) = 460 \sqrt[3]{I}$$

where I is the intensity (MKS) of the 1-kHz tone. For any other frequency, refer to Fig. 11.10 and make a plot of L_N versus L_I . It will be seen that for any frequency the resulting curve will be very close to a straight line with a slope of +1. Fitting this

¹⁵ For a standardized engineering approximation, ISO R131-1959, see Problem 11.20.

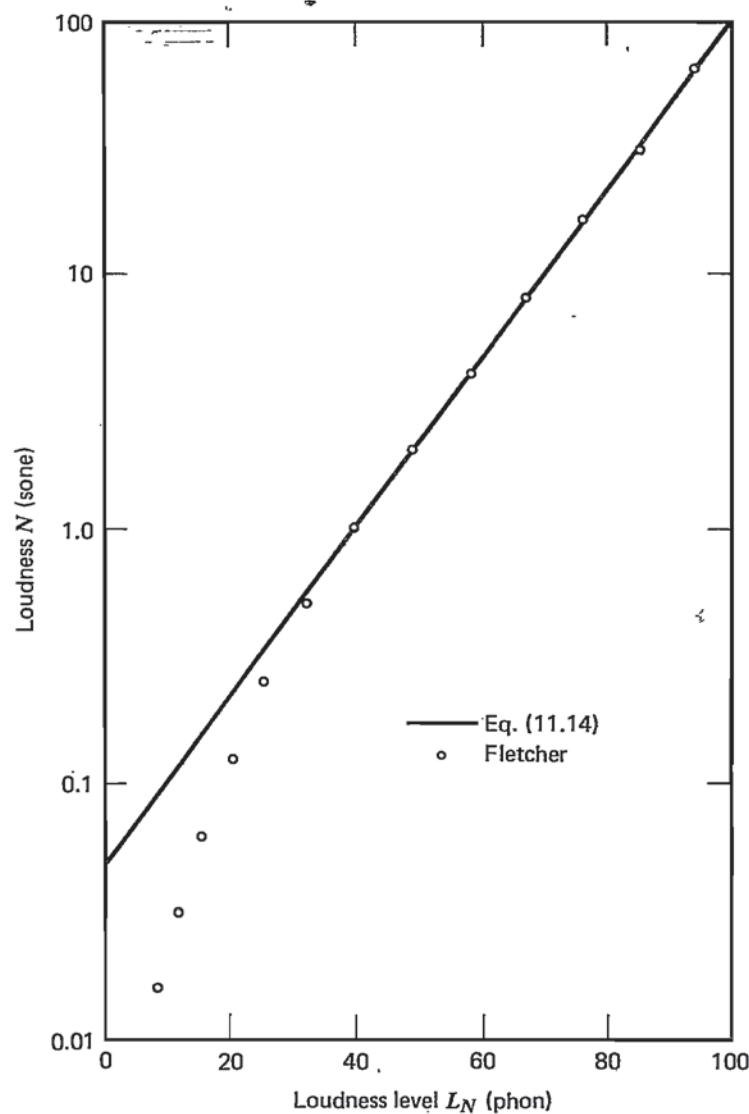


Fig. 11.13. Loudness versus loudness level.

curve with a straight line of this slope over the loudness levels of interest will yield

$$L_N \doteq L_I + 30 \log F(f)$$

where $F(f)$ is an empirically determined parameter depending only on frequency. Substitution into (11.14) gives

$$N \doteq 460F(f)\sqrt[3]{I} \quad (11.15)$$

This relation is one example of a general psychophysical "power law" of sensation postulated by Stevens.¹⁶

$$\begin{aligned} S &= G(\phi - \phi_T)^v & \phi &\geq \phi_T \\ &= 0 & \phi &< \phi_T \end{aligned}$$

¹⁶ Stevens, *Science*, 133, 80 (1961).

where S is the sensation, ϕ the stimulus, ϕ_T the threshold value, and G and v are constants depending on the quantities represented by S and ϕ . Equation (11.15) is quite accurate for the range 500 Hz to 5 kHz and moderate loudness levels.

If two or more tones are sounded simultaneously, the total loudness depends on whether they lie within a critical band.

1. Tones lying within one critical bandwidth are sensed according to the overall power, so their intensities add and the loudness is given by

$$N(\text{critical band}) = 460F(f) \left(\sum_i I_i \right)^{1/3}$$

2. Tones differing by more than the relevant critical bands are sensed as well-separated regions on the basilar membrane and the loudness add, $N = \sum N_i$.

3. If the tones are *considerably* different in their loudness, or widely separated in frequency, the evaluation of loudness becomes difficult, often tending to be based on the loudest of the tones.

As an example of the second case, let us determine the composite loudness of six pure tones of frequencies 125, 250, 500, 1000, 2000, and 4000 Hz and all having an *SPL* of 60 dB. This computation is summarized in Table 11.3. The result of 34 sone

Table 11.3. Sample calculation of loudness

Frequency f (Hz)	Intensity Level L_I (dB)	Loudness Level L_N (phon)	Loudness N (sone)
125	60	55	3.2
250	60	62	5.4
500	60	63	5.9
1000	60	60	4.7
2000	60	62	5.4
4000	60	69	9.3
Total			33.9

is equivalent to about 84 phon. This corresponds to an intensity level of 84 dB for a 1000-Hz tone. Since the sum of the intensities of the six tones corresponds to an intensity level of only 68 dB, it is apparent that the sound energy appears to be louder when it is distributed over a wide range of frequencies rather than lying within one critical bandwidth.

Calculations of the overall loudness of broadband noise (or even more complicated combinations of tones) becomes quite complex. The loudness becomes a function of the masking interactions among the sounds in the various critical bandwidths. The procedure has been systematized by Stevens through a number of permutations, and the methodology along with the associated tables necessary for the calculation is contained in reference 17.

¹⁷ Stevens, J. *Acous. Soc. Am.*, **51**, 575 (1972).

11.9
is its
varic
wave
mode
most
frequ
40 ph
10 p
betw
loud
the in
F
pitch
pure
this
soun
tone
fixed

scale
maint

11.9 PITCH AND FREQUENCY. Another subjective characteristic of a sound is its *pitch*. Pitch, like loudness, is a complex characteristic and is dependent on various physical quantities. While determined *primarily* by frequency, intensity and waveform are also influential. If a pure tone of about 100 Hz is sounded at a moderate and then at a high loudness level, the louder sound has a lower pitch. The most pronounced *decrease* in pitch with increasing loudness occurs for tones with frequencies below about 300 Hz. When the loudness of such a tone is increased from 40 phon to 100 phon, it may be necessary to increase the frequency by as much as 10 percent in order to maintain the pitch at a constant value.* For frequencies between 500 and 3000 Hz, the pitch of a tone is relatively independent of its loudness. For tones with frequencies above 4 kHz, the pitch *increases* with loudness, the increase being greater for higher frequencies.

For any particular loudness it is possible to assign numbers to the perceived pitches describing how "high" they sound, thereby establishing a pitch scale for pure tones. The reference frequency is usually 1 kHz, and the tone corresponding to this frequency is said to have a pitch of 1000 *mel*. A tone whose pitch is 500 mel sounds half as high, and a tone of 2000 mel sounds twice as high, as the 1000 mel tone. A curve giving the relationship between subjective pitch and frequency at a fixed loudness is shown in Fig. 11.14.

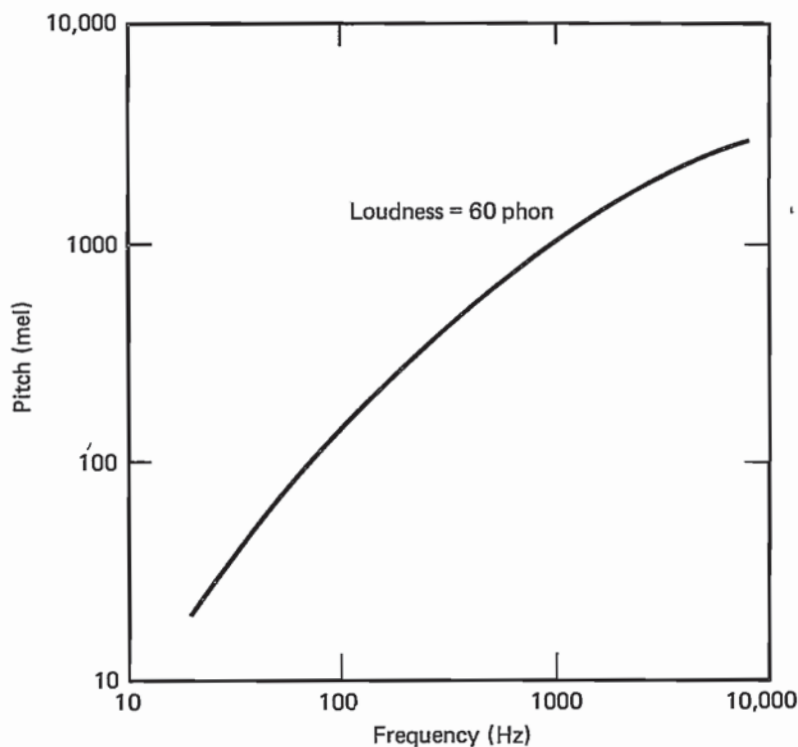


Fig. 11.14. Pitch as a function of frequency for a 60-phon pure tone.

* As noted earlier, the fundamental frequencies of adjacent semitones in the even-tempered musical scale are related by $f_1/f_2 = 2^{1/12} = 1.059$. Thus, the 10 percent change in frequency necessary to maintain the same pitch corresponds to a change of nearly a whole tone.

Musical sounds of differing loudnesses generate much smaller changes in pitch, usually not more than one-fifth as much as for similar pure tones. Furthermore, there are much smaller discrepancies between the fundamental frequency and pitch. Both these effects occur because musical sounds are usually rich in harmonics, some of which may have amplitudes exceeding that of the fundamental. Even if the fundamental lies in the frequency range where a pure tone shows a large decrease in pitch with loudness, the harmonics will have frequencies for which the pitch changes very little, so that the ear, aided by the consonances of all the harmonics, judges the sound as remaining at essentially the same pitch. The pitch of a complex sound is thus determined by the harmonics, which explains why a radio or TV, whose low-frequency response may extend down to only 150 Hz or so, still gives a reasonable approximation of the original sound (at least it is not unrecognizable).

11.10 THE VOICE. The acoustic energy associated with speech originates in the chest muscles which, by contraction, force air from the lungs up through the various components of the vocal mechanism (Fig. 11.15). This steady stream of air may be looked on as a *carrier* of

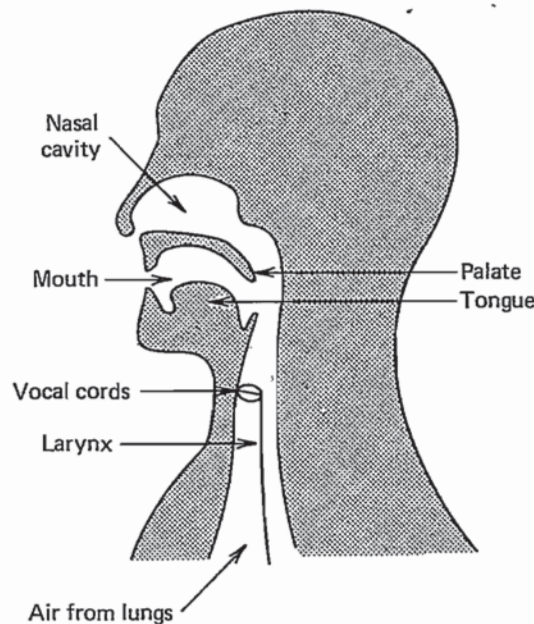


Fig. 11.15. Sectional view of the head, showing the important elements of the voice mechanism.

energy that must be modulated in its velocity and corresponding pressure in order to produce sounds. The requisite modulation is accomplished in two basic ways, leading respectively to *voiced* and *unvoiced* sounds.

Voiced sounds include the vowels of ordinary speech as well as tones characteristic of the singing voice. The primary modulating agent for voiced sounds is the *larynx*, across which are stretched the *vocal cords*. The vocal cords are two membranelike bands forming a diaphragm with a slitlike opening that modulates the airstream as it vibrates open and shut. The length of this opening (about 2.5 cm in men and 1.5 cm in women) and the tension to which the

vocal cords are stretched determine the fundamental frequency of this modulation. The shorter and lighter vocal cords possessed by most women vibrate almost twice as rapidly as those possessed by men. This accounts for the higher pitch of most women's voices. The action of the vocal cords produces a *sawtooth* pressure waveform containing many harmonics.

The resonating cavities and orifices of the nose, mouth, and throat form an acoustic filtering network altering the relative amounts of the harmonics. Many of these are controllable at will (changing the position of the tongue or altering the configuration of the lips) and thereby a wide variety of voiced sounds may be produced.

It is also possible for the voice mechanism to produce unvoiced sounds. For instance, a steady forcible exhaling of the breath will produce a hissing sound caused by turbulences set up in the flow of air through the numerous irregularities along the vocal tract. These sounds include such unvoiced *fricative* consonants as *f* and *s* as well as the unvoiced *stop* consonants

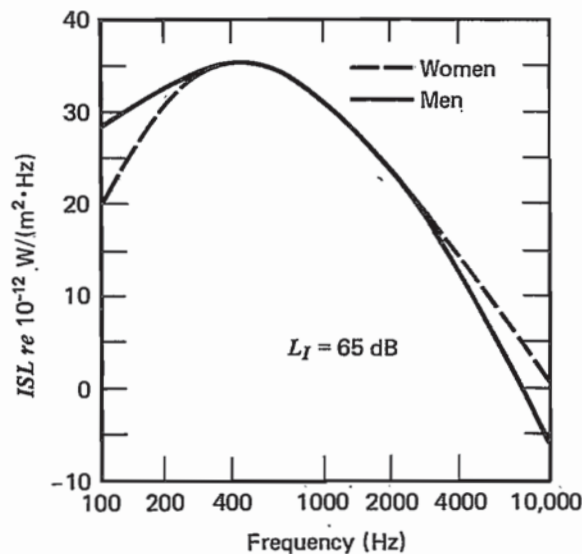


Fig. 11.16. Spectra levels for average conversational speech measured 1 m from the mouth.

p, *t*, and *k*. Here the sounds are produced by modulating the airstream with the lips, teeth, or tongue. As with the voiced sounds, conscious control of the tongue and lips alters the resonances of the cavities and orifices, producing a wide variety of unvoiced sounds so that recognizable speech can be generated—*whispering*. Spectrum analysis of the unvoiced sounds reveals a band of practically continuous frequency coverage largely confined to the upper portion of the audible range.

The distribution of average speech power with frequency has been measured by several investigators.¹⁸ Typical average speech spectra are shown in Fig. 11.16. The overall L_I at 1 m is about 65 dB.

The acoustic power, averaged over 2 to 4 s, generated by a speaker at conversational level is about $10 \mu\text{W}$. Very loud talking generates about $200 \mu\text{W}$ and shouting about $1000 \mu\text{W}$. By contrast, whispering generates about $0.001 \mu\text{W}$. When the power of conversational-level speech is averaged over a time interval that is short compared to the duration of a

¹⁸ Dunn and White, *J. Acous. Soc. Am.*, **11**, 278 (1940).

syllable (about 0.2 s), sizable fluctuations in level are observed as different speech sounds are uttered. For example, the power of the vowel *o* as in *low* is about $50 \mu\text{W}$, whereas the weak consonant *v* has an average power of only $0.03 \mu\text{W}$.

PROBLEMS

11.1 An acoustic signal consists of three tones, each of different frequency and different effective pressure amplitude: $P_1 = 5 \times 10^{-2} \text{ Pa}$, $f_1 = 104 \text{ Hz}$; $P_2 = 7 \times 10^{-2} \text{ Pa}$, $f_2 = 190 \text{ Hz}$; $P_3 = 0.1 \text{ Pa}$, $f_3 = 237 \text{ Hz}$. Find the intensity in the following bands: (a) 100 to 110 Hz, (b) 100 to 150 Hz, and (c) 150 to 300 Hz.

11.2 The results of a noise analysis are as follows:

Filter #	f_l	f_u	$V = \text{effective voltage output}$
1	100	200	7.1 mV
2	200	400	6.3
3	400	800	11.2
4	800	1600	8.9
5	1600	3200	11.2
6	3200	6400	7.9

(a) If the sensitivity of the receiver is $5 \times 10^{-2} \text{ V/Pa}$, find the effective pressure of the sound within the bandwidth of each filter. (b) Find the intensity of the sound within the bandwidth of each filter. (c) Find the band level (*re* $20 \mu\text{Pa}$) for each filter. (d) Find the pressure spectrum level (*re* $20 \mu\text{Pa/Hz}^{1/2}$) for the bandwidth of each filter. (e) Use the intensities obtained in (b) to find the band level between 100 and 6400 Hz. (f) Use the band levels from (c) to find the total band level between 100 and 6400 Hz. Compare this to the answer for (e).

11.3. (a) A tone with a sound pressure level of 140 dB *re* $1 \mu\text{Pa}$ is superimposed on a background noise with a constant pressure spectrum level of 150 dB *re* $1 \mu\text{Pa}$. Calculate the band levels obtained when the tone and the background are combined in filters of bandwidths 1, 10, or 100 Hz. Find the averaged pressure spectrum level in each of these filters. (b) Repeat for a tone with a sound pressure level of 150 dB *re* $1 \mu\text{Pa}$. (c) Repeat for a tone with a sound pressure level of 160 dB *re* $1 \mu\text{Pa}$. (d) Comment on the effect of increasing the sound pressure level of the tone.

11.4. A noise has a spectrum such that the intensity I_1 in each 1-Hz band is given by $I_1 = 10^{-6}/f \text{ W/m}^2$ where f is the center frequency of the band in hertz. (a) Compute the intensity spectrum level at 100, 500, and 1000 Hz. (b) What is the intensity level of the entire band of frequencies between 100 and 1000 Hz.

11.5. A noise is represented by an rms acoustic pressure $P_1 = 500/f \mu\text{bar}$, where P_1 is the pressure in a 1-Hz band centered on the frequency f in hertz. (a) Derive a general expression for the pressure spectrum level PSL of this sound. (b) How does the pressure spectrum level change with frequency expressed in decibels per octave? (c) What is the band level of this noise in a band of 50-Hz width centered on a frequency of 2500 Hz?

11.6. A subject is presented with samples of noise and samples of noise with a signal. In each of the following cases, qualitatively explain what happens to $P(D)$, $P(FA)$, and d' . (a) The mean amplitude of the signal gets weaker and $P(FA)$ does not change. (b) The subject is instructed to say there is a signal present only "when he is sure this is true" rather than "when it may be true." (c) The mean amplitude of the noise gradually gets smaller. (d) The subject is instructed "don't give so many false alarms."

11.5
yes-no
with no
 $P(FA)$

11.4
choice,
amplitude
present
detected
each, an

11.3
integrat
with no
law det

11.2
effective
pressure
 $0.41 \times$
signal
 $P(D) =$
calculat
ms and

11.1
in a fac
that the

11.
detector
that int
fixed d
 $P(D) =$

(a) Fin
(b) Det
(c) Det
calcula
Hz, and

11
sound
sound
5000-Hz

11.
 $p = P$
where
having

11.7. Two experiments are designed to detect the presence of a signal in noise. One is a yes-no task and the other is a 2AFC task, and both use the same samples of noise and signal with noise. In the first experiment $P(FA) = 0.002$ when $P(D) = 0.5$. For the same $P(D)$, what $P(FA)$ will be expected in the second experiment?

11.8. In a yes-no task with specific instructions to each of the subjects for making the choice, three subjects listen to samples of Gaussian noise with or without a signal of constant amplitude such that $d' = 1$. Subject 1 detected the signal 10 percent of the time when it was present, subject 2 detected the signal 70 percent of the time it was present, and subject 3 detected the signal 40 percent of the time it was present. (a) Estimate the respective $P(FA)$ for each, and (b) list the subjects in order of increasing desire to avoid a false detection.

11.9. A detection system has fixed bandwidth w and a postdetection filter of fixed integration time T_s , but the processing time τ is adjustable. For samples of noise and signal with noise with fixed d' , qualitatively explain how the detection threshold changes for square-law detection.

11.10. A signal consists of a tone burst of duration $\tau = 1$ s, a frequency 200 Hz, and an effective pressure amplitude of 0.02 Pa. The noise with which it is combined has a mean pressure amplitude of 2.83×10^{-2} Pa, a bandwidth of 100 Hz, and a standard deviation of 0.41×10^{-2} Pa. (a) Calculate the sound pressure levels *re* 20 μ Pa of the signal, noise, and signal with noise. (b) Calculate d' for this case assuming a yes-no task. (c) Find $P(FA)$ for $P(D) = 0.5$. (d) If the detector is a square-law processor with a postdetection filter with $T_s = \tau$, calculate the detection threshold for the specified $P(FA)$ and $P(D)$. (e) Repeat (d) if $T_s = 500$ ms and if $T_s = 2$ s.

11.11. What must be the sound pressure level of a 200-Hz pure tone if it is to be audible in a factory noise spectrum level of 73 dB *re* 20 μ Pa/Hz^{1/2} between 100 and 300 Hz? Assume that the critical ratio is the operative.

11.12. Assume that the ear can be modeled by a collection of parallel square-law detectors, each of which has a bandwidth given by w_{cr} and is followed by a postdetection filter that integrates over time T_s . In three separate experiments, using sample of 200-Hz signals of fixed duration τ in Gaussian noise, the following relationships were found to result in $P(D) = 0.6$ and $P(FA) = 0.05$:

Experiment	τ	$A_{S,N}$	A_N	σ
1	4 s	3.44×10^{-2} Pa	3×10^{-2} Pa	0.22×10^{-2} Pa
2	1	3.44×10^{-2}	3×10^{-2}	0.22×10^{-2}
3	0.1	4.51×10^{-2}	3×10^{-2}	0.22×10^{-2}

(a) Find the behavior of the detection threshold DT' as a function of τ for fixed d' . (b) Determine the detection threshold of the processor with filter for each experiment. (c) Determine the integration time T_s . (d) Assuming that T_s is independent of frequency, calculate the resultant DT' for tones of 1-s duration at 100 Hz, 200 Hz, 500 Hz, 1000 Hz, 2000 Hz, and 5000 Hz for the same $P(D)$ and $P(FA)$.

11.13. Use Figs. 11.9 and 11.11 to answer the following questions. (a) What must be the sound pressure level of a 1000-Hz tone if it is to be heard in the presence of a 2000-Hz tone of sound pressure level 80 dB *re* 20 μ Pa? (b) What must be the sound pressure level of a 5000-Hz tone in order to be heard in the presence of the 2000-Hz tone?

11.14. (a) Let the pressure exerted by acoustic waves on the eardrum be represented by $p = P \cos \omega t$. Assuming that the subjective response r may be represented by $r = a_1 p + a_2 p^2$, where all the a 's are constants, prove that the response contains a constant term, and terms having angular frequencies ω and 2ω . Compute the amplitude of each term as a function of P

and the constants. (b) If the incident sound consists of two wave motions of frequency ω_1 and ω_2 which produce a pressure $p = P_1 \cos \omega_1 t + P_2 \cos \omega_2 t$, determine the frequencies and amplitudes of the response.

11.15. (a) Determine the loudness level and loudness of a pure tone having a frequency of 100 Hz and an intensity level of 60 dB *re* 1 pW/m². (b) To what intensity level must this tone be reduced in order to reduce its loudness to one-tenth of the value obtained above? (c) To what intensity level must it be increased in order to raise its loudness to ten times the value determined in part (a)?

11.16. Six pure tones have the following frequencies and intensity levels (*re* 1 pW/m²): 50 Hz at 85 dB, 100 Hz at 80 dB, 200 Hz at 75 dB, 500 Hz at 80 dB, 1000 Hz at 75 dB, and 10,000 Hz at 70 dB. (a) Determine the loudness level for each of the tones. (b) Assume the intensity level of each of the tones to be decreased by 30 dB. Determine the new values of the loudness level for each of the tones.

11.17. (a) Compute the overall intensity level for the six tones of Problem 11.16(a). (b) What is the total loudness in sone of these six tones? (c) What is the required intensity level of a single 1000-Hz tone having the loudness of part (b)?

11.18. Three pure tones have the following frequencies and sound pressure levels *re* 0.0002 μ bar: 100 Hz at 60 dB, 200 Hz at 60 dB, and 500 Hz at 55 dB. (a) Which tone is the loudest? (b) What is the overall sound pressure level of these three tones when sounded simultaneously? (c) What is their total loudness level in phons?

11.19. An approximation is that a 10-dB increase in loudness level is equivalent to twice the loudness. Plot this on Fig. 11.13 normalized so that 1 sone and 40 phon are equivalent, and compare with the plot of (11.14).

11.20. For engineering applications, ISO R131-1959 recommends

$$N = 0.0625 \times 10^{0.03 L_N}$$

 Plot this on Fig. 11.13, and compare with the results of Problem 11.19.