# The Physics and Psychophysics of Music
## An Introduction

*Fourth Edition*

Juan G. Roederer

# The Physics and Psychophysics of Music

## An Introduction

*Fourth Edition*

Springer

Juan G. Roederer
Geophysical Institute
University of Alaska
Fairbanks, AK 99775-7320
USA
jgr@gi.alaska.edu

ISBN: 978-0-387-09470-0 e-ISBN: 978-0-387-09474-8
DOI: 10.1007/978-0-387-09474-8

Library of Congress Control Number: 2008937029



*Cover credit:*
Organ: Groote Kerk, Haarlem (The Netherlands)
Photo by the author

Infant: EEG measurements of brain reactions to music
Photo courtesy of Laurel Trainor, McMaster Institute for Music and the Mind,
McMaster University, Canada

Printed on acid-free paper

springer.com

*Dedicated to the memory of my dear parents, who awakened and nurtured my love for science and music*

# Preface

This introductory text deals with the physical systems and biological processes that intervene in what we broadly call "music." We shall analyze what objective, physical properties of sound patterns are associated with what subjective, psychological sensations of music. We shall describe how these sound patterns are actually produced in musical instruments, how they propagate through the environment, and how they are detected by the ear and interpreted in the brain. We shall do all this by using the physicist's language and his method of thought and analysis—without, however, using complicated mathematics. Although no previous knowledge of physics, physiology, and neurobiology is required, it is assumed that the reader possesses high-school education and is familiar with basic aspects of music, in particular with musical notation, scales and intervals, musical instruments and typical musical "sensations."

Books are readily available on the fundamentals of physics of music (e.g., Benade, 1990; Pierce, 1983; Fletcher and Rossing, 1998; Johnston, 2003) and psychoacoustics, music psychology and perception (e.g., Plomp, 1976; Deutsch, 1982a; Zatorre and Peretz, 2001; Hartmann, 2005). An excellent text on musical acoustics is that of Sundberg (1991), still most useful 17 years later; comprehensive discussions of recent researches on pitch perception and related auditory mechanisms can be found in Plack et al. (2005). The purpose of the present volume is not to duplicate but to synthesize and complement existing literature. Indeed, my original goal in writing this book in the seventies was to weave a close mesh between the disciplines of physics, acoustics, psychophysics, and neurobiology and produce a *single-authored* truly interdisciplinary text on what is called "the science of music"—and this is still the goal of this fourth edition! I also hope that it will convey to the reader a bit of what I call "the music of science," that is, the beauty and excitement of scientific research, reasoning and understanding.

After the first 1973 edition, several reprints and two revised editions were published, as were translations into German, Japanese, Spanish and Portuguese. These are all personally most gratifying indicators, especially in view of the fact that the subject in question was always more of a hobby for me (being a space physicist), than an official occupation! This Fourth Edition was prepared

under the motto "*if it ain't broke, don't fix it*". Indeed, based on the fact that the previous edition has been called a "classic" by some reviewers, I felt that the main pedagogical structure of the book should be maintained intact, and that the only major changes should be restricted to updating some critical points, especially in the psychophysical and neurobiological areas. As a matter of fact, I find it rather remarkable that many statements that were mere conjectures or speculations in the previous edition, have been verified in measurements and experiments and now can be presented as scientific facts in the text.

One of the most painful parts of writing a book is deciding what topics should be left out, or grossly neglected, in view of the stringent limitations of space. No matter what the author does, there will always be someone bitterly complaining about this or that omission. Let me list here some of the subjects that were deliberately neglected or omitted—without venturing a justification. In the discussion of the generation of musical tones mainly basic mechanisms are analyzed, to the detriment of the presentation of concrete musical situations. The human voice has been all but left out and so have discussions of inharmonic tones (bells and percussion instruments) and electronic tone generation; computer-generated music is not even mentioned. On the psychoacoustic side, only the perception of single or multiple sinusoidal tones is discussed, with no word on noise-band or pulse stimuli experiments. There is only very little on rhythm, stereo perception, and historical development. Finally, this being a book on an eminently interdisciplinary subject intended mainly for students from all disciplines and university levels, including those in lower division, many subjects had to be simplified considerably—and I apologize to the experts in the various disciplinary areas for occasionally sacrificing parochial detail for the benefit of ecumenical understanding. For the same reason, in the literature references priority was given to the quotation of reviews and comprehensive articles in sources of more widespread availability to the intended readership, rather than articles in specialized journals. Detailed references of original articles can be found in many of the quoted reviews.

The first edition was an offspring of a syllabus published by the University of Denver for the students in a "Physics of Music" course, introduced at that university more than 35 years ago, which quickly turned into a "Physics and Psychophysics of Music" course. In addition to regular class work, the students were required to perform a series of acoustical and psychoacoustical experiments in a modest laboratory. Conducting such experiments, some of which will be described here, is essential for a clear comprehension of the principal concepts involved. Unfortunately they often require electronic equipment that is not readily available, even in well-equipped physics departments. I ask that the readers trust the description of the experiments and believe that they really do turn out the way I say they do! Whenever possible I shall indicate how a given experiment can be performed with the aid of ordinary musical equipment. For a list of possible errata, visit my personal Web page.

I am grateful to the director of the Geophysical Institute, Professor Roger Smith, for institutional support of my work, and to my wife Beatriz for her understanding and tolerance of the "extracurricular" time spent on rewriting this book.

Juan G. Roederer
Geophysical Institute, University of Alaska-Fairbanks
http://www.gi.alaska.edu/~Roederer
March 2008

# Contents

# 1
# The Science of Music and the Music of Science: A Multidisciplinary Overview

*"He who understands nothing but chemistry does not truly understand chemistry either"*

Georg Christoph Lichtenberg, physicist and satirical writer (1742–1799)

## 1.1 The Intervening Physical Systems

Imagine yourself in a concert hall listening to a soloist performing. Let us identify the systems that are relevant to the music you hear. First, obviously, we have the player and the instrument that "makes" the music. Second, we have the air in the hall that transmits the sound into all directions. Third, there is you, the listener. In other words, we have the chain of systems: *instrument* → *air* → *listener*. What links them while music is being played? A certain type and form of vibrations called sound, which propagates from one point to another in the form of waves and to which our ear is sensitive. (There are many other types and forms of vibrations that we cannot detect at all, or that we may detect, but with other senses such as touch or vision.)

The physicist uses more general terms to describe the three systems listed above. She calls them: *source* → *medium* → *receptor*. This chain of systems appears in the study of other physical interaction processes involving light, radio waves, electric currents, cosmic ray particles, etc. The source emits, the medium transmits, the receptor detects, registers, or, in general, is affected in some specific way. What is emitted, transmitted, and detected is energy—in one of its multiple forms, depending on the particular case envisaged. In the case of sound waves, it is elastic energy, because it involves oscillations of pressure, i.e., rapidly alternating compressions and expansions of air.[1] The patterns in which this energy is conveyed represent acoustic *information*—linking certain oscillation patterns at the source with intended effects at the receptor (also expressed in the form of oscillations). We thus say that a sound wave is a *carrier* of information, which may represent the content and meaning of speech and music (the energy conveyed is important, but does not *define* the words spoken or the music being played!).

[1] Sound, of course, also propagates through liquids and solids.

J.G. Roederer, *The Physics and Psychophysics of Music*,
DOI: 10.1007/978-0-387-09474-8_1, © Springer Science+Business Media, LLC 2008

Let us have a second, closer look at the systems involved in music. At the source, i.e., the musical instrument, we identify several distinct physical components:

1. The *primary excitation mechanism* that must be activated by the player, such as the bowing or the plucking action on a violin string, the air stream blown against a wedge in the flute, the reed in a clarinet, and the player's lips on a brass instrument, or in the case of a singer, the vocal folds in the larynx.[2] This excitation mechanism acts as the primary acoustic energy source.
2. The fundamental *vibrating element* which, when excited by the primary mechanism, is capable of sustaining well-defined vibration modes of specific frequencies, such as the strings of a violin, the air column in the bore of a wind instrument or organ pipe. This vibrating element actually determines the musical pitch of the tone and, as a fortunate bonus, provides the upper harmonics needed to impart a certain characteristic quality or timbre to the tone. In addition, it may serve as a vibration energy storage device. In wind instruments, it also controls the primary excitation mechanism through feedback coupling (strong in woodwinds, weak in brasses, and nonexistent in the harmonium and the human voice).
3. Many instruments have an additional *resonator* (sound board of a piano, body of a string instrument, bell of a wind instrument, buccopharyngeal cavity) whose function is to convert more efficiently the oscillations of the primary vibrating element (string, air column) into sound vibrations of the surrounding air and to give the tone its final timbre.

In the medium, too, we must make a distinction: We have the *medium proper* that transmits the sound and its *boundaries*, i.e., the walls, the ceiling, the floor, the people in the audience, etc., which strongly affect the sound propagation by reflection and absorption of the sound waves and whose configuration determines the quality of room acoustics (reverberation, echo).

Finally, in the listener, we single out the following principal components: (1) The outer ear with the *eardrum*, which picks up the pressure oscillations of the sound wave reaching the ear, converting them into mechanical vibrations that are transmitted via a link of three tiny bones to (2) the inner ear, or *cochlea*, in which the vibrations are sorted out according to frequency ranges, picked up by receptor cells, and converted into electrical nerve impulses. (3) The *auditory nervous system* transmits the neural signals to the brain where the acoustic information is processed, displayed as a neural image of auditory features in certain areas of the cerebral cortex, identified, stored in the memory, and eventually transferred to other centers of the brain for further cognitive processing and affective response. These latter stages lead to the conscious perception of musical sounds.

[2] To make the description complete we ought to add the following "components" of the player: the frontal lobes of his brain that tell the motor cortex to send commands to the specific muscles with which he activates the musical instrument or his vocal tract, the feedback from ears and muscles that aids him in controlling his performance, etc. However, in this book we shall leave the player completely out of the picture.

TABLE 1.1. Physical and biological systems relevant to music and their overall functions.

| | System | Function |
|---|---|---|
| Source | Excitation mechanism | Acoustic energy supply |
| | Vibrating element | Determination of fundamental tone characteristics |
| | Resonator | Final determination of tone characteristics |
| | | Conversion into air pressure oscillations (vibration patterns of sound waves) |
| Medium | Medium proper | Sound propagation |
| | Boundaries | Reflection, refraction, absorption |
| Receptor | Eardrum | Conversion into mechanical oscillations |
| | Inner ear | Primary frequency sorting |
| | | Conversion into nerve impulses |
| | Nervous system | Acoustic information processing |
| | | Transfer to specific brain centers |
| | | Cognitive processes and affective response |

Notice that we may replace the listener by a recording device such as a magnetic tape or digital disc recorder, or a photoelectric record on film, and still recognize at least three of the subsystems: The mechanical detection and subsequent conversion into electrical signals in the microphone, deliberate or accidental transformations or processing in the electronic circuitry, and memory storage on tape, disc, or film, respectively. The first system i.e. the instrument, of course, also may be replaced by an electronic playing device, in which we can easily recognize both the primary excitation mechanism and the vibrating element in the speaker. We may summarize this discussion in Table 1.1.

The main aim of this book is to analyze comprehensively what happens at each stage shown in this table and during each transition from one stage to the next, when music is being played on real instruments. However, we will not deal with electronic sound generation and recording, nor with the human voice.

## 1.2 Characteristic Attributes of Musical Sounds

Subjects from all cultures agree that there are three primary sensations associated with a single sustained, constantly sounding musical tone: *pitch*, *loudness*, and *timbre*.[3] We shall not attempt to formally define these subjective attributes or

[3]The sometimes quoted sensations of volume and density (or brightness) are composite concepts that can be "resolved" into a combination of pitch and loudness effects (lowering of pitch with simultaneous increase of loudness leads to a sensation of increased volume; rising pitch with simultaneous increase of loudness leads to increased density or brightness). They will not be considered in this book.

*psychophysical magnitudes*; we shall just note that pitch is frequently described as the sensation of "altitude" or "height," and loudness the sensation of "strength" or "intensity" of a tone. Timbre, or tone quality, is what enables us to distinguish among sounds from different kinds of instruments even if their pitch and loudness were the same. The unambiguous association of these three qualities to a given sound is what distinguishes a musical *tone* from "noise"; although we can definitely assign loudness to a given noise, it is far more difficult to assign a unique pitch or timbre to it.

The assignment of the sensations pitch, loudness, and timbre to a musical tone is the result of complex physical mechanisms in the ear and information-processing operations in the nervous system. As we shall discuss in Section 1.4, it is subjective and inaccessible to direct physical measurement. However, each one of these primary sensations can be associated, in principle, to a well-defined physical quantity of the original stimulus, the sound wave, which can be measured and expressed numerically by *physical* methods. Indeed, as we shall discuss in detail in Chaps. 2, 3, and 4, respectively, the sensation of pitch is primarily associated to the *fundamental frequency* (repetition rate of the vibration pattern in harmonic tones, described by the number of oscillation patterns per second), loudness to *intensity* (energy flow or pressure oscillation amplitude of the sound wave reaching the ear), and timbre to the "spectrum," or proportion in which other, higher, frequencies called upper *harmonics* appear mixed with each other.

This, however, is a far too simplistic picture. First, the pitch of a complex musical tone can be heard clearly even if the fundamental is absent (Sect. 2.7); it changes slightly when the loudness changes, and the same note may lead to a slightly different pitch sensation in one ear than in the other. Second, the sensation of loudness of a tone of constant physical intensity will appear to vary if we change the frequency, and the loudness of a superposition of several tones of different pitch each (e.g., a chord) is not related in a simple way to the sum of sound energy flows from each component; for a succession of tones of very short duration; on the other hand (e.g., staccato play), the perceived loudness also depends on how long each tone actually lasts (Sect. 3.4). Third, refined timbre perception as required for musical instrument recognition is a process that utilizes much more information than just the spectrum of a tone; the transient attack and decay characteristics are equally important (Sect. 4.8), as one may easily verify by trying to recognize musical instruments while listening to a magnetic tape played backwards.

To complicate the picture even further, there is a "top-down" influence of knowledge-driven processes in the brain, which introduces a heavily context-dependent bias in actual music perception. For instance, the tones of a given instrument may have spectral characteristics that change appreciably throughout the compass of the instrument, and the spectral composition of a given tone may change considerably from point to point in a music hall (Sect. 4.7)—yet they are recognized without hesitation as pertaining to the same instrument. Or, conversely, a highly trained musician may have greatest difficulty in matching the exact pitch of a single electronically generated tone deprived of upper harmonics, fed to her

ears through headphones, because her central nervous system is lacking some key additional information that normally comes with the "real" sounds with which she is familiar.

Another relevant physical characteristic of a tone is the spatial direction from which the corresponding sound wave is arriving. What matters here is the minute time difference between the acoustic signals detected at each ear, which depends on the direction of incidence. This time difference is measured and coded by the nervous system to yield the sensation of tone *directionality*, stereophony, or lateralization (Sect. 2.9).

When two or more tones are sounded simultaneously, our brain is capable of singling them out individually, within certain limitations. New, less well-defined but nevertheless musically very important subjective sensations appear in connection with two or more superposed tones, collectively leading to the concept of harmony. Among them are the "static" sensations of *consonance* and *dissonance* describing the pleasing or irritating character of certain superpositions of tones, respectively (Sect. 5.2); the "dynamic" sensation of the urge to *resolve* a given dissonant interval or chord (Sect. 5.5); the peculiar effect of *beats* (Sect. 2.4); and the different character of *major and minor chords*. In particular, as we shall see in Sect. 5.2, as the most "perfect" musical interval, the octave has a unique property: The pitches of two tones that are one or more octaves apart are perceived as belonging to the same pitch "family." As a result, all notes differing by one or more octaves are designated with the same name. This circular property of pitch (return to the same "family" after one octave when one moves up or down in pitch) is called *chroma*; it has intrigued people for thousands of years, yet today finds its explanation in physical/physiological/neural processes in the auditory system. All these "higher order" yet still fundamental musical sensations are universal, experienced by individuals from all cultures since very early age.

The correlation of pitch, loudness, and static aspects of timbre with specific physical characteristics of single tones is "universal"—i.e. independent of the cultural conditioning of a given individual. This even applies to the chroma and the preeminent roles of the octave and the fifth as perfect consonances. Such universal subjective attributes must be natural consequences of information-processing mechanisms in the acoustic neural system and hence, the result of evolution, not culture (see Sect. 5.5 and Appendix II). Even the existence of certain musical scales seems universal. Indeed, this is supported by recent archeological finds that indicate that musical scales already were in use in upper Paleolithic times (e.g., d'Errico et al., 2003), between 27,000 and 21,000 years ago (Fig. 1.1).

In all of this, of course, we only have been talking about the building blocks, i.e. , the common "infrastructure" of music. Actual music depends on how this infrastructure is *used*, that is, on how melodies, harmonies, and rhythm are put together. Here too, exist some basic rules, to be analyzed throughout the book, which emerge from the physiological and neural functions of the human auditory system. But as this assemblage becomes increasingly varied and complex, more and more it is influenced by the "environment," i.e. , the development of a

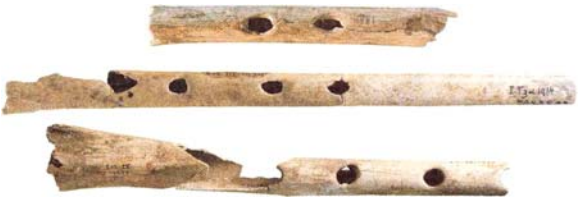


FIGURE 1.1 Pipes made of bird bone, dating from 27,000 and 21,000 years before present. (Source: Francesco d'Errico, Institut de Préhistoire et de Géologie du Quaternaire, Université Bordeaux 1, France; permission gratefully acknowledged.)

particular musical culture. As the brain is increasingly exposed to a repertoire of tone assemblages, context dependence takes over.

## 1.3 The Time Element in Music

A steady sound, with constant frequency, intensity, and spectrum is annoying. Moreover, after a while, our conscious present would not register it anymore. Only when that sound is turned off, may we suddenly realize that it had been there (Sect. 2.9). Music is made up of tones whose physical characteristics change with time in a certain fashion. It is only this time dependence that makes a perceived sound "musical" in the true sense. In general, we shall henceforth call a time sequence of individual tones or tone superpositions a *musical message*. Such a musical message may be "meaningful" (once called a "tonal Gestalt") if it carries *information* that in some way elicits a reaction in our brain that goes beyond merely noticing it, i.e. that triggers a series of brain operations involving analysis, association with previously stored messages, storage in the memory, and emotional response.

A *melody* is the simplest example of a musical message. Some attributes of meaningful musical messages are key elements in western music: tonality and leading note (domination of a single tone in the sequence), the sense of return to the tonic, modulation, and rhythm (Sect. 5.5). A fundamental characteristic of a melody is that the succession of tones proceeds in discrete, finite steps of pitch in practically all musical cultures. This means that out of the infinite number of available frequencies, our auditory system prefers to single out discrete values corresponding to the notes of a *musical scale*, even though we are able to detect frequency changes that are much smaller than the basic step of any musical scale (Sect. 5.3). Another characteristic is that the neural mechanism that analyzes a musical message pays attention only to the *transitions* of pitch; "absolute" pitch identification (perfect pitch) is lost at an early age in most individuals.

Let us examine the time element in music more closely. There are three distinct time scales on which time variations of psychoacoustic relevance occur.

First, we have the "microscopic" time scale of the actual vibrations of a sound wave, covering a range of periods from about 0.00007 to 0.05 s. Then there is an "intermediate" range centered at about one-tenth of a second, in which some transient changes such as tone attack and decay occur, representing the time variations of the microscopic features. Finally, we have the "macroscopic" time scale, ranging from about 0.1 s upward, corresponding to common musical tone durations, successions, and rhythm. It is important to note that each typical time scale has a particular processing level with a specific function in the auditory system. (1) The microscopic vibrations are detected and coded in the *inner ear* (Sect. 2.8) and mainly lead to the primary tone sensations (pitch, loudness, and timbre). (2) The intermediate or transient variations seem to affect mainly processing mechanisms in the *neural pathways* from the ear to the auditory areas of the brain (Sect. 2.9) and provide additional cues for quality perception, tone identification, and discrimination (e.g., Sect. 4.9). (3) The macroscopic time changes are processed at the highest neural level—the *cerebral cortex*;[4] they determine the actual musical message and its cognitive attributes (Sect. 4.9). The higher we move up through these processing stages in the auditory pathway, the more difficult it becomes to define and identify the psychological attributes to which this processing leads and the more everything is influenced by the context in which the tone appears, i.e., by learning and cultural conditioning, as well as by the current emotional and behavioral state of the individual. But even this context dependence is, to a considerable extent, controlled by the universal way the human brain processes acoustic information (Sect. 5.6).

For more than 100 years, musicologists have bitterly complained that physics of music and psychoacoustics have been restricted mainly to the study of production and perception of steady, constant tones or esoteric, laboratory-generated tone complexes. Their complaints are well founded, but the reasons for such a restriction are well founded too. As explained above, the processing of tone sequences occurs at the highest level of the central nervous system, involving a complex and still little-explored chain of mechanisms. Before these can be tackled scientifically, all contributing basic building blocks—the fundamental simple physical and psychoacoustic mechanisms—must be clearly understood. However, we should point out that the noninvasive techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) are indeed providing fundamental new insights concerning the *neural correlates* of "real music" perception (Sect. 4.9), i.e. the specific neural activity and interactions involved in musical information processing.

[4]The folded outer layer of white neural tissue in which the fundamental sensory and cognitive information processing takes place (see Sect. 5.6). With a few exceptions, we will not deal with specific brain anatomy and neurophysiology; there are many traditional and modern books on these subjects available in medical libraries (e.g., Brodal, 1969; Hohne, 2001).

## 1.4 Physics and Psychophysics

We may describe the principal objective of physics in the following way: To provide methods by means of which one can quantitatively predict the evolution of a given physical system (or "retrodict" its past history), based on the conditions in which the system is found at any one given time. For instance, given an automobile of a certain mass and specifying the braking forces, physics allows us to predict how long it will take to bring the car to a halt and where it will come to a stop, provided we specify the position and the speed at the initial instant of time. Given the mass, length, and tension of a violin string, physics predicts the possible frequencies with which the string will vibrate if plucked or hit in a certain manner (Sect. 4.3). Given the shape and dimensions of an organ pipe and the composition and the temperature of the gas inside (air), physics predicts the frequencies of the fundamental and overtones of the sound emitted when it is blown (Sect. 4.5).

In classical physics, "to predict" means to provide a mathematical framework, a series of algorithms, equations or "recipes" which, based on the physical laws that govern the system under analysis, establish mathematical relationships between the values of the physical magnitudes that characterize the system at any given instant of time (position and speed in the case of the car; frequency and amplitude of oscillation in the other two examples). These relations are then used to find out what the values are and how they change with time.

In order to establish the physical laws that govern a given system, we must first observe the system and make quantitative *measurements* of relevant physical magnitudes to find out their causal interrelationships experimentally. A physical law expresses a certain relationship that is common to many different physical systems and independent of particular circumstances. For instance, the laws of gravitation are valid here on Earth, for the solar system, for a star orbiting a galaxy and elsewhere else in the universe. Newton's laws of motion apply to all bodies, regardless of their chemical composition, color, temperature, speed, size, or position.

Most of the actual systems studied in physics—even the simple and familiar examples given above—are so complex that accurate and detailed predictions are impossible. Thus, we must make approximations and devise simplified *models* that represent a given system only by its main features. The ubiquitous "mass point" to which a physical body is often reduced in introductory physics courses—be it a planet, an automobile, or a gas molecule—is the most simplified model of all! Likewise, the study of vibrating strings and organ pipes begins by assuming that these strings and pipes are infinitesimally thin objects; later, the model is refined by giving them a more realistic cylindrical (or conical) form (Chap. 4). Many times it is necessary to break up the system under study into a series of more elementary subsystems physically interacting with each other, each one governed by a well-defined set of physical laws.

Turning to psychophysics, as happens with physics in general, it tries to make predictions on the response of a specific system subjected to given initial conditions. The system under consideration is a subject's (or an animal's) *sensory system* (receptor organ and related parts of the nervous system), the conditions

are determined by the *physical input stimuli*, and the response is expressed by the *psychological sensations* evoked in the brain and reported by a human subject or manifested in the sensory-specific behavior of an animal. In particular, *psychoacoustics*, a branch of psychophysics, is the study that links acoustic stimuli with auditory sensations. Again like physics, psychophysics requires that the causal relationship between physical stimulus input and psychological (or behavioral) output be established through experimentation and measurement, and it must make simplifying assumptions and devise models in order to be able to establish quantitative mathematical relationships and venture into the business of prediction-making. In the early times of psychophysics, the empirical input–output relationships were condensed into so-called psychophysical laws, treating the intervening "hardware" as a black box. Today, psychophysical models take into account the physiological functions of the sensory organs and pertinent parts of the nervous system.

Unlike classical physics, but strikingly similar to quantum physics,[5] most measurement processes in psychophysics will substantially perturb the system under observation (e.g., a subject reporting the sensations caused by a given physical stimulus, an animal trained to respond in certain fashion to certain stimuli), and little can be done to eliminate said perturbation completely. As a consequence of all this, the result of a psychophysical measurement does not reflect the state of the system per se, but rather, the more complex state of "a system *under observation*." Unlike classical physics, but strikingly similar to quantum physics, psychophysical predictions cannot be expected to be exact or unique—only the likelihood of an

[5]The physics of daily life's world, or *classical physics*, assumes that both, measurements and predictions should always be exact and unique, the only limitations and errors being those caused by the imperfection of our measuring methods and numerical calculations (or, in the case of chaotic systems like a pinball machine, by the physical impossibility of reproducing *exactly* the same initial conditions). In the atomic and subatomic domain, however, this view is no longer tenable. Nature is such that no matter how much we try to improve our techniques, most measurements will always be of limited accuracy, and only *probabilities*, that is, likelihood, can be predicted for the values of physical magnitudes in the atomic domain. For instance, it is impossible to predict *when* a given radioactive nucleus will decay (even if we had been waiting a terribly long time), or exactly *where* an electron of given energy will be found at a given time during its journey from the cathode to the TV monitor screen—only probabilities can be specified. An entirely new physics had to be developed in the early 1900's, fit to describe atomic and subatomic systems—the so-called *quantum mechanics*. When we try to apply to the quantum domain the ways of thinking that our brain has acquired during its interaction with the macroscopic classical world and try to imagine what must be happening "inside" a quantum system while it remains unobserved, we have to invoke a paradoxical, counterintuitive, and often outright spooky behavior if we want to "explain" the results of a measurement. Yet quantum mechanics has been extraordinarily successful, and we must resign ourselves to the fact that we cannot find out, not even in principle, what *exactly* happens inside a quantum system while it is left alone between measurements—the only extractable information being that coming from a far more complex entity, namely "the quantum system *under observation*".

outcome, i.e., its probability value, can be determined.[6] Unlike classical physics, but strikingly similar to quantum physics, one and the same input stimulus can lead to different discrete outputs, as in the multiple ambiguous pitch sensations of certain pure tone superpositions (Appendix II). In general, psychophysics requires experimentation with many different equivalent systems (subjects) exposed to identical conditions, and a statistical interpretation of the results.

Quite obviously, there are some limits to these analogies. In physics, the process or "recipe" of the measurement which defines a given physical magnitude, such as the length, mass, or velocity of an object, can be formulated in a rigorous, unambiguous way. As long as we deal with physiological output, such as neural impulse rate, amplitude of evoked goose bumps or increase in heartbeat rate, psychophysical measurements can be expressed in a rigorous, quantitative way too. But in psychoacoustics, how do we define and measure the subjective sensations of pitch, loudness, timbre or—to make it even trickier—the magnitude that represents the urge to bring a given melody to its tonic conclusion? Or how would we arrange measurements on "internal hearing," i.e., the action of provoking musical tone images by volition, without external stimuli? Could this be done only with fMRI techniques or by implantation of microelectrodes into brain cells? As we shall see in Sect. 5.6, such procedures tell us about the location of the neuropsychological processes involved, but they still would not provide any quantitative information on the actual *feelings* experienced by the subject!

Many sensations can be *classified* into more or less well-defined types (called sensory qualities if they are caused by the same sense organ)—the fact that people do report to each other on pitch, loudness, tone quality, consonance, etc., without much mutual misunderstanding with regard to the meaning of these concepts, is an example. Furthermore, two sensations belonging to the same type, experienced one immediately following the other, can in general be ordered by the experiencing subject as to whether the specific attribute of one is felt to be "greater" (or "higher," "stronger," "brighter," "more pronounced," etc.), "equal," or "less" than the other. For instance, when presented with two tones in a succession in a forced-choice experiment, the subject must judge whether the second tone was of higher, equal, or lower pitch than the first one (e.g., Sect. 2.4). Another example of ordering is the following: Presented with the choice of three complex tones of the same pitch and loudness, he may order them in pairs by judging which two tones have the most similar timbre and which the most dissimilar one (Sect. 4.8). One of the fundamental tasks of psychophysics is the determination, for each type of sensation, of the minimum detectable value (or threshold value) of the physical magnitude responsible for the stimulus, the minimum detectable change or *difference limen* (DL—also called "just noticeable difference"), and the minimum discrimination between two simultaneous sensations of the same type (MD) (Sects. 2.3

[6] We must emphasize that these are only *analogies*. Quantum physics as such does not play an explicit role in integral nervous system function (only in the chemical and electrochemical reactions inside neurons and between them).

and 3.4). In general, psychoacoustic measurements with human subjects involve exposure to electronically generated sounds fed into headphones in an acoustically isolated room (anechoic chamber). The subjects are then asked to follow a strict protocol of listening to probe tones and comparing them with reference tones, and reporting the results of their sensations in as much an objective way as possible.

The ability, possessed by all individuals, to classify and order subjective sensations gives subjective sensations a status almost equivalent to that of a physical magnitude and justifies the introduction of the term *psychophysical magnitude*. What we must not expect a priori is that individuals can judge without previous training whether a sensation is "twice" or "half," or any other *numerical* factor that of a reference unit sensation. There are situations, however, in which it is possible to learn to make quantitative estimates of psychophysical magnitudes on a statistical basis and, in some circumstances, the brain may become very good at it. The visual sense is an example. After sufficient experience, the estimation of the size of objects can become highly accurate, provided enough information about the given object is available; judgments such as "twice as long" or "half as tall" are made without hesitation. It is quite clear from this example that a "unit" and the corresponding psychophysical process of comparison have been built into the brain only through *experience and learning*, in multiple contacts with the original physical magnitudes. The same can be achieved with other psychophysical sensations such as loudness; it is necessary to acquire through learning the ability of comparison and quantitative judgment. The fact that musicians all over the world use a common loudness notation (Sect. 3.4) is a self-evident example. And the fact that we can judge the dampened sound of a full organ chord listened to from outside the church, or that of a band playing in the distance as "fortissimo," is a clear example that loudness is a context-dependent psychophysical quantity.

Here we come to the perhaps most crucial differences between physics and psychophysics: (1) repeated measurements of the same kind may *condition* the response of the psychophysical system under observation; the brain has the ability of learning, gradually changing the response to the same input stimulus, as the number of similar exposures increases. (2) The degree of *motivation* of the subject under study and the consequences thereof, mental or physical, may interfere in a highly unpredictable way with the measurements. (3) An individual may be cued by the experimenter to focus, in her perception, on some specific ranges or contexts of the stimulus, and the results may reveal specific sensory ambiguities. As a consequence of the first point, a statistical psychophysical study with one single individual exposed to repeated "measurements" may not be identical to a statistical study involving one single measurement performed on many different individuals (exactly as it happens with the measurement of a quantum system!). This difference is due not only to differences among individuals, but also to the conditioning that takes place in the case of repeated exposures. In summary, the very complex feedback loops in the nervous system and the strategy of the brain of predicting in the short term what is to come (and then making corrections if

the prediction turns out wrong) make psychoacoustic measurements particularly tricky to plan, set up, and interpret.

## 1.5 Psychophysics and Neuroscience

Psychophysics is part of a larger, more encompassing discipline, namely neuroscience. For instance, *Psychoacoustics*, only addresses the question of *why* we hear what we hear when we are exposed to a given acoustic stimulus—but it does not deal with the *meaning* of acoustic input, leaving out all higher-level processes of cognition, emotional response, and behavior. Neuroscience or, more specifically, *systems neuroscience*[7] is the discipline that studies the functions of the neural system linking the *information* received from environment and body with the full cognitive, emotional, and behavioral output. Like physics, it also works with models. These are mainly models of functional interrelationships (e.g., information flowcharts) and, at the microscopic level, models of neural networks; although such models are only idealizations and approximations, the intervening neuroanatomical parts and physiological processes are taken into account realistically (Sect. 2.8).

The main system under study is, obviously, the brain. In brief, the most important "higher functions" of an *animal brain*—mainly its cerebral cortex—are environmental representation and prediction, and the planning of behavioral response, with the goal of maximizing the chances of survival and perpetuation of the species. To accomplish this, the brain must, in the long term, acquire the necessary sensory information to make "floor plans" of spatial surroundings and discover cause-and-effect relationships in the occurrence of temporal events, and, in the short term, assess the current state of environment and body, identify relevant features or changes, make short-term predictions based on experience (learned information) and instinct (genetic information), and execute a behavioral response that is likely to be beneficial for the organism (Sect. 4.9). The overall guidance and motivation to carry out these tasks is controlled by the *limbic system*, a phylogenetically old part of the brain (which in the popular literature is sometimes called "our lizard brain"), consisting of a group of nuclei sitting deep inside, but intimately connected with the cortex. The limbic system dispenses signals that make up the affective state of the organism (pleasure or pain, fear or boldness, love or hate, anxiety or hope, happiness or sadness, etc.). Sections 4.8 and 5.6 will deal in detail with brain function and its relevance to music perception.

The *human brain* can go "off-line," work on its own output, and plan a behavioral response which is completely independent of the current state of environment

[7] In earlier editions of this book we used the term "neuropsychology," but in some clinical communities this term is reserved for the study of the effects of lesions on specific brain functions. Neurobiology is also a commonly used term, but it encompasses more than the study of brain function.

and body, with a goal disconnected from the instantaneous requirements of survival (Sect. 5.6). It can recall information at will without external or somatic stimulation, analyze it, and store in memory modified versions thereof for later use—we call this *the human thinking process*. In addition, because of these "internal command" abilities, the human brain can overrule the dictates of the limbic system—a diet is a good example!—and also engage in information-processing operations for which it did not originally evolve—abstract mathematics and music are good examples!

All perceptual and cognitive brain functions are based on *electrical impulses* generated, transmitted, and transformed by neurons, the basic constituent elements of the nervous system (Sect. 2.8). There are more than ten billion of these cells in the human brain; one neuron can be connected to hundreds, even thousands of others, and each cerebral operation, however "simple," normally involves millions of neurons. It is in the architecture of synaptic interconnections of this conglomerate of neurons and their activation by electrical impulses that the mysteries of memory, consciousness, thinking, and feelings are buried (Sect. 5.6). Every brain operation, such as the recognition of a face that is being seen, the imagination of a musical sound, or the pleasure experienced by eating chocolate, is defined by a very specific *distribution in space and time of electrical neural activity*. The above-mentioned representation of the environment, or for that matter any mental image, even a totally abstract thought, is nothing but the appearance of a distribution of neural impulses in certain areas of the cortex that, while incredibly complex, contains patterns that are absolutely *specific* to what is being represented or imagined (its *neural correlate*).[8]

Because of the complexity involved, there is no hope, at least for the moment, to determine the full, detailed neural pattern experimentally and represent it in a mathematically tractable form. However, as we shall see in Sect. 2.8, it is possible to interrogate individual neurons with the implantation of microelectrodes registering the electric spikes of their activity in laboratory animals or in human brains during neurosurgery. On the other hand, it is possible to register average changes in the collective activity of hundreds, thousands, or millions of neurons by using the noninvasive tomographic imaging techniques of *functional magnetic resonance imaging* or fMRI and *positron emission tomography* or PET, or the older electric and magnetic *encephalography* (EEG and MEG, respectively) (Sects. 2.8 and 5.6). Comparison of clinical studies of patients with localized *brain lesions*, later identified in detail in an autopsy, was historically the first method used to identify the functions of specific brain regions.

[8]Note carefully that these patterns, although absolutely specific, do not bear any "pictorial" resemblance with what they represent! When you see a tree, think of a tree or dream of a tree nothing that resembles the form of a tree pops up in your brain—only a horribly complex distribution of neural activity that is always the same, specific to the cognition of a tree (Sect. 5.6).

The human brain is the most complex information system in the Universe as we presently know it. It is thus quite understandable that any scientist, let alone any scientifically untrained persons, have greatest difficulty in understanding why, despite this complexity, the function of our own brain appears to us so "simple" and as "one single whole" of which we feel totally in control (this is called "the natural simplicity of mental function" and "the unitary nature of conscious experience," respectively). Likewise, it is quite understandable that we have greatest difficulty in accepting the fact that to describe scientifically the function of the human brain in modern neuroscience, there is no need to invoke any separate, physically indefinable and immeasurable, concepts such as the "mind" or the "soul"!

## 1.6 Neuroscience and Informatics[9]

In the preceding sections, we have mentioned the concept of "information" several times, in different contexts. For instance, a musical message is, by the very meaning of that word, information (Sect. 1.3). But what *is* information? The mere asking of such a question seems absurd. Aren't we living in the "Information Age"? Information is shaping human society. Not just in recent times—it has been doing so since the beginning of the human race; information-processing power is what distinguishes us from animals. Much later in human evolution, great inventions facilitating the spread of information such as the ancient petroglyphs, Gutenberg's movable printing type, photography, sound recording, wireless communications, the computer, and the Internet have brought about explosive, revolutionary developments. Information, whether good, accidentally wrong, or deliberately false, whether educational, artistic, entertaining, or erotic, is now a trillion dollar business.

Information-processing machines are getting faster, better, cheaper, and smaller. Yet, as mentioned in the preceding section, the most complex, most sophisticated, most exquisite information-processing machine that has been in use more or less in its present shape for tens of thousands of years, and will remain so for a long time, is the human brain. Every task that the brain executes is an information-processing task—however simple, however complex. Our own self-consciousness, without which we wouldn't be humans, involves an interplay in real time of information from the past (instincts and experience), from the present (state of the organism and environment), and about the future (desires and goals)—an interplay incomprehensively complex yet so totally coherent that, as mentioned above, it appears to us as "just one process": the awareness of our one-and-only self and the feeling of being in total, effortless control of it.

[9]Condensed from Roederer (2005).

This very circumstance presents a big problem to scientists when it comes to understanding the concept of information in a truly objective way. Because "Information is Us," we are so strongly biased that we have the greatest difficulty in detaching ourselves from our own experience with information whenever we try to look at this concept scientifically. Like pornography, "we know it when we see it"—but we cannot easily define it!

In common parlance, information is used as a synonym of many different words: Message, news, data, instruction, announcement, answer, knowledge, characterization, etc. In science however, we usually think of the concept "information" as a statement that answers a pre-formulated question (e.g., what is the mass of this object?) or defines the outcome of some expected alternatives (e.g., the result of a throw of dice). In physics, the alternatives are often the possible states of a physical system (e.g., the many stable vibration modes of a string or organ pipe), and information usually comes as a statement describing the result of a measurement (e.g., "it's the third harmonic"). In communications technology, the alternatives are usually messages from a given, known pool of possibilities (letters of an alphabet, words of a language). We shall use the term *informatics* to designate the study of all aspects of information.[10]

In the 1940 s, Claude Shannon (Shannon and Weaver, 1949) developed what is called the C*lassical Theory of Information* which works with mathematical expressions for concepts like the "novelty value" of one given alternative,[11] the "expected average information gain" in a process that has different possible outcomes[12] or the degree of uncertainty of a set of possible outcomes. And in terms of a quantitative measure of information, everybody knows that the answer to a "yes or no" question or the resolution of any two equally-likely alternatives represents one *bit* (short for "binary unit") of information.

Traditional information theory is not interested in the meaning conveyed by information, the purpose of sending it, the motivation to acquire it, or the potential effect it may have on the recipient. Therefore, it does not give a universal and objective definition of the concept of information applicable to *all* sciences—it is mainly focused on communications, control systems, and computers and quite generally only deals with mathematical expressions involving the *amount*

[10]This term has not gained the popularity in the United States as it has in Europe and elsewhere.

[11]For instance, in a throw of two dice there is only *one* way of getting a total of 12 points (two sixes) whereas there are five different ways of getting 6, so the novelty value of obtaining 12 points must be higher than the novelty value of getting 6. The less probable an alternative, the higher the novelty value when it occurs.

[12]For instance, the expected average information gain of a set of alternatives in which all but one have zero chance to appear, is zero (because we already know what will come out!); the "expected average information gain" of a loaded coin is less than that of a fair coin (because we can guess the outcome with a better chance of success); a fair coin represents maximum uncertainty, therefore maximum information gain (one *bit*) once the outcome is known.

of information contained in a given message. This presents a serious problem when information is used in biology, brain science, sociology—and in music!

So, what is this powerful yet "ethereal" something that resides in CD's, in music scores, is carried by sound waves, is acquired by our senses, triggers our enjoyment—or sits in the genome and directs the construction and performance of an organism? It is *not* the digital pits on the CD, the notes on the pentagram, the air pressure oscillations in a sound wave, the neural activity in the brain, or the chemical bases of the DNA molecule—these all *express* information, but they are not *the* information. Just shuffle them around or change their order ever so slightly and you may get noise, nonsense, or destroy an intended effect or function! On the other hand, information can take many forms and still mean the same—what counts in the end is what information *does*, not how it looks or sounds, how much it is, or what it is made of. Information has always a *purpose*, and the purpose is, without exception, to cause some specific and consistent *change* somewhere, sometime—a change that otherwise would not occur or would happen only by chance.

A fundamental property of information is that it is *the mere shape or pattern of something*—not energy or forces—that triggers this specific change, and can do so consistently over and over again (of course, forces and energy are necessary in order to effect the change, but they are subservient to the purpose of the information in question). However, it is important to emphasize again that the pattern alone or the material it is made of is *not* the information per se, although we are often tempted to think that way. What counts is the unique cause-effect relationship between the pattern and a specific physical response to it. There is no such thing as "information per se" in isolation.

Information always requires a *source* or *sender* (where the original pattern is located or generated) and a *recipient* (where the intended change is supposed to occur). It must be *transmitted* from one to the other. And for the specific change to occur, a specific mechanism must exist and be activated. We usually call this latter action *information processing*. Information can be *stored* and *retrieved*, either in the form of the original pattern, or of some transformation of it. It is always the intended effect that ultimately identifies information. In short, information is the overarching concept that represents the unique *correspondence* between a certain pattern in a source and an intended change in a recipient; this has been called the *pragmatic aspect of information* (Küppers, 1990). Thus defined, information only plays a role in *life systems*, with their unique capability of entertaining *information-driven interactions* with the environment and each other in order to counteract the "normal" (often detrimental) course of physical events. In a purely physical, inanimate, macroscopic world, happenings are driven by forces that are the result of individual interactions between particles and fields; information only appears when a living being, for instance a human, intervenes (scent marks, books, computers, robots, etc.). Looking back at the first paragraphs of this chapter and Table 1.1, it should be evident that *music is information*—information of a very special kind (Sect. 5.8), linking very specific patterns of physical vibration with very specific patterns of neural responses in the brain.

There are many acoustic patterns which do not elicit any specific response beyond the lower auditory areas (i.e., beyond giving the sensation that "something is sounding")—they trigger no specific information-driven interaction and thus carry no pragmatic information. But many acoustic patterns do—for instance speech sounds, provided you know the language involved, or environmental sounds, provided you have knowledge of what is producing them. There is one class of acoustic patterns, however, which may elicit specific neural responses at higher levels in the brain *without* any previous learning requirements, engaging inborn information-processing mechanisms: the superpositions and sequences of periodic tones which make up the *music* in all cultures (Sect. 5.8). The question of what information is involved in music has thus become a question of identifying the relevant higher level neural response patterns it elicits in the human brain and their behavioral consequences. Expressed in terms of informatics, in this book, we will study those acoustic *patterns* that make a sound wave "music," the physical mechanisms by which they are generated and transmitted, and the *correspondence* between the characteristics of such patterns, the *changes* that they cause in the information-processing systems in the listener's ear and brain, and the resulting sensations and feelings.

## 1.7 Informatics and Music: Why Is There Music?

The previous discussion may have irritated some readers. Music, they will say, is "pure aesthetics," a manifestation of the innate and sublime human comprehension of beauty rather than the mere effect of "cold information," embedded in certain air pressure waves, on a complex network of billions of nerve cells. However, as already implied in Sect. 1.5, even aesthetic feelings are related to neural information processing (Sect. 5.6). The characteristic blend of regular, ordered patterns alternated with surprise and uncertainty, common to all sensorial input judged as "aesthetic," may be a manifestation of the curious, yet fundamental drive of humans to exercise their complex neural network with biologically nonessential information-processing operations of changing or alternating complexity. Indeed, artistic expression is perhaps the most human of all intellectual capabilities; whereas it can be argued that cognition and the ability to communicate are only higher in degree in humans than in animals, artistic creativity and appreciation are absolutely unique to human beings.[13]

Indeed, music is ubiquitous in human society, and as historical and archeological evidence shows, it must have been around for a very long time (Fig. 1.1 and Sect. 5.6). Indeed, we can affirm that, in parallel with the statements about the "Information Society" in the first paragraph of the previous section, we are also a "Musical Society"! Musical information-processing distinguishes us humans

[13] Obviously, we do not subscribe to the belief that plants, cows or chickens, when exposed to this or that kind of music, raise their productivity because of artistic appreciation!

from animals; music is everywhere, whether we like it or not; music has become a multi-billion dollar business; we dedicate an appreciable proportion of our personal time to music; and music has always been at the forefront of technology—from crafting the pre-Paleolithic flutes of Fig. 1.1 to inventing the sophisticated wind and tracker machinery of Renaissance organs, to the digital electronics for contemporary sound synthesis and reproduction systems. Yet, there is little doubt that the very first "musical instrument" would have been the human voice (see below) and, from an anthropological point of view, it is reasonable to assume that singing may have predated instrument crafting by hundred thousand years or more.

But do we really know what music is? Like the concept of information, we know what it is, but it is not easy to define! When an animal listens to environmental sounds, it does so in response to an innate drive to become aware of its surroundings (see Sect. 1.5). When we speak, we transmit messages that trigger specific reactions in the recipient's brain, even if the information conveyed represents totally abstract concepts. The purpose of speech is, indeed, that of causing a very specific change in the recipient's *state of knowledge*. In summary, listening confers a survival advantage to all higher animals, and interpreting acoustic information offered by speech is of fundamental biological importance to human beings. Now, what information does music transmit? As mentioned in Sect. 1.2, in nearly all cultures, music consists of organized, structured, rhythmic successions and superpositions of tones, selected from a very limited repertoire of the discrete pitches of some musical scale. Typical environmental sounds offer no equivalent structures,[14] and imitating environmental sounds hardly ever was a prime force driving the development of a musical culture except perhaps at its earliest stages (e.g., for hunting strategies). Yet, if music does not seem to convey biologically relevant information, *why* does it affect us? Crying babies calm down at the simple musical tones of their mothers' song; beautiful passages can give us goose bumps, awful ones can move us to rage. A military march may elicit pride in one's country or loyal submission to a cruel dictator. Why do these things happen? *Why is there music*?

As we shall discuss in Sect. 5.8, music may be a natural co-product of the evolution of *human language*. In this evolution, which undoubtedly was an essential factor in the development of homo sapiens, a neural network emerged capable of executing the extraordinarily complex acoustic information processing, analysis, storage, and retrieval operations necessary for phonetic/phonemic recognition, voice identification, comprehension of speech, and cognition of the content. Language endowed humans with a mechanism that increased immeasurably their memory and associated information-handling capacity by allowing the reduction of complex images of environmental scenes, objects, and their causal

[14]Bird song is music to *us*, but to the birds it is very concrete and "down-to-earth" information such as "This territory is claimed," or "This male is looking for a mate!"

interrelations to short symbolic representations. In the course of this evolution, a most remarkable division of tasks between the two cerebral hemispheres developed (Sect. 5.7). The left hemisphere (in about 97% of all persons) mainly executes short-term temporal operations such as are required for verbal comprehension and other short-term sequencing operations of crucial importance to thinking. The right hemisphere cooperates with the performance of spatial integrations and long-term time representations. Holistic operations of the right hemisphere intervene in pictorial imaging—and music perception. Indeed, as we shall see throughout this book, music perception does involve holistic tasks ranging from the analysis of spatial excitation patterns along the auditory receptor organ, caused by musical tones and tone superpositions, to the analysis of long-term time patterns of melodic lines.

Why do human beings respond *emotionally* to music, why can music elicit squirts of endorphins that stimulate the pleasure centers of the human brain (Sect. 5.6), and why are we motivated to *create* music? In other words, what is, or was, the biological survival value of music that led to music loving in the course of human evolution? What was the biological *purpose* of crafting the early instruments shown in Fig. 1.1? It is not difficult to speculate about the origin of the motivation to perform certain actions that have no immediate biological purpose, such as climbing a mountain (instinct to explore), playing soccer (training in skilled movement, instinct to lead or win), or enjoying the view of a sunset (imminence of rest, expectation of the shelter of darkness). But what was the genetic advantage to early hominids of stringing together "abstract" musical tones and forms? Of course, this question must be considered part of a more encompassing question related to the emergence of aesthetic motivation, affective response, and creativity.

The motivation to listen to music may well be the result of an inborn drive, with consequent limbic reward, to train at an early age in the highly sophisticated auditory analysis operations expected for speech perception and language (Sect. 5.8)—not unlike an animal's play being the manifestation of an inborn motivation to develop or improve skilled movements required for preying and self-defense. Infants born without a drive to listen attentively, or born to mothers lacking a drive to vocalize simple musical sounds during early mother-child social bonding, would have had a decisive communications disadvantage for survival in the early human environment. Isn't the universality of the first speech vocalizations of babies and lullabies of mothers all over the world a convincing argument? Language perception develops spontaneously, without effort and so does musical form perception—practically at the same time. Amazing experiments presently being conducted with months-old infants (Fig. 1.2) are providing supporting arguments in favor of an inborn predisposition for musical message processing (Sect. 5.8).

There is no unanimity regarding the language-music co-evolution, though. Most of the opposition comes from anthropologists and behavioral psychologists. However, the large number of basic universal musical characteristics that can be

FIGURE 1.2 Infant "wired up" with 126 EEG electrodes to determine the effects on neural activity of different sounds.
(Source: Laurel Trainor, McMaster Institute for Music and the Mind, McMaster University, Canada; permission gratefully acknowledged.)

linked to specific operational modes of the sensory and cerebral information processing systems makes it highly unlikely that "music is just a happy evolutionary accident," as some opponents claim.

After the initial phase of the evolution of music, it is logical to assume that the major or minor degree of complexity of a tone-message identification, the degree of success, and related limbic rewards in prediction-making operations carried out by the brain to expedite this identification process, as well as the type of associations evoked by comparison with stored information on previous experiences, would all contribute to the "cause" of musical sensations evoked by a given musical message. If this is true, it would be obvious that both innate neural mechanisms (primary processing operations) *and* cultural conditioning (stored messages and learned processing operations) determine our affective response to music (Sects. 5.6 and 5.8). And, finally, this cultural conditioning may be reinforced by the effectiveness of music in achieving behavioral coherence in masses of people, as demonstrated by its age-old role in superstitious and sexual rites, religion, ideological proselytism, military arousal, even anti-social behavior.

Nearly 90% of this first chapter dealt with perception, yet half of this book is dedicated to the physics of music: sound waves, tone generation in musical instruments, and sound propagation. The reason for this thematic imbalance in the introduction is simple. Physical acoustics is an old science (Cohen and Drabkin, 1948)—it began 2600 years ago with Pythagoras! Many of its current aspects, from the basics of how musical instruments work, to the intricacies of electroacoustics are known to many lay people. Psychoacoustics, instead, is

new—barely 160 years young—and the study of music perception has only now, in the last few decades, been developing vigorously;[15] the results are only now beginning to reach the nonspecialist and the general public. I just wanted to place the science of music in the proper perspective—the perspective of the *music of science*!

[15]The author had the privilege of actively promoting this multidisciplinary development, as testified by the world-renowned music psychologist Diana Deutsch who writes in her guest editorial for the 20th anniversary of the international journal *Music Perception* (Deutsch, 2004):

...*"A series of interdisciplinary workshops on the Physical and Neuropsychological Foundations of Music organized by Juan Roederer were held in Ossiach, Austria* [between 1973 and 1985 during the Carinthian Summer Festivals]*, and it was at these workshops that many of us learned for the first time, and with great excitement, about studies on music that were being carried out in each other's fields. It became clear at these exhilarating workshops that an interdisciplinary study of music, with input from music theorists, composers, psychologists, linguists, neuroscientists, computer scientists, and others, was not only viable but even necessary to advance the understanding of music"*.

# 2
# Sound Vibrations, Pure Tones, and the Perception of Pitch

*" With over a million essential moving parts, the auditory receptor organ, or cochlea, is the most complex mechanical apparatus in the human body"*

A. J. Hudspeth (reference Hudspeth, 1985).

We hear a sound when the eardrum is set into a characteristic type of motion called *vibration*. This vibration is caused by small pressure oscillations of the air in the auditory canal associated to an incoming sound wave. In this chapter we shall first discuss the fundamentals of periodic vibratory motion in general and then focus on the effects of eardrum vibrations on our sensation of hearing. We shall not worry at this stage about *how* the eardrum is set into motion. To that effect, let us imagine that we put on headphones and listen to tones generated therein. In the lower frequency range, the eardrums will very closely follow the vibrations of the headphone diaphragms. This approach of introducing the subject is somewhat unorthodox. But it will enable us to plunge straight into the study of some of the key concepts associated with sound vibration and sound perception without spending first a long time on sound waves and sound generation. From the practical point of view, this approach has one drawback: The experiments that we shall present and analyze in this chapter necessarily require electronic generation of sound rather than natural production with real musical instruments. Whenever possible, however, we shall indicate how a given experiment could be performed with real instruments.

## 2.1 Motion and Vibration

Motion means *change of position* of a given body with respect to some reference body. If the moving body is very small with respect to the reference body, or with respect to the dimensions of the spatial domain covered in its motion, so that its shape is practically irrelevant, the problem is reduced to the description of the motion of a *point* in space. This is why such a small body is often called a material point or a particle. On the other hand, if the body is not small, but if from the particular circumstances we know beforehand that all points of the body are confined to move along straight lines parallel to each other ("rectilinear translation"), it, too, will suffice to specify the motion of just *one* given point of the body. This is a "one-dimensional" case of motion, and the position of the given


J.G. Roederer, *The Physics and Psychophysics of Music*,
DOI: 10.1007/978-0-387-09474-8_2, 

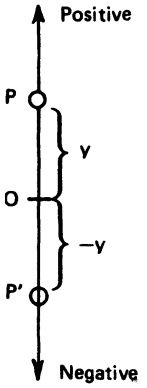



FIGURE 2.1 Instantaneous positions of a point moving on a straight line. *y*: coordinate; *O*: fixed reference point.

point of the body (and, hence, that of the whole body) is completely specified by just *one* number - the distance to a fixed reference point.

In this book we shall only deal with one-dimensional motions. Let us assume that our material point moves along a vertical line (Fig. 2.1). We shall designate the reference point on that line with the letter *O*. Any fixed point can serve as a reference point, although, for convenience, we sometimes may select a particular one (such as the equilibrium position for a given oscillatory motion). We indicate the position of a material point *P* by the distance *y* to the reference point *O* (Fig. 2.1). *y* is also called the *displacement* of *P* with respect to *O*, or the *coordinate* of *P*. We must use both positive and negative numbers to distinguish between the two sides with respect to *O*.[1]

The material point *P* is in motion with respect to *O* when its position *y* changes with time. We shall indicate time by the letter *t*. It is measured with a clock—and it, too, requires that we specify a "reference" instant of time at $t = 0$. Motion can be represented mathematically in two ways: analytically, using so-called functional relationships, and geometrically, using a graphic representation. We shall only use the geometric method. To represent a one-dimensional motion graphically, we introduce two axes perpendicular to each other, one representing the time *t*, the other the coordinate *y* (Fig. 2.2). For both we have to indicate clearly the *scale*, that is, the unit intervals (of time and displacement, respectively). A motion can be represented by plotting for each instant of time *t*, the distance *y* at which the particle is momentarily located. Each point of the ensuing curve, such as $S_1$ (Fig. 2.2), tells us that at time $t = t_1$ the particle *P* is at a distance $y_1$ from *O*, that

[1] In science, the *metric system* is used to measure distances. The unit of length is the meter (1 m = 3.28 feet); several decimal submultiples (e.g., the centimeter = 0.01 m = 0.394 inch, or the millimeter = 0.001 m) and multiples (e.g., the kilometer = 1000 m = 0.625 mile) are also used.

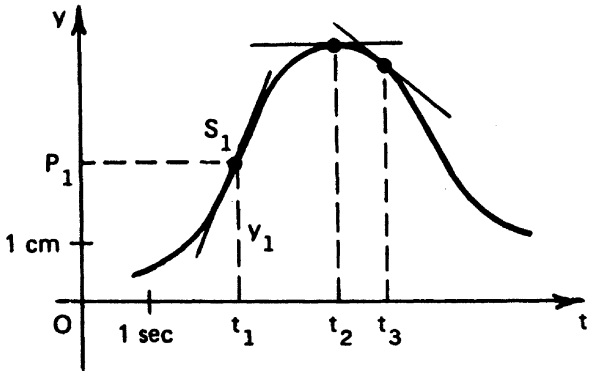



FIGURE 2.2 Graphic representation of the motion of a point. The heavy curve is *not* the trajectory! Rather, it determines the positions of the point at different times $t$, as it moves along a *straight* line (coordinates represented by the $y$ axis).

is, at position $P_1$. Note very carefully that in this graph the material point does *not* move along the curve of points S! This curve is just a human-made "aid" that helps us to find the position $y$ along the vertical axis of the particle at any time $t$.

The graph shown in Fig. 2.2 also gives information on the velocity of the material point, i.e., the rate of change of its position. This is determined by the *slope* of the curve in the graph: at $t_1$ the particle is moving at a certain rate upward, at $t_3$ it is moving downward at a slower rate. At $t_2$ it is momentarily at rest, reversing its direction.

There is a certain class of motions in which the material point follows a pattern in time that is repeated over and over again. This is called *periodic motion* or *vibration*. It is the type of motion of greatest importance to physics of music. In order to have a truly periodic motion, a body not only has to repeatedly come back to the same position, it has to do so at exactly equal intervals of time and exactly repeat the same type of motion in between. The interval of time after which the pattern of motion is repeated is called the *period* (Fig. 2.3). We denote it by the Greek letter tau ($\tau$). During one period, the motion may be very simple (Fig. 2.3(a)) or rather complicated (Fig. 2.3(b)).[2] The elementary pattern of motion that occurs during one period and that is repeated over and over again is called a *cycle*.

There are mechanical and electronic devices that can automatically plot the graph of a periodic motion. In a *chart recorder*, the pen reproduces in the $y$ direction the periodic motion that is to be described, while it is writing on a strip of paper that is moving perpendicularly to the $y$ axis at a constant speed. Since we

[2]It is a good exercise for understanding graphs like Fig. 2.2 or 2.3 to mimic the depicted motion with an equivalent up-and-down movement of your hand representing the changing distances (along the $y$ axis) as time goes on ($t$ axis).

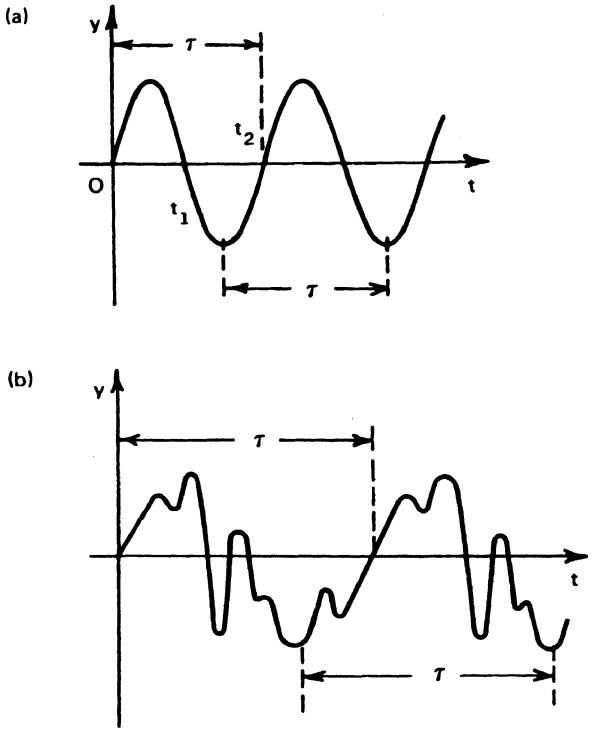



FIGURE 2.3 Graphic representation of (a) a simple periodic motion; (b) a complex periodic motion ($\tau$: period).

know this speed, we can assign a *time* scale to the axis along the paper strip. The curve obtained is the graphic representation of the motion. This method is not practical for the registration of acoustic vibrations though. They have such short periods that it would be impossible to displace a pen fast enough to reproduce this type of vibration. An electronic device called an *oscilloscope* serves the purpose. In essence, it consists of a very narrow beam of electrons (elementary particles of negative electric charge) that impinge on a TV screen giving a clearly visible light spot. This beam can be deflected both in the vertical and horizontal directions. The vertical motion is controlled by a signal proportional to the vibration which we want to display (for instance, the vibration of the diaphragm of a microphone). The horizontal motion is a continuous sweep to the right with constant speed, equivalent to the motion of the paper strip in a chart recorder, thus representing a time scale. The luminous point on the screen thus describes the graph of the motion during one sweep. If the image of the luminous point is retained long enough, it appears as a continuous curve on the screen. Since the screen is only of limited size, the horizontal motion is instantaneously reset to the origin whenever the beam reaches the right edge of the screen, and the sweep starts again. To

represent a periodic motion, the sweep must be synchronized with the period $\tau$ or one of its multiples.

## 2.2 Simple Harmonic Motion

The question now arises as to which is the "*simplest*" kind of periodic motion. There are many familiar examples: the back and forth oscillation of a pendulum, the up and down motion of a spring, the oscillations of molecules, etc. Their motions have something important in common: they all can be represented as the projection of a uniform circular motion onto one diameter of the circle (Fig. 2.4).[3] When point $R$ turns around uniformly (with period or, i.e., once every or seconds) the projection point $P$ moves up and down along the $y$ axis with what is called a *simple harmonic motion* (see graph at right-hand side of Fig. 2.4). This is also called a *sinusoidal motion* (because $y$ can be represented analytically by a trigonometric function called sine).

Note that a simple harmonic motion represents a vibration that is symmetric with respect to point $O$, which is called the equilibrium position. The maximum displacement $A$ (either up or down) is called *amplitude.* $\tau$ is the *period* of the harmonic motion. There is one more parameter describing simple harmonic motion that is a little more difficult to understand. Consider Fig. 2.4: at the initial instant $t = 0$, the particle (projection of the rotating point $R$) is located at position $P$. We may now envisage a second case of harmonic motion with the *same* period $t$ and the *same* amplitude $A$, but in which the particle starts from a *different position* $Q$ (Fig. 2.5). The resulting motion obviously will be different, not in form or type, but in relative "timing." Indeed, as seen in Fig. 2.5, both particles will pass through a given position (e.g., the origin $O$) at different times ($t_1$, $t_2$). Conversely, both particles will in general be at different positions at one given time (e.g., $P$ and $Q$ at $t = 0$). If we again imagine the motion of the second particle

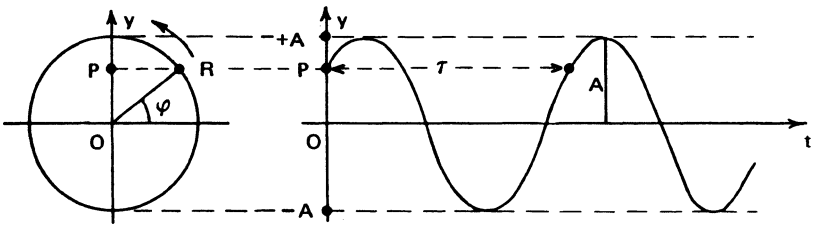



FIGURE 2.4 Simple harmonic motion or *sinusoidal* motion (represented on right graph) obtained as the projection of a point in uniform circular motion onto a diameter ($\varphi$: phase; $A$: amplitude; $\tau$: period).

[3]Note carefully that the construction at the left-hand side of Fig. 2.4 is *auxiliary*; the only real motion is the periodic up-and-down motion of the particle $P$ along the $y$ axis.

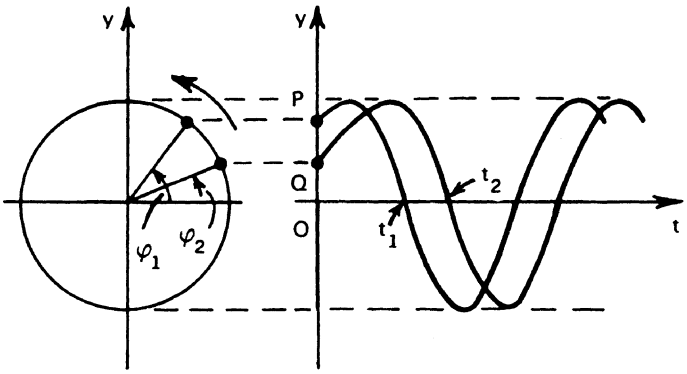



FIGURE 2.5 Graphic representation of the harmonic motion of two points with the same amplitude and frequency but different phases $\varphi_1$ and $\varphi_2$.

$Q$ as the projection of a uniform circular motion (Fig. 2.5), we realize that both cases pertain to different *angular positions* $\varphi_1$, $\varphi_2$ of the associated rotating points on the circle. The angle $\varphi$ is called the *phase* of a simple harmonic motion; the difference $\varphi_1 - \varphi_2$ (Fig. 2.5), which remains constant in this example, is called the *phase difference* between the two harmonic motions.[4]

In summary, a given "pure," or harmonic, vibration is specified by the values of three parameters: the *period* $\tau$, the *amplitude* $A$, and the *phase* $\varphi$ (Fig. 2.4). They all, but especially the first two, play a key role in the perception of musical sounds.

Simple harmonic motions occur practically everywhere in the universe: Vibrations of the constituents of the atoms, of the atoms as a whole in a crystal, of elastic bodies, etc., can all be described in first approximation as simple harmonic motions. But there is another even more powerful reason for considering simple harmonic motion as the most basic periodic motion of all: it can be shown mathematically that *any kind of periodic motion, however complicated, can be described as a sum of simple harmonic vibrations*. We shall deal with this fundamental property later on in detail (Chap. 4). It is indeed of capital importance to music.

## 2.3 Acoustic Vibrations and Pure Tone Sensations

When the eardrum is set into periodic motion, its mechanical vibrations are converted in the inner ear into electrical nerve impulses that are signaled to the brain and interpreted as sound, provided the period and the amplitude of the vibrations

[4]Use your two hands to mimic the two up-and-down motions represented on the right-hand side of Fig. 2.5! Do the same for different phase differences, for example, 180° (opposite phases), 0° (in-phase), 90°, etc.

lie within certain limits. In general, the ear is an extremely sensitive device: vibration amplitudes of the eardrum as small as $10^{-7}$ cm can be detected, and so can vibrations with periods as short as $7 \times 10^{-5}$ s.[5]

We now introduce a quantity that is used more frequently than the period $\tau$, called the *frequency*:

$$f = 1/\tau \tag{2.1}$$

Physically, $f$ represents the number of repetitions of the vibration pattern, or cycles, in the unit interval of time. The reason for preferring $f$ to $\tau$ is that the frequency increases when our sensation of "tone height" or pitch increases. If $\tau$ is given in seconds, $f$ is expressed in *cycles per second*. This unit is called hertz (Hz), in honor of Heinrich Hertz, a famous German physicist. Air pressure vibrations in the interval 20–15,000 Hz are sensed as sound by a normal person. Both the lower and particularly the upper limit depend on tone loudness and vary considerably from person to person as well as with age.

When a sound causes a simple harmonic motion of the eardrum with constant characteristics (frequency, amplitude, phase), we hear what is called a *pure tone*. A pure tone sounds dull, and music is not made up of single pure tones. However, as stated in the introduction to this chapter, for a better understanding of complex sounds, it is advisable to deal first with pure or simple tones only. Pure tones have to be generated with electronic oscillators; there is no musical instrument that produces them (and even for electronically generated pure tones, there is no guarantee that they will be pure when they actually reach the ear). In any case, since the flute is the instrument whose sound approximates that of a pure, sinusoidal tone more than any other instrument, especially in the upper register, several (but not all) of the experiments referred to in this chapter could indeed be performed at home using one or, as a matter of fact, two flutes—played by experts, though!

When we listen to a pure tone whose frequency and amplitude can be changed at will, we verify a correspondence between *pitch* and *frequency* and between *loudness* and *amplitude.* In this chapter, we only consider pitch.

The simple harmonic oscillations of the eardrum are transmitted by a chain of three tiny bones in the middle ear called "hammer," "anvil," and "stirrup" (or, in more erudite parlance, malleus, incus, and stapes, which means exactly the same thing in Latin) to the entrance ("oval window") of the inner ear proper (Fig. 2.6(a)). The marble-sized *cochlea* is a tunnel spiraling as a snail shell through the human temporal bone. This cavity, shown in Fig. 2.6(b) in a highly simplified, stretched out version, is partitioned into two channels, the scala vestibuli and scala timpani, filled with an incompressible fluid, the *perilymph* (a direct filtrate of the cerebrospinal fluid). Both channels behave as one hydrodynamic system,

[5] In this book we shall use the exponent notation: $10^{+n} = 100\leftarrow n$ zeros $\rightarrow 00$. $10^{-n}$ is just $1/10^{+n}$, that is, a decimal fraction given by one unit in the $n$th decimal position.

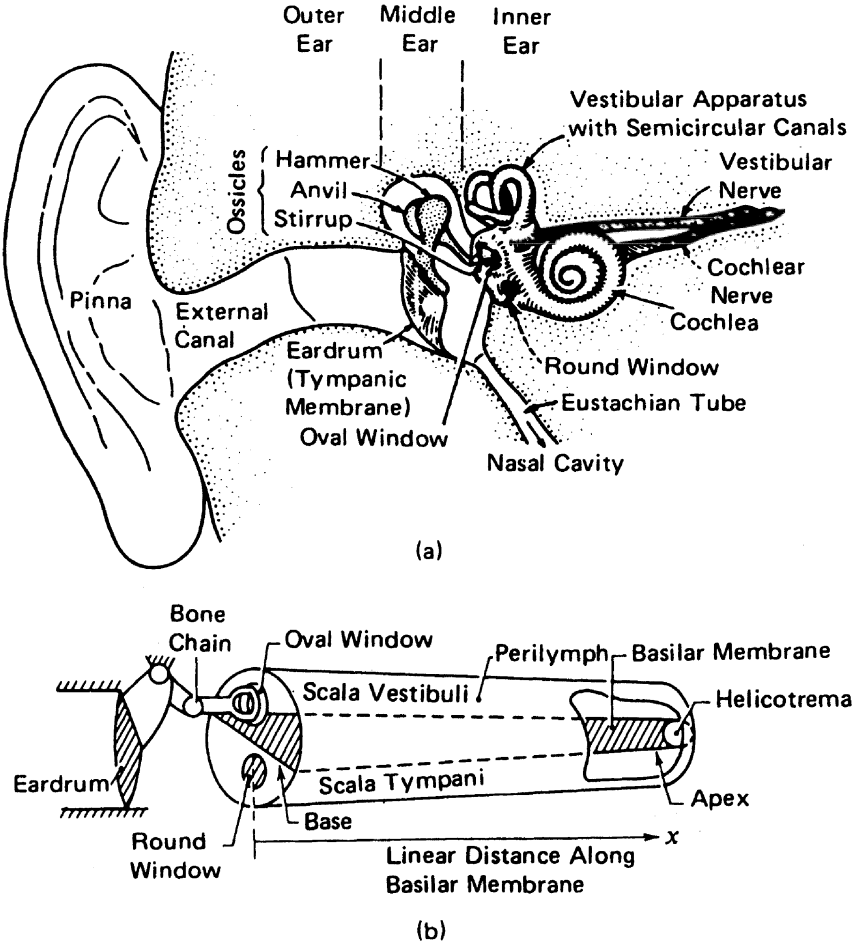



FIGURE 2.6 (a) Schematic view of the ear (Flanagan, 1972; Fig. 4) (not in scale); (b) the cochlea shown stretched out (highly simplified).

because they are connected at the far end, or apex, by a small hole in the partition called *helicotrema;* the lower section is sealed off with an elastic membrane at the "round window" (Fig. 2.6(b)). The partition separating both scalae is in itself a highly structured duct of triangular cross section (also called scala media; Fig. 2.7(a)), filled with another fluid, the *endolymph*. Its boundaries are the *basilar membrane* which holds the sensory organ proper (*organ of Corti*), Reissner's membrane, which serves to separate endolymph from perilymph, and the rigid lateral wall of the cochlea.

The elasticity of the basilar membrane determines the cochlea's basic hydromechanical properties. In the human adult, the membrane is about 34 mm long from the base (the input end) to the apex; because of its gradual change

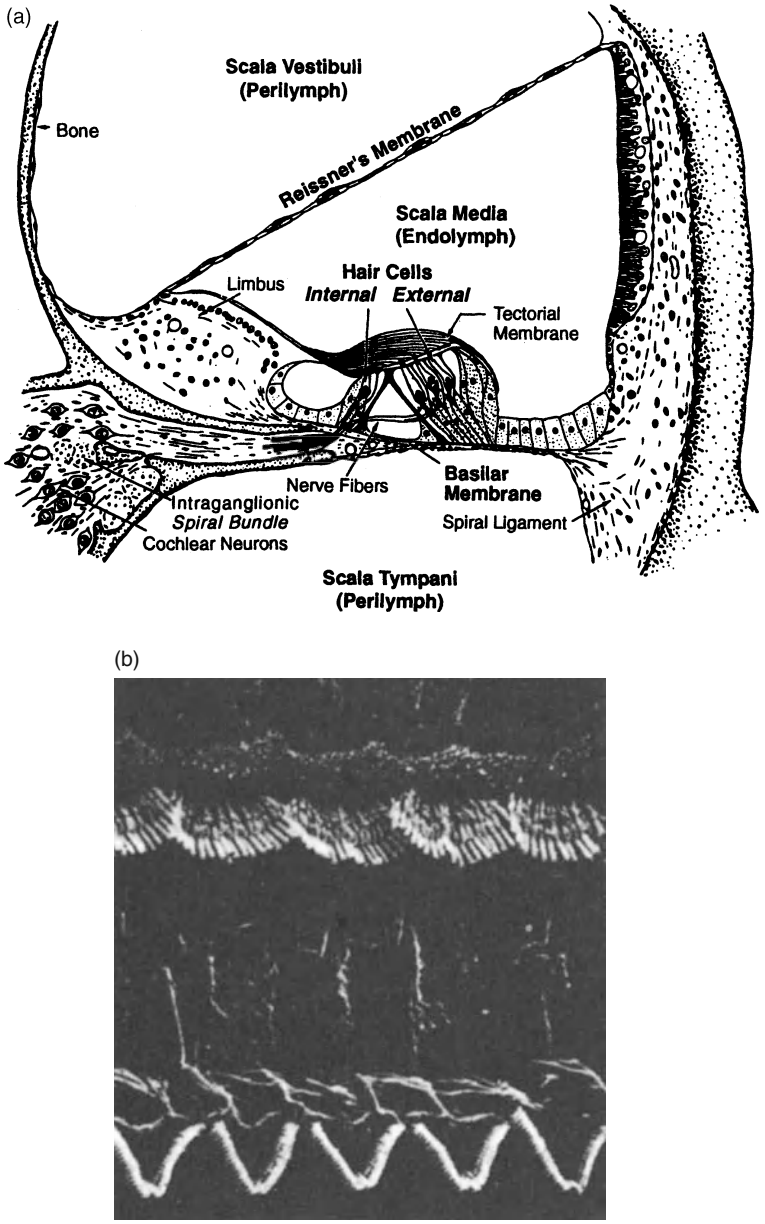



FIGURE 2.7 (a) Cross section of the organ of Corti (after Davis (1962)). (b) Scanning electron micrograph (Bredberg et al., 1970) of the stereocilia of inner row (*top*) and outer row (*bottom*—only one of the three rows is shown) hair cells on the basilar membrane of a guinea pig. (These animals, as well as chinchillas and cats have peripheral acoustic systems very similar to humans and are the laboratory animals most frequently used in hearing research.)

in width and thickness, there is a 10,000-fold decrease in stiffness from base to apex, which gives the basilar membrane its fundamental frequency analyzing function. Vibrations transmitted by the bone chain to the oval window are converted into pressure oscillations of the perilymph fluid in the scala vestibuli. The ensuing pressure differences across the cochlear partition between the two scalae flex the basilar membrane up and down setting it into motion like a waving flag. As this wave travels toward the apex, its amplitude increases to a maximum at a given place, which depends on the input frequency, and then dies down, very quickly toward the apex. On the other hand, the membrane also twists transversally in a very complicated way, which is believed to play a fundamental role in the hydromechanical stimulation of the approximately 16,000 receptor units, called *hair cells*, arranged in one "inner" row and three "outer" rows along the basilar membrane. These cells pick up the motions of the latter and impart signals to the nerve cells, or *neurons*, that are in contact with them. The name "hair cell" comes from the fact that at its top, there is a bundle of 20–300 tiny processes called *stereocilia* (Fig. 2.7(b)) protruding into the endolymphatic fluid, whose deflection triggers a chain of electrochemical processes in the hair cell and its surroundings that culminate in the generation of electrical signals in the acoustic nerve. The *tectorial membrane* is a gelatinous tissue suspended in the endolymph above the organ of Corti (Fig. 2.7(a)), into which the cilia of outer hair cells are inserted; it plays a key role in stimulating and receiving the motile action of the latter. We shall return to the cochlear function in more detail in Sects. 2.8 and 3.6.

The remarkable fact is that for a *pure tone* of given frequency, the maximum basilar membrane oscillations occur only in a limited region of the membrane, *whose position depends on the frequency of the tone*. In other words, for each frequency there is a region of maximum stimulation, or "resonance region," on the basilar membrane. The lower the frequency of the tone, the closer to the apex (Fig. 2.6(b)) lies the region of activated hair cells (where the membrane is most flexible). The higher the frequency, the closer to the entrance (oval window) it is located (where the membrane is stiffest). *The spatial position $x$ along the basilar membrane* (Fig. 2.6(b)) *of the responding hair cells and associated neurons determines the primary sensation of pitch* (also called spectral pitch). A change in frequency of the pure tone causes a shift of the position of the activated region; this shift is then interpreted as a change in pitch. We say that the primary information on tone frequency is encoded by the sensorial organ on the basilar membrane in the form of *spatial location* of the activated neurons. Depending on which group of neural fibers is activated, the pitch will appear to us as low or high.

Figure 2.8 shows how the position $x$ (measured from the base, Fig. 2.6(b)) of the region of maximum sensitivity varies with the frequency of a pure, sinusoidal tone, for an average adult person (von Békésy, 1960). Several important conclusions can be drawn. First of all, note that the musically most important range of frequencies (approximately 20–4000 Hz) covers roughly two-thirds of the extension of the basilar membrane (12–35 mm from the base). The large remaining portion of the frequency scale (4000–16,000 Hz, not shown beyond 5000 Hz in Fig. 2.8) is squeezed into the remaining one-third. Second, notice the significant

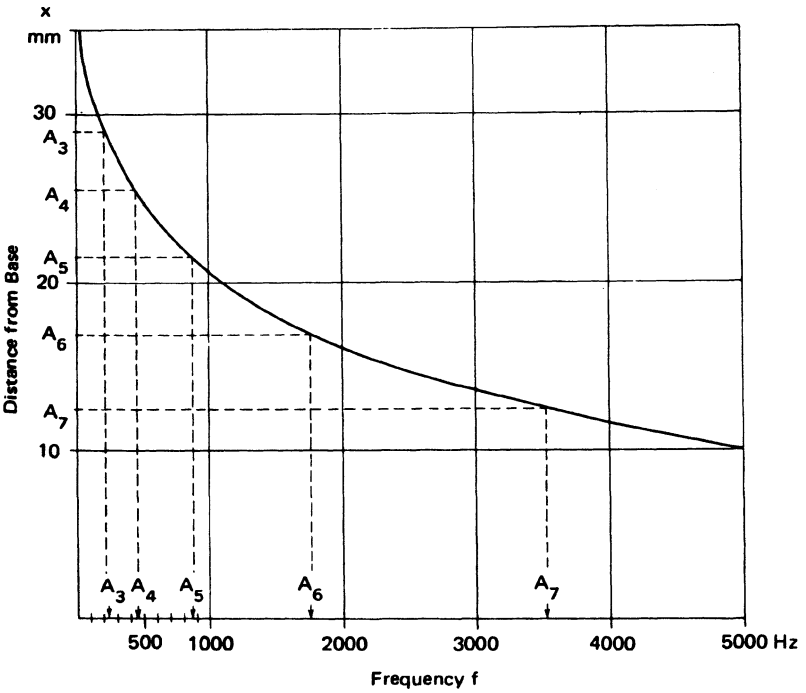



FIGURE 2.8 Position of the resonance maximum on the basilar membrane (after von Békésy (1960)) for a pure tone of frequency $f$ (linear scales).

fact that, whenever the frequency of a tone is doubled, that is, the pitch jumps one octave, the corresponding resonance region is displaced by a roughly constant amount of 3.5–4 mm, no matter whether this frequency jump is from 220 to 440 Hz, from 1760 to 3520 Hz, or, as a matter of fact, from 5000 to 10,000 Hz. In general, whenever the frequency $f$ is multiplied by a given factor, the position $x$ of the resonance region is not multiplied but simply shifted a certain amount. In other words, it is frequency *ratios*, not their differences, which determine the displacement of the resonance region along the basilar membrane. A relationship of this kind is called "logarithmic" (Sect. 3.4).

The above results come from physiological measurements performed on dead (but well preserved) animals (von Békésy, 1960). Today, such measurements can be performed on living cochleas with microscopic laser beams or through the Mössbauer effect in which a minute mass of radioactive substance (Cobalt 57) is implanted on the basilar membrane. The tiny displacements of the membrane can then be detected indirectly by measuring the frequency shift (Doppler effect) of the reflected laser beam or the gamma rays emitted by the radioactive substance. More on this in Sect. 3.6.

We now consider the psychophysical magnitude pitch, associated to a pure tone of frequency $f$. In Sect. 1.4, we mentioned that a psychophysical magnitude

cannot be measured in the same quantitative manner as a physical magnitude like frequency. Only a certain *order* can be established by the experiencing individual between two sensations of the same kind presented in immediate succession. For some sensations, quantitative estimates are possible only after the brain somehow has been trained to perform the necessary operations (e.g., a child learning to estimate the size of the objects she sees).

According to the description given above, the primary function of the inner ear (cochlea) is to convert a vibration pattern in time (the in-out motion of the eardrum) into a vibration pattern in space (the up-down motion along the basilar membrane), and this, in turn, into a spatial pattern of neural activity (the distribution of electrical signals across the acoustic nerve).

Let us consider an individual's ability to establish a relative order of pitch when two pure tones (of the same intensity) are presented one after the other. There is a natural limit: when the difference in frequency between the two tones is small, below a certain value, both tones are judged as having "the same pitch." This is true with the order judgments for all psychophysical magnitudes: whenever the variation of an original physical stimulus lies within a certain *difference limen* (DL) or "just noticeable difference" (JND, a term used in the older literature), the associated sensation is judged as remaining "the same"; as soon as the variation exceeds the DL, a change in sensation is detected. Notice that the DL relates to a *physical* magnitude (the stimulus) and is expressed by a number.[6]

The degree of sensitivity of the primary pitch perception mechanism to frequency changes, or *frequency resolution* capability, depends on the frequency, intensity, and duration of the tone in question—and on the suddenness of the frequency change. It varies greatly from person to person, is a function of musical training, and, unfortunately, *depends considerably on the method of measurement employed*. Figure 2.9 shows the average DL in frequency for pure tones of constant intensity (80 decibels, Sect. 3.4), whose frequency was slowly and continuously modulated up and down (Zwicker, Flottorp, and Stevens, 1957). This graph shows for instance, that for a tone of 2000 Hz a change of 10 Hz—that is, of only 0.5%—can already be detected. This is a very small fraction of a semitone! *Sudden* changes in frequency are detected with even lower DL—up to 30 times smaller than the values shown in Fig. 2.9 (Rakowski, 1971). Frequency resolution becomes worse at low frequencies (e.g., 3% at 100 Hz in Fig. 2.9). It also decreases with decreasing tone duration once the latter falls below about one-tenth of a second. In contrast, frequency resolution is roughly independent

[6]This quantity is measurable only in statistical terms in a so-called *forced-choice psychometric experiment*: when the frequency difference between the two tones is small and the listeners are asked to identify the one with higher pitch, they only can guess and their scores will be 50% correct. When the frequency difference is large enough, 100% of the answers will be right. 75% of correct answers lie halfway between guessing and perfect response, and is then taken as the DL in frequency (Hartmann, 1996).

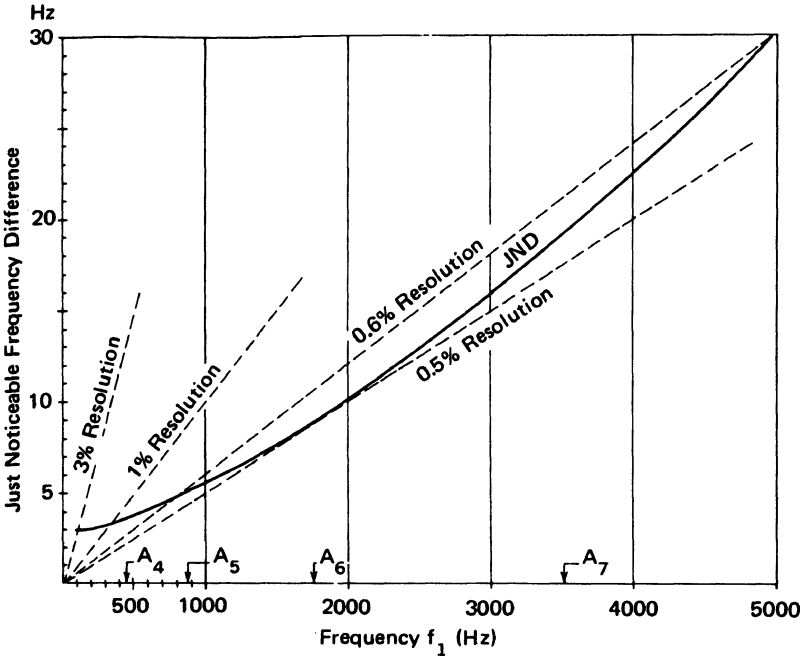



FIGURE 2.9 Frequency difference limen (or "just noticeable difference") for a pure tone of frequency $f_1$ (linear scales), as determined with a slowly frequency-modulated signal (after Zwicker, Flottorp, and Stevens (1957)).

of amplitude (loudness). For a detailed discussion of DL see Zwicker and Fastl (1999), Hartmann (2005), and Terhardt (1998).

Since the inception of psychophysics, psychologists have been tempted to consider the minimum perceptible change in sensation caused by a just noticeable difference of stimulus as the natural unit with which to measure the corresponding psychophysical magnitude. The minimum perceptible change in pitch has been used to construct a "subjective scale of pitch" (Stevens, Volkmann, and Newman, 1937). However, as we shall see later, because the *octave* plays such an overriding role as a natural pitch interval, and because all musical scales developed in total independence of the attempts to establish a subjective scale of pitch, the latter has not found a direct practical application in music (see however, Sect. 5.4).

## 2.4 Superposition of Pure Tones: First-Order Beats and the Critical Band

We said before that single pure tones sound very dull. Things become a little livelier as soon as we superpose two pure tones by sounding them together. In this section, we shall analyze the fundamental characteristics of the superposition

of two pure tones. We will meet some very fundamental concepts of physics of music and psychoacoustics.

There are two kinds of superposition effects, depending on where they are processed in the listener's auditory system. If the processing is *mechanical*, occurring in the cochlear fluid and along the basilar membrane, we call them "first-order superposition effects," mainly because they are clearly distinguishable and of fairly easy access to psychoacoustic experimentation. "Second-order" superposition effects are the result of *neural* processing and are more difficult to detect, describe, and measure unambiguously. In this section, we shall focus on first-order effects only.

Let us first discuss the physical meaning of "superposition of sound." The eardrum moves in and out commanded by the pressure variations of the air in the auditory canal. If it is ordered to oscillate with a pure harmonic motion of given frequency and amplitude, we hear a pure tone of certain pitch and loudness. If now, *two* pure tones of different characteristics are sounded together (e.g., by listening to two independent sources at the same time), the eardrum reacts as if it were executing at the same time two independent commands, one given by each pure tone. The resulting motion is the sum of the individual motions that would occur if each pure sound were present alone, in the absence of the other. Not only does the eardrum behave this way, but also the medium and all other vibrating components (this, however, is not true if the amplitudes are very large). This effect is called a *linear superposition* of two vibrations. Linear superposition of two vibrations is a technical term that means "peaceful coexistence": one component vibration does not perturb the affairs of the other, and the resulting superposition simply follows the dictations from each component simultaneously. In a *nonlinear* superposition, the dictation from one component would depend on whatever the other one has to say, and vice versa.

We start our discussion with the analysis of the superposition of two simple harmonic motions with *equal frequency and equal phase* (zero phase difference, Sect. 2.2). It can be shown graphically (Fig. 2.10) and also analytically that, in this case, we again obtain a simple harmonic motion of the same frequency, the same phase, but with amplitude which is the sum of the amplitudes of the two component vibrations. If the two component oscillations of given frequency have *different phase*, their superposition still turns out to be a simple harmonic motion of the same frequency, but the amplitude will *not* be given anymore by the sum of the component amplitudes. In particular, if the amplitudes of the component vibrations are equal and their phase difference $\varphi$ is 180°, both oscillations will annihilate each other and no sound at all will be heard. This is called *destructive interference* and plays an important role in room acoustics. In summary, when two pure tones of equal frequency reach the eardrum, we perceive only *one* tone of given pitch (corresponding to the frequency of the component tones) and loudness (controlled by the amplitudes of the superposed tones *and* their phase difference).

We now consider the superposition of two simple tones of the same *amplitude* but with *slightly different frequencies,* $f_1$ and $f_2 = f_1 + \Delta f$. The frequency difference

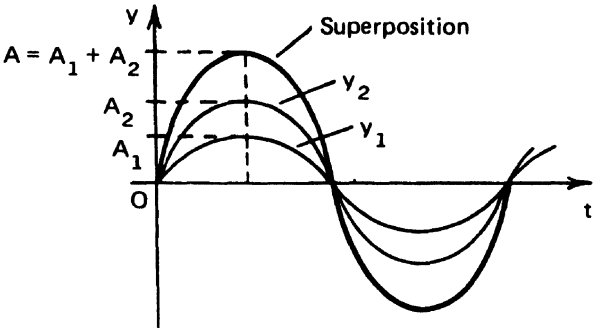



FIGURE 2.10 Superposition of two sinusoidal vibrations of equal phase and frequency.

$\Delta f$ is small; let us assume that it is positive (the tone corresponding to $f_2$ is slightly sharper than that of $f_1$). The vibration pattern of the eardrum will be given by the sum of the patterns of each component tone (Fig. 2.11). The result of the superposition (heavy curve) is an oscillation of period and frequency *intermediate* between $f_1$ and $f_2$, and of slowly modulated amplitude. Note, in this figure, the slowly changing phase difference between the component tones $y_1$ and $y_2$: they start in phase (0° phase difference, as in Fig. 2.10) at the instant $t = 0$, then $y_2$ starts leading in phase (ahead of $y_1$ until both are completely out of phase (180° phase difference) at the instant corresponding to C. The phase difference keeps increasing until it reaches $360° = 0°$ at instant $\tau_B$. This continuous, slow phase shift is responsible for the changing amplitude of the resultant oscillation: the broken curves in Fig. 2.11 represent the *amplitude envelope* of the resulting vibration (see also Fig. 2.17(A)).

What is the resulting tone sensation in this case? First of all, note very carefully that the eardrum will follow the oscillation as prescribed by the *heavy* curve in Fig. 2.11. The eardrum "does not know" and "does not care" about the fact that

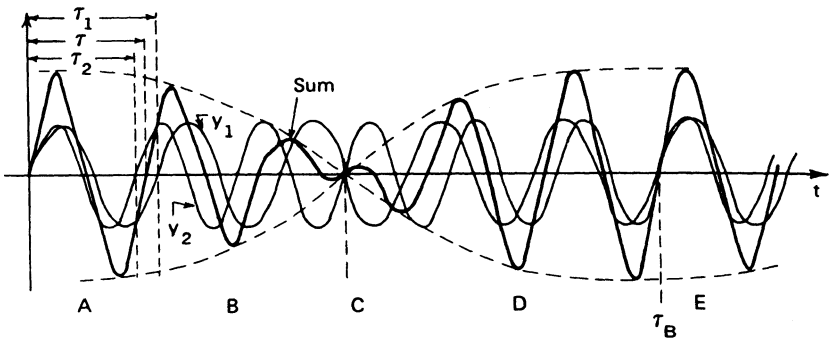



FIGURE 2.11 Superposition of two sinusoidal oscillations of slightly different periods $\tau_1$ and $\tau_2$ corresponding to the frequencies $f_1$ and $f_2$.

this pattern really is the result of the sum of two others. It has just *one* vibration pattern, of varying amplitude. A most remarkable thing happens in the cochlear fluid: this rather complicated, but single, vibration pattern at the oval window gives rise to *two* resonance regions of the basilar membrane. *If the frequency difference $\Delta f$ between the two component tones is large enough*, the corresponding resonance regions are sufficiently separated from each other; each one will oscillate with a frequency corresponding to the component tone (light curves in Fig. 2.11), and *we hear two separate tones of constant loudness*, with pitches corresponding to each one of the original tones. This property of the cochlea to disentangle a complex vibration pattern caused by a tone superposition into the original pure tone components is called *frequency discrimination*. It is a mechanical process, controlled by the hydrodynamic and elastic properties of the inner ear constituents. On the other hand, *if the frequency difference $\Delta f$ is smaller than a certain amount*, the resonance regions overlap, and *we hear only one tone of intermediate pitch with modulated or "beating" loudness*. In this case, the overlapping resonance region of the basilar membrane follows a vibration pattern essentially identical to that of the eardrum (heavy curve in Fig. 2.11). The amplitude modulation of the vibration pattern (envelope shown in Fig. 2.11) causes the perceived loudness modulation. We call this phenomenon "first-order beats." These are the ordinary beats, well known to every musician.

The frequency of the resulting vibration pattern of two tones of very similar frequencies $f_1$ and $f_2$ is equal to the average value:

$$f = \frac{f_1 + f_2}{2} = f_1 + \frac{\Delta f}{2} \tag{2.2}$$

The time interval $\tau_B$ (Fig. 2.11), after which the resulting amplitude attains the initial value, is called the beat period. The beat frequency $f_B = 1/\tau_B$ (number of amplitude changes per second) turns out to be given by the difference:

$$f_B = f_2 - f_1 = \Delta f \tag{2.3}$$

It makes no difference whether $f_2$ is greater than $f_1$, or vice versa. Beats will be heard in either case, and their frequency will always be given by the frequency difference of the component tones (relation (2.3) really must be taken in *absolute value*, i.e., positive only). The closer together the frequencies $f_2$ and $f_1$ are, the "slower" the beats will result. If $f_2$ becomes equal to $f_1$, the beats disappear completely: both component tones sound in *unison.*

Let us summarize the tone sensations evoked by the superposition of two pure tones of equal amplitude and of frequencies $f_1$ and $f_2 = f_1 + \Delta f$, respectively. To that effect, let us assume that we hold $f_1$ steady and increase $f_2$ slowly from $f_1$ (unison, $\Delta f = 0$) to higher values. (Nothing would change qualitatively in what follows, if we were to decrease $f_2$.) At unison, we hear one single tone of pitch corresponding to $f_1$ and a loudness that will depend on the particular phase difference between the two simple tones. When we slightly increase the frequency $f_2$,

we continue hearing *one* single tone, but of slightly *higher pitch*, corresponding to the average frequency $f = f_1 + \Delta f/2$ (2.2).[7] The loudness of this tone will be beating with a frequency $f_B = \Delta f$ (2.3). These beats increase in frequency as $f_2$ moves away from $f_1$ ($\Delta f$ increases); as long as $\Delta f$ is less than about 10 Hz, they are perceived very clearly. When the frequency difference $\Delta f$ exceeds, say, 15 Hz, the beat sensation disappears, giving way to a quite characteristic *roughness* or unpleasantness of the resulting tone sensation. When $\Delta f$ surpasses the so-called *limit of frequency discrimination* $\Delta f_D$ (not to be confused with the limen of frequency resolution DL of Fig. 2.9), we suddenly distinguish *two* separate tones, of pitches corresponding to $f_1$ and $f_2$. At that moment, both resonance regions on the basilar membrane have separated from each other sufficiently to give two distinct pitch signals. However, at that limit, the sensation of roughness still persists, especially in the low pitch range. Only after surpassing a yet larger frequency difference $\Delta f_{CB}$ called the *critical band*, the roughness sensation disappears, and both pure tones sound smooth and pleasing. This transition from "roughness" to "smoothness" is in reality more gradual; the critical band as defined here only represents the *approximate* frequency separation at which this transition takes place.

All these results are easily verified using two electronic "sine-wave generators" of variable frequency whose output is combined and fed monaurally into each ear with headphones. But they can also be verified, at least qualitatively, with two flutes played simultaneously in the upper register by expert players. While one flautist maintains a fixed tone (holding the pitch very steady), the other plays the same *written* note out of tune (e.g., pulling out or pushing in step by step the mouthpiece). Beats, roughness, and tone discrimination can be explored reasonably well.

Figure 2.12 is an attempt (not in scale) to depict the above results comprehensively. The heavy lines represent the frequencies of tones (or beats) that are *actually* heard. Tone $f_1$ is that of fixed frequency, $f_2$ corresponds to the tone whose frequency is gradually changed (increased or decreased). The "fused" tone corresponds to the single tone sensation (of intermediate frequency) that is perceived as long as $f_2$ lies within the limit of frequency discrimination of $f_1$. Notice the extension of the critical band on either side of the unison ($\Delta f = 0$). We must emphasize again that this transition from roughness to smoothness is not at all sudden, as one might be tempted to conclude from Fig. 2.12, but rather gradual. A detailed and far more rigorous discussion is given in Zwicker and Fastl (1999).

The limit for pitch discrimination and the critical band depend strongly on the average frequency $(f_1 + f_2)/2$ of the two tones (called the *center frequency* of the two-tone stimulus). They are relatively independent of amplitude, but may vary considerably from individual to individual. The critical band is related to several

[7] How do we verify that, indeed, the pitch sensation of the resulting tone *does* correspond to a tone of frequency $f = f_1 + \Delta f/2$? This is accomplished with *pitch matching* experiments: the subject is presented alternatively with a *reference* tone of controllable pitch and is requested to "zero in" the frequency of the latter until he senses "equal pitch" with respect to the original tone.

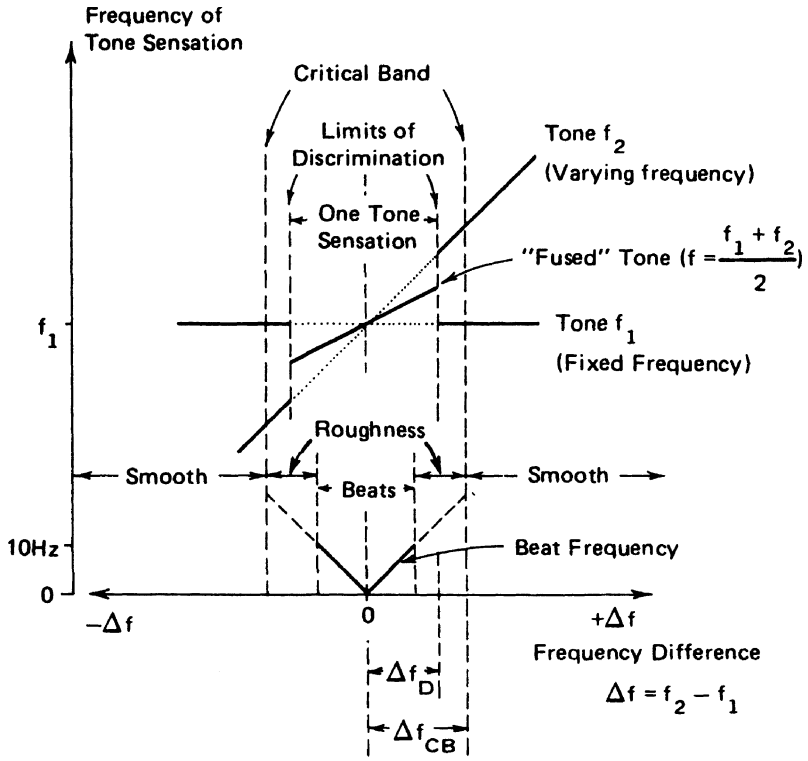



FIGURE 2.12 Schematic representation of the frequency (heavy lines) corresponding to the tone sensations evoked by the superposition of two pure tones of nearly frequencies $f_1$ and $f_2 = f_1 + \Delta f$.

other psychoacoustic phenomena, and there are many (and, as a matter of fact, far more precise) ways to define it experimentally (Sect. 3.4). Figure 2.13 shows the dependence of pitch discrimination $\Delta f_D$ (Plomp, 1964) and critical band $\Delta f_{CB}$ (Zwicker, Flottorp, and Stevens, 1957) with the center frequency of the component tones. For reference, the frequency differences that correspond to the musical intervals of a semitone, a whole tone, and a minor third are shown with broken lines. For instance, two tones in the neighborhood of 2000 Hz must be at least 200 Hz apart to be discriminated, and more than 300 Hz apart to sound smoothly. Note the remarkable fact that the limit of pitch discrimination is larger than a halftone,[8] and even larger than a whole tone at both extremes of very high and

[8]This may come as a surprise to musicians: they will claim that they can hear out very well the two component tones when a minor second is played on musical instruments! The point is this: the results shown in Fig. 2.13 only apply to *pure* tone superpositions, sounding *steadily* with constant intensity. When a musical interval is played with *real*

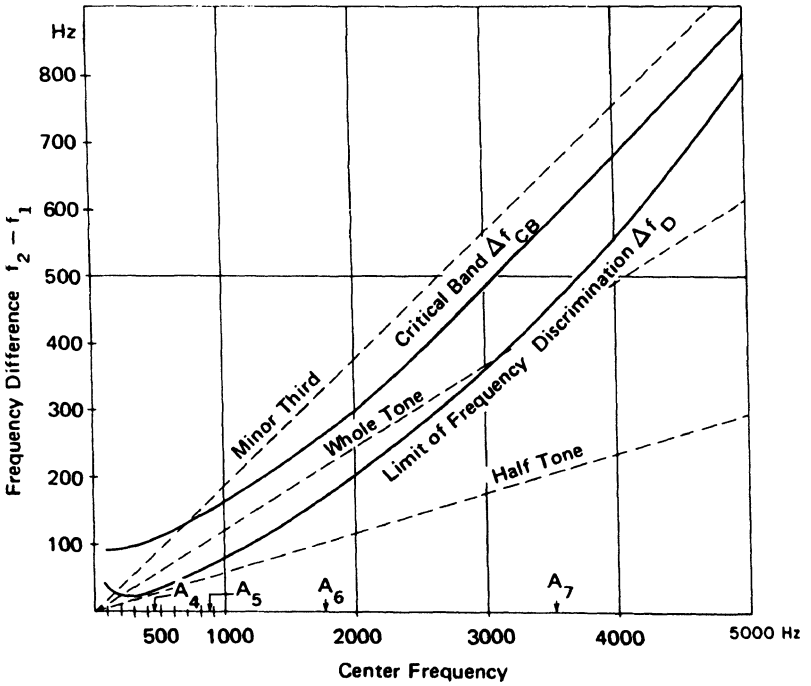



FIGURE 2.13 Critical bandwidth $\Delta f_{CB}$ (after Zwicker, Flottorp, and Stevens (1957)) and limit of frequency discrimination $\Delta f_D$ (Plomp, 1964) as a function of the center frequency of a two-tone stimulus (linear scales). The frequency difference corresponding to three musical intervals is shown for comparison).

very low frequencies. Note also the interesting fact that, in the high pitch range, the critical band lies between the frequency difference that corresponds to a whole tone interval (qualified as a "dissonance") and that of a minor third (termed a "consonance")—that is, roughly *extending over one-third of an octave*. In the low frequency range, there is an important departure: frequency discrimination and critical band are larger than a minor third (and even a major third). This is why thirds in general are not used in the deep bass register!

Compare Fig. 2.13 with Fig. 2.9: The limit for frequency discrimination $\Delta f_D$ is roughly 30 times larger than the DL for frequency resolution. In other words, we can detect very minute frequency changes of one *single* pure tone, but it takes

instruments, the tones are not simple tones, they do not sound steadily, and a stereo effect is present. All this gives additional cues to the auditory system that are efficiently used for tone discrimination.

an appreciable frequency difference between *two* pure tones sounding simultaneously, to hear out each component separately.[9]

What are the implications of these results for the theory of hearing? The existence of a finite limit for tone discrimination is an indication that *the activated region on the basilar membrane corresponding to a pure tone must have a finite spatial extension*. Otherwise, if it were perfectly "sharp," two superposed tones would always be heard as two separate tones as long as their frequencies differed from each other—no matter how small that difference—and no beat sensation would ever arise. Actually, the fact that the roughness sensation persists even beyond the discrimination limit, is an indication that the two activated regions still overlap or interact to a certain degree, at least until the critical band frequency difference is reached. An illustrative experiment is the following: Feeding each one of the two tones $f_1$ and $f_2$ *dichotically* into a different ear, the primary beat or roughness sensation disappears at once, both tones can be discriminated even if the frequency difference is way below $\Delta f_D$, and their combined effect sounds smooth at all times! The moment we switch back to a monaural input, the beats or roughness come back. Of course, what happens in the dichotic case is that there is only *one* activated region on each basilar membrane with no chance for overlapping signals in the cochlea;[10] hence no first-order beats or roughness.

At this stage the reader may wonder: If the region activated on the basilar membrane by *one* pure tone of *one* frequency is spatially spread, covering a certain finite range $\Delta x$ along the membrane, how come we hear only *one* pitch and not a whole "smear" over all those pitches that would correspond to the different positions within $\Delta x$ that have been activated? Unfortunately, we must defer the answer to later sections (e.g., Sect. 3.6). Let us just anticipate here that a so-called "*sharpening*" process takes place in which the activity collected along the whole region $\Delta x$ is "focused" or "funneled" into a much more limited number of responding

[9]There is an equivalent experiment that can be performed with the sense of *touch* to point out the difference between "resolution" and "discrimination." Ask somebody to touch the skin of your underarm for about 1 s on a fixed point with a pointed pencil *while you look away*. Then ask the person to repeat this at gradually displaced positions. It will require a certain small but finite minimum distance before you can tell that the position of touch has changed—this is the DL for localization of a single touch sensation, or "touch resolution." Now ask the person to use two pencils and determine how far both touching points must be from each other before you can identify *two* touch sensations. This is the minimum distance for "touch discrimination," which turns out to be considerably larger than the DL for touch resolution. Both touch resolution and discrimination vary along the different parts of the body. The equivalence between touch and hearing experiments is not at all casual: the basilar membrane is, from the point of view of biological evolution, a piece of epithelial tissue (skin) with a greatly magnified touch sensitivity! This analogy has been profusely used by von Békésy (1960) in his superb experiments.

[10]There is, however, an overlap of *neural signals* in the upper stages of the neural pathway, giving rise to "second order" effects, to be discussed in Sects. 2.6–2.9.

neurons. The beat phenomenon plays an important role in music.[11] Whenever beats occur, they are processed by the brain giving us sensations that may range from displeasing or irritating to pleasing or soothing, depending on the beat frequency and the musical circumstances under which they occur. The peculiar, displeasing sound of an instrument out-of-tune with the accompaniment is caused by beats. The ugly sound of out-of-tune strings in a mediocre high school orchestra is ugly, in part, because of the beats, and the "funny" sound of a saloon piano is caused by beats between deliberately out-of-tune pairs or triplets of strings in the middle and upper register. The fact that beats disappear completely when two tones have exactly the same frequency ($f_1 = f_2$) plays a key role for the process of *tuning* an instrument. If we want to adjust the frequency of a given tone to be exactly equal to the frequency of a given standard (e.g., a tuning fork), we do this by listening to beats and "zeroing in" the frequency until the beats have completely disappeared.

The critical band, too, plays a key role in the perception of music. We shall discuss this concept in later sections in more detail. For the time being, let us just remark that the critical band represents a sort of "information collection and integration unit" on the basilar membrane. The experimental fact that the critical band frequency extension $\Delta f_{CB}$ is roughly independent of sound amplitude or loudness is a strong indication that it must be related to some inherent property of the structure of the sensorial organ on the basilar membrane, rather than to the wave form in the cochlear fluid. Indeed, if one converts the frequency extension $\Delta f_{CB}$ shown in Fig. 2.13 into spatial extension along the basilar membrane by using Fig. 2.8, one obtains an almost constant value of about 1.2 mm for the critical band. An even more significant result is the following: The critical band turns out to correspond to an extension on the basilar membrane "serviced" by the roughly constant number of about 1300 receptor cells (out of a total of about 16,000 on the membrane) (Zwislocki, 1965), independently of the particular center frequency (i.e. position on the membrane) involved. A complex auditory stimulus (e.g., from two pure tones) whose components are spread over a frequency extension that lies *within* the critical band causes a subjective sensation (e.g., roughness, in our example) that usually is quite different from a case in which the extension *exceeds* that of the critical band (smoothness in the two-tone example). This is true for a variety of phenomena. It plays an important role in the perception of tone quality (Sect. 4.8) and provides the basis for a theory of consonance and dissonance of musical intervals (Sect. 5.2).

[11] In this chapter we have discussed the case of beats between *pure* tones only. As we shall see later, they similarly occur for the complex tones of real musical instruments.

## 2.5 Other First-Order Effects: Combination Tones and Aural Harmonics

So far, we have been analyzing the superposition effects of two pure tones whose frequencies were not too different from each other (Fig. 2.12). What happens with our tone sensations when the frequency of the variable tone $f_2$ increases beyond the critical band, while $f_1$ is kept constant? The ensuing effects may be classified into two categories, depending on whether they originate in the ear or in the neural system, respectively. In this section, we shall focus on a phenomenon belonging to the first category, the perception *of combination tones*. These tones are additional pitch sensations that appear when two pure tones *of* frequencies $f_1$ and $f_2$ are sounded together;[12] they are most easily perceived if the latter are of high intensity level and correspond to frequencies that differ from both $f_1$ and $f_2$, as can be established easily with pitch-matching or pitch-cancellation experiments (Goldstein, 1970). *They are not present* in *the original sound stimulus*—they appear as the result of a so-called nonlinear distortion of the acoustic signal in the ear.

Let us repeat the experiment corresponding to Fig. 2.12 concerning the superposition of two pure tones, but this time, turning up the loudness considerably and sweeping the frequency $f_2$ slowly up and down between the unison $f_1$ and the octave of frequency $2f_1$. While we do this, we pay careful attention to the evoked pitch sensations. Of course, we will hear both the tone constant pitch corresponding to the frequency $f_1$ and the variable tone $f_2$. But in addition, one clearly picks up one or more lower pitch tones sweeping up and down, depending on how we vary the frequency $f_2$. In particular, when $f_2$ sweeps upward away from $f_1$, we hear a tone of rising pitch starting from very deep. When $f_2$ sweeps downward starting at the octave $f_2 = 2f_1$, we also hear a tone of rising pitch starting from very deep. And paying even more attention, more than one low pitch tone may be heard at the same time. These tones, which do not exist at all in the original sound, are the combination tones.

Perhaps, the most easily identifiable combination tone is one whose frequency is given by the difference of the component frequencies, at *high* intensity level:

$$f_{C1} = f_2 - f_1 \tag{2.4}$$

It is also called the *difference tone*. Notice that for values of $f_2$ very close to $f_1, f_{C1}$ is nothing but the beat frequency (2.3). $f_{C1}$ must be at least 20–30 Hz to be heard as a tone. As $f_2$ rises, $f_{C1}$ increases, too. When $f_2$ is an octave above $f_2$, $f_{C1} = 2f_1 - f_1 = f_1$—that is, the difference tone coincides with the lower component $f_1$. When $f_2$ is halfway between $f_1$ and $2f_1$—that is, $f_2 = 3/2\,f_1$ (a musical interval called *the fifth*)—the difference tone has a frequency $f_{C1} = 3/2\,f_1 - f_1 = 1/2\,f_1$ corresponding to a pitch one octave below that of $f_1$.

[12]The lower frequency tone is sometimes called the *root* of the musical interval.

The two other combination tones that are most easily identified (Plomp, 1965), *even at low intensity levels of* the original tones, correspond to the frequencies

$$f_{C2} = 2f_1 - f_2 \tag{2.5}$$

$$f_{C3} = 3f_1 - 2f_2 \tag{2.6}$$

Both tones $f_{C2}$ and $f_{C3}$ decrease in pitch when $f_2$ increases from unison toward the fifth, and they are most easily heard when $f_2$ lies between about $1.1f_1$ and $1.3f_1$. At high intensity of the original tones, they also can be perceived quite well as low pitch sensations near the octave and the fifth respectively. Note that the tones $f_{C2}$ and $f_{C1}$ coincide in frequency $= 1/2\, f_1$ when $f_2$ is at the fifth $3/2\, f_1$. In Fig. 2.14, we summarize the first-order tone sensations evoked by the superposition of two pure tones of frequency $f_1$ and $f_2$. Notice that Fig. 2.12 is nothing but a "close-up" picture of what happens when the frequency $f_2$ is very near $f_1$ (hatched area in Fig. 2.14). The portions of the combination tones shown with heavier track are the ones easiest to be heard out (the actual extent depends on intensity).

How are these extra tone sensations generated? As pointed out above, they are *not* present in the original sound vibration of the eardrum. Careful experiments

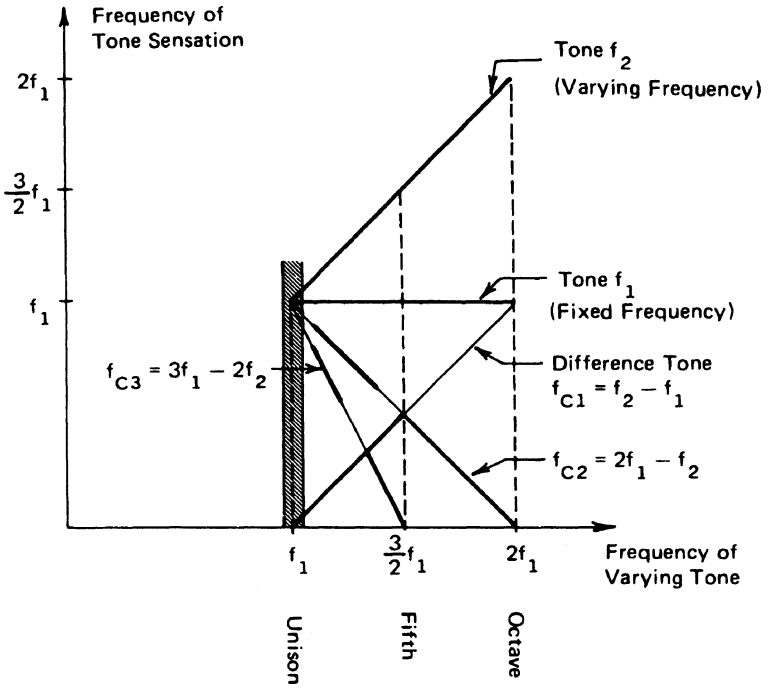



FIGURE 2.14 Frequencies of the combination tones $f_{C1}, f_{C2}, f_{C3}$, evoked by a two-tone superposition $(f_1, f_2)$. Heavy lines: most easily detected ranges of combination tones.

conducted on animals have shown that combination tone frequencies are not even present at the entrance of the cochlea (oval window membrane, Fig. 2.6); on the other hand, from direct neural pulse measurements (Goldstein, 1970), one must conclude that there *are* indeed activated regions on the basilar membrane at the positions corresponding to the frequencies of the combination tones. They are thought to be caused by a "nonlinear" distortion of the primary wave form stimulus in the cochlea. It can be shown mathematically that, indeed, when two harmonic (sinusoidal) oscillations of different frequencies $f_1$ and $f_2$ enter a transducer of distorted (nonlinear) response, the output will contain, in addition to the original frequencies $f_1$ and $f_2$, all linear combinations of the type $f_2 - f_1, 2f_1 - f_2, 3f_1 - 2f_2, f_2 + f_1, 2f_1 + f_2$, etc. Later experiments (Smoorenburg, 1972) however, indicate that the difference tone (2.4) and the two other combination tones (2.5, 2.6) must originate in mutually independent cochlear mechanisms respectively. The intensity threshold for the generation of the former (2.5) is considerably higher than that of the latter, and roughly independent of the frequency ratio $f_2/f_1$. On the other hand, the intensity of the latter (2.6) increases as $f_2/f_1$ approaches unity.[13]

It is interesting to note that, because of nonlinear distortion, even one *single* tone of frequency $f_1$ will give rise to additional pitch sensations when it is very loud. These additional tones, called *aural harmonics*, correspond to frequencies that are integer multiples of the original frequency: $2f_1$, $3f_1$, $4f_1$, etc.

Although all experiments pertinent to this section are most appropriately done with electronic tone generators, it is possible, at least qualitatively, to explore combination tones and aural harmonics using some musical instruments capable of emitting steady sounds at high intensity level. As a matter of fact, perhaps the most adequate "instrument" for this purpose is a dog whistle whose (very high) pitch can be varied. A simple homemade experiment on combination tones can be done by blowing two such whistles at the same time—one at constant pitch, the other one sweeping the frequency away from and back to unison—and listening to low pitch tone sensations. Combination tones do not play an important role in music because of the relatively high intensity of the original tones required for their elicitation. They must, however, appear everywhere when rock music is listened to at the usual, exaggerated (and physiologically damaging) loudness levels, especially with earphones!

"Fake" combination tones can be easily generated in electronic organs and low quality hi-fi amplifiers and speakers. In these cases it is a nonlinear distortion in the electronic circuitry and mechanical system of the speaker that generates these parasitical frequencies. In particular, the difference tone can be produced and listened to quite clearly with an electronic organ: turn the loudness up, pull 8' flute stops, play back and forth the sequence shown on the upper staff of Fig. 2.15, and listen for the low pitch tones indicated on the bottom staff.

[13]The reasons for this difference in behavior are not clearly understood. Nor is it known why the combinations $f_2 + f_1, 2f_1 + f_2$, etc., do not appear as tone sensations.

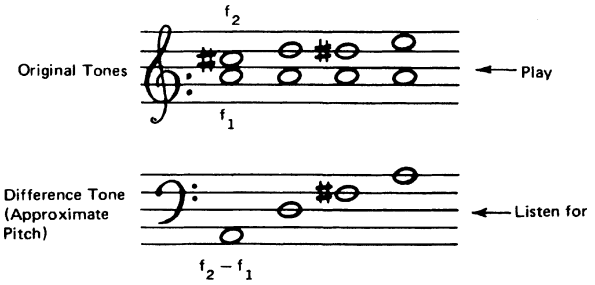



FIGURE 2.15 Difference tones heard (lower staff) when the two-tone superpositions shown on the upper staff are played (loud!).

Some of the difference tones thus generated are out of tune because of the equal-temperament tuning of the instrument (Sect. 5.3). We must point out again that what is heard in this experiment is a *fake* combination tone, in the sense that the low pitch sensation is generated in the speaker and *not* in the ear. It is quite clear from this example why electronic circuitry and speakers of hi-fi systems and electronic organs should have a "very" linear response.

## 2.6 Second-Order Effects: Beats of Mistuned Consonances

We now repeat the experiment of the preceding section with two electronically generated pure tones, but, this time, ignoring possible combination tone sensations. We feed both tones at *low* intensity level into the same ear; tone $f_1$ is held at constant frequency, whereas $f_2$ can again be varied at liberty. The amplitudes of both pure tones are held constant throughout the experiment. When we sweep $f_2$ slowly upward, we notice something peculiar when we pass through the neighborhood of the octave $f_2 = 2f_1$: a distinct beating sensation, quite different from the first-order beats near unison, but clearly noticeable. When $f_2$ is exactly equal to $2f_1$, this beat sensation disappears. It reappears as soon as we mistune the octave, that is, when $f_2$ is set at $f_2 = 2f_1 + \varepsilon$, where $\varepsilon$ (epsilon) represents only a few Hertz. The beat frequency turns out to be equal to $\varepsilon$. It is difficult to describe "what" is beating. Most people describe it is a beat in tone "quality." We call these *second-order* beats; some prefer the name "subjective beats." They are the result of neural processing.

It is instructive to watch the vibration pattern on the oscilloscope while one listens to second-order beats. This pattern is seen to change in exact synchronism with the beat sensations. Obviously, our auditory system must somehow be able to detect these changes of the *form* of a vibration pattern. Figure 2.16 shows several vibration patterns corresponding to the superposition of a fundamental tone of frequency $f_1$ and its octave $f_2 = 2f_1$ (of smaller amplitude), for four different

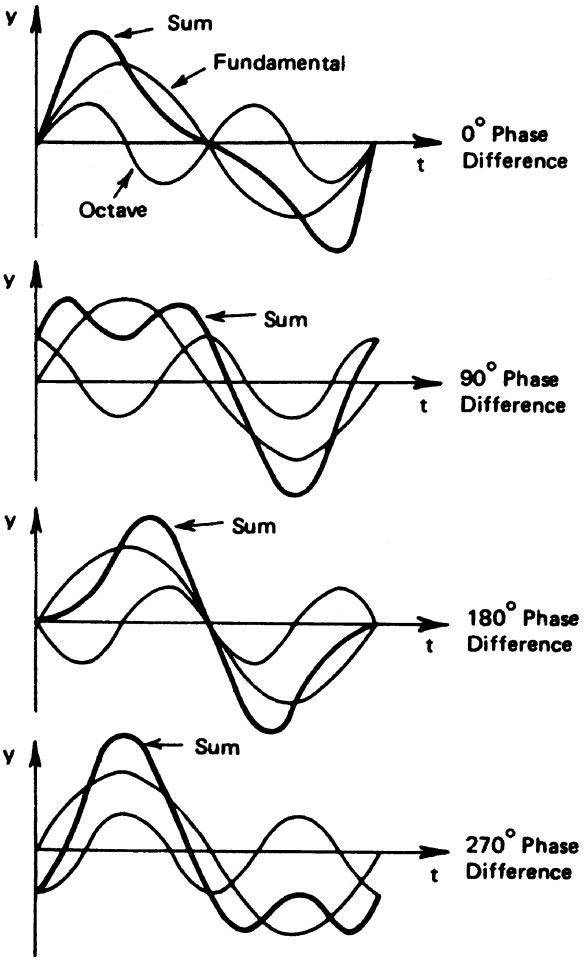



FIGURE 2.16 Octave superpositions of two pure tones, shown for four different phase differences.

values of the phase difference. As long as the octave is perfectly in tune, the phase difference remains constant and the image on the oscilloscope static; any of the four superpositions sounds like the other—our ear does not distinguish one case from another. But when we throw $f_2$ slightly out of tune: $f_2 = 2f_1 + \varepsilon$, the mutual phase relationship will change continuously with time and the resulting vibration pattern will gradually undergo a shift from one of the forms shown in Fig. 2.16 to the next. It can be shown mathematically that this cycle of changing vibration pattern repeats with a frequency $\varepsilon$, the amount by which the upper tone is out of tune from the octave. This obviously means that the ear is sensitive to a

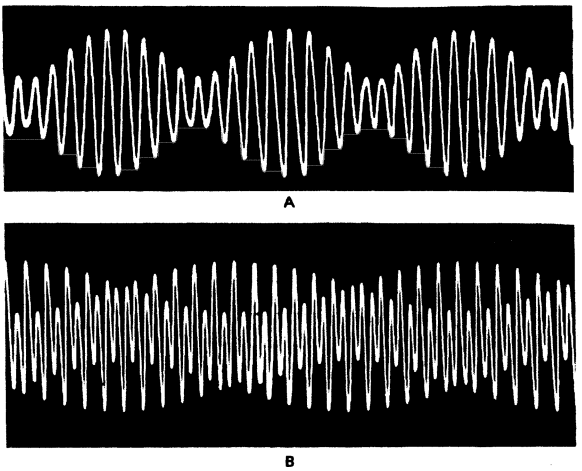



FIGURE 2.17 Comparison of first-order and second-order beats. (A) First-order beats (mistuned unison); amplitude modulation with no change in vibration pattern form. (B) Second-order beats (mistuned octave); pattern modulation with little change in average amplitude.

slowly changing phase difference between two tones.[14] An equivalent statement is: *The auditory system is capable of detecting cyclic changes in vibration pattern forms*. Notice carefully that there is no appreciable "average" change in amplitude from pattern to pattern in Fig. 2.16—quite contrary to what happens with first-order beats, which are cyclic changes of vibration pattern amplitude (Fig. 2.11). Figure 2.17 shows two actual oscilloscope pictures confronting first-order beats near unison and second-order beats of a mistuned octave. Note the amplitude modulation of the former and the vibration pattern modulation in the latter. It is important to take into account that the second-order beat sensation only appears in the low frequency range of the original two-tone stimulus. When $f_1$ (and $f_2$) exceeds about 1500 Hz, second-order beats cannot be perceived (Plomp, 1967a).

Now, we turn again to our experimental setup and explore the whole frequency range between unison and the octave. We discover that there are other pairs of values for $f_2$ and $f_1$, that is, other musical intervals, in whose neighborhood beat sensations appear, though much less perceptible than for the octave. Two such "beat holes," as we may call them, can be found centered at the frequencies $f_2 = 3/2 f_1$ and $f_2 = 4/3 f_1$ corresponding to the musical intervals fifth and fourth respectively. Again, watching the vibration pattern on the oscilloscope at the same

[14] *Sudden* changes of phase (e.g., presentations of the octave stimuli represented in Fig. 2.16 alternated with a reference octave stimulus) are also detected. The degree of detectability has a maximum for a 180° phase difference with that of the reference signal (Raiford and Schubert, 1971).

time as we listen, we realize that for a mistuned fifth ($f_2 = 3/2 f_1 + \varepsilon$) and a mistuned fourth ($f_2 = 4/3 f_1 + \varepsilon$), the vibration pattern form is not static (as happens with an exact fifth or fourth, i.e., for $\varepsilon = 0$), but changes periodically in form (not in amplitude). The second-order beats are "faster" than those of the octave (for the fifth, the beat frequency is $f_B = 2\varepsilon$, for the fourth $f_B = 3\varepsilon$). This is not the only reason why they are more difficult to pick up: the vibration pattern itself gets more and more complicated (i.e., departs more and more from that of a simple harmonic motion) as we go from the octave (Fig. 2.16) to the fifth and to the fourth. The more complex a vibration pattern, the more difficult it is for the auditory system to detect its periodic change. For a detailed discussion of beats of mistuned consonances, see Plomp (1976).

There is an optimal relationship between the intensities of the component tones for which second-order beats are most pronounced, which always places the *higher pitch tone at a lower intensity* (Plomp, 1967a). Finally, it is important to note that second-order beats are also perceived when each of the component tones is fed separately *into a different ear*. In that case, a strange sensation of spatial "rotation" of the sound image "inside" the head is experienced (Sect. 2.9).

Second-order beats of mistuned consonances of *pure* tones do not play an important role in music (mainly because pure tones do not). But they are an important ingredient to the understanding of the processing mechanism of musical sound (Sect. 2.8).

## 2.7 Fundamental Tracking

We now introduce another series of psychoacoustic experiments that have been of crucial importance for the theories of auditory perception. Let us consider two pure tones, a perfect fifth apart, of frequency $f_1$ and $f_2 = 3/2 f_1$. Figure 2.18 shows the resulting vibration (sum) for one particular phase relationship. Note that the pattern repeats the exact shape after a time $\tau_0$, which is twice as long as $\tau_1$, the period of the lower pitch tone. This means that the *repetition rate* $f_0 = 1/\tau_0$ of the vibration pattern of a fifth is equal to one-half the frequency of the lower tone:

$$f_0 = \frac{1}{2} f_1 \tag{2.7a}$$

We call this repetition rate the "fundamental frequency" of the vibration pattern. In this case, it lies one octave below $f_1$. If we consider two tones forming a fourth ($f_2 = 4/3 f_1$), we can plot the resulting vibration pattern in the same way as was done for the fifth. The resulting repetition rate is now

$$f_0 = \frac{1}{3} f_1 \tag{2.7b}$$

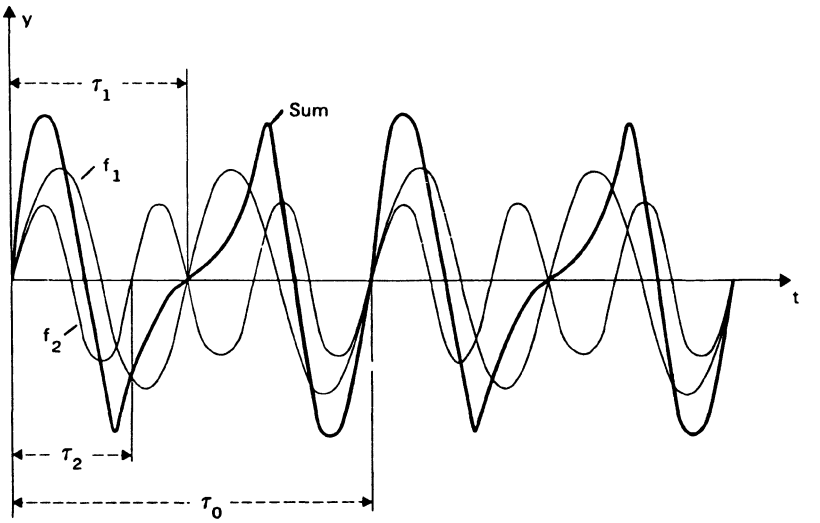



FIGURE 2.18 Superposition of two pure tones a musical fifth apart (for a given phase relationship). $\tau_0$: repetition period of the resulting sound.

that is, a twelfth below the lowest tone. For a major third ($f_1 = 5/4/f_1$), the repetition rate lies two octaves below $f_1$:

$$f_0 = \frac{1}{4} f_1 \tag{2.7c}$$

*Our auditory system turns out to be sensitive to these repetition rates*. Indeed, careful experiments have been performed in which the subjects were exposed to short sequences of stimuli made up of pairs of simultaneously sounding pure tones a fifth, a fourth, a third, etc., apart (Houtsma and Goldstein, 1972). These subjects were cued into identifying a *single* basic pitch of the "melody." Most of them did indeed single out a pitch that is matched by a frequency given by relations (2.7a), (2.7b), or (2.7c), respectively![15] It is important to point out that this pitch identification experiment demands that the two-tone complexes be presented as a *time sequence* or melody. (When confronted with a steady-sounding pair of pure tones, our auditory system fails to "seek" a single pitch sensation; it very rapidly refocuses its attention in order to discriminate the spectral pitches of both pure tone components as explained in Sect. 2.4.)

[15]Please note that this experiment *must* be performed with pairs of sinusoidal, electronically generated tones—it will *not work* on the piano or on any other musical instrument. See, however, subsequent remark on organs.

Note that the repetition rates (2.7a)–(2.7c) of the above two-tone complexes are identical to the frequencies of the corresponding difference tones (e.g., see fourth, second, and first cases in Fig. 2.15). However, the experiments have shown that repetition rate detection is successful even if the intensities of the two tones $f_1$ and $f_2$ are low, way below the threshold for combination tone production. A difference tone (2.4) is thus ruled out (Plomp, 1967b). Actually, repetition rate detection has been used in music for many centuries (and wrongly attributed to a combination tone effect). For instance, since the end of the 16th century, many organs include a stop (the "5 1/3 foot fifth") composed of pipes sounding a fifth higher than the pitch of the written note actually played. The purpose is to stimulate or reinforce the bass one octave below (Eq. (2.7a)) the pitch of the written note (i.e., to reinforce the 16′ sound of the organ). Of even older usage is the 10 2/3 foot fifth in the pedals, a "cost-effective" stop which, when used in combination with 16′ stops, simulates the 32′ bass (two octaves below the written note, requiring very tall and expensive pipes).

The tone of frequency $f_o$ (2.7) is not present as an original component. This tone is called the *missing fundamental* (for reasons that will become apparent below); the corresponding pitch sensation is called *periodicity pitch, subjective pitch, residue tone*, or *virtual pitch*. This pitch sensation should be clearly distinguished from the primary or spectral pitch of each one of the two original pure single-frequency tones. Experiments have shown that for normal loudness levels the frequency $f_o$ is *not* present in the cochlear fluid oscillations either (whereas combination tones are). Indeed, the region of the basilar membrane corresponding to the frequency $f_o$ (Fig. 2.8) may be saturated (masked) with a band of noise (sound of an infinite number of component frequencies lying within a given range) so that any additional excitation of that region would pass unnoticed—yet the missing fundamental will still be heard (Small, 1970). Or, one may introduce an extra tone slightly out of tune with $f_o$; first-order beats should appear if the missing fundamental tone $f_o$ really did exist in the cochlea—yet no beat sensation is felt. A more drastic effect is that *the missing fundamental is perceived even if the two component tones are fed in dichotically*, one into each ear (Houtsma and Goldstein, 1972). All of this indicates that the missing fundamental, or periodicity pitch, must be the result of neural processing at a higher level.

Subjective pitch detection, that is, the capability of our auditory system to identify the repetition rate of the incoming, unanalyzed vibration pattern, only works in the lower (but musically most important) frequency range, below about 1500 Hz. The more complex the vibration pattern, that is, the smaller the interval between the component tones, the more difficult it is for the auditory system to identify the missing fundamental, and the more ambiguous the subjective pitch becomes.

Let us now "turn around" relations (2.7) and find out which pairs of pure tone frequencies give rise to the *same* repetition rate or fundamental frequency $f_o$. We obtain:

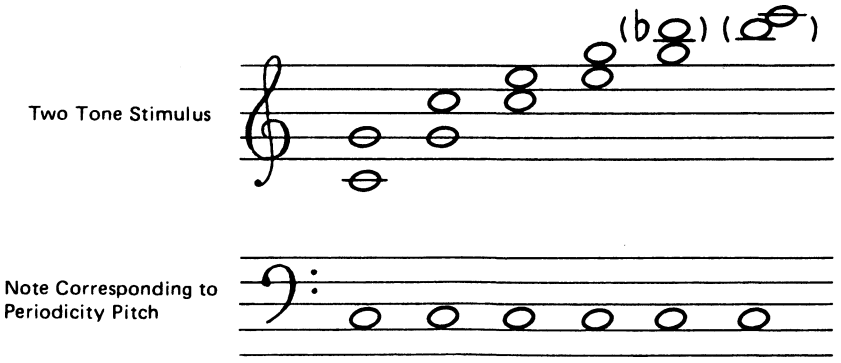



FIGURE 2.19 Two-tone stimuli (*upper staff*) that give rise to the same periodicity pitch (*lower staff*). The *b*-flat marked between parentheses most be tuned flat with respect to any scale in use (Sect. 5.3) in order to yield a C as a periodicity pitch.

$$\underbrace{2f_0 \text{ and } 3f_0}_{\text{fifth}} \qquad \underbrace{3f_0 \text{ and } 4f_0}_{\text{fourth}}$$

etc.

$$\underbrace{4f_0 \text{ and } 5f_0}_{\text{major third}} \qquad \underbrace{5f_0 \text{ and } 6f_0}_{\text{minor third}}$$

In other words, if $f_0$ corresponds to the note shown on the lower staff in Fig. 2.19, the musical intervals shown in the upper staff yield that same note as a subjective pitch sensation. It is important to think of the notes in Fig. 2.19 as representing *pure* tones of just one frequency each, *not* as tones produced by real musical instruments.

The individual components of frequency $2f_0, 3f_0, 4f_0, 5f_0, \ldots$, etc., are called the *upper harmonics* of the fundamental frequency $f_0$. Upper harmonic frequencies are integer multiples of the fundamental frequency. Any two successive tones of the upper harmonic series form a pair with the *same* repetition rate or fundamental frequency $f_0$. Therefore, all upper harmonics, if sounded together, will produce one single subjective pitch sensation corresponding to $f_0$—*even if that latter frequency is totally absent in the multitone stimulus!* This is the reason why, in the above examples, $f_0$ has also been called the "missing fundamental," and why the perception of this repetition rate is called *fundamental tracking*. Note once more the striking property of this set of pure tones of frequencies $2f_0, 3f_0, 4f_0, \ldots, nf_0 \ldots$—out of the infinite variety of imaginable pure tone superpositions, this is the one and only whose components, taken in pairs of contiguous frequencies, yield one and the same repetition rate. Conversely, this is the reason why *any* periodic tone with a complex but repetitive vibration pattern (of repetition rate $f_0$) is made up of a superposition of pure tones of frequencies $nf_0$ ($n$ = whole number) (see Sect. 4.2).

The above-mentioned psychoacoustic experiments with two-tone complexes have been extended to include melodies or sequences composed of *multitone*

complexes starting on the $n$th harmonic (i.e., superpositions of pure tones of frequencies $nf_0$, $(n + 1)f_0$, $(n + 2)f_0$, etc.). Although, again, the tone of fundamental frequency is missing, the subjective pitch assigned to these tone complexes always corresponds to $f_0$. As a matter of fact, the more harmonics are included, the clearer the periodicity pitch is heard (unless the starting harmonic order is very high, i.e., $n$ is large). The most crucial pairs of neighboring harmonics for periodicity pitch determination are those around $n = 4$ (Ritsma, 1967). Because real musical tones happen to be made up of a superposition of harmonics (Chap. 4), *fundamental tracking is the auditory mechanism that enables us to assign a unique pitch sensation to the complex tone of a musical instrument.*[16]

It is important to understand the full implication of fundamental tracking for the theory of hearing. A brief analysis of Smoorenburg's (1970) historic pitch-matching experiments will serve that purpose.[17] Consider a short duration two-tone stimulus whose component frequencies $f_a$ and $f_b$ differ by a fixed amount $\Delta f = (f_b - f_a)$. When it is presented in some specific context, about half of the test subjects perceive a clearly identifiable low subjective "residue" pitch (the others only seem to be able to hear one or both pitches of the original stimulus, i.e., they were listening analytically rather than synthetically). The experiments show that if $f_a$ and $f_b$ correspond to *two neighboring harmonics* of a complex tone of order $n$ and $n + 1$, the subjective pitch, when perceived, is that of the missing fundamental $f_1$ $(= f_a/n = \Delta f)$. For instance, if $f_a = 800$ Hz and $f_b = 1000$ Hz ($n = 4$ and $\Delta f = 200$ Hz), the residue pitch heard is that of a 200 Hz note. Figure 2.20(a) shows the vibration pattern of the two-tone stimulus, reminiscent of the first-order beat phenomenon (Sect. 2.4). In this case, however, the amplitude modulation (change of the "envelope" of the curve) is very rapid (200 times per second) and is not perceived as a beat. Rather, what is perceived (by about 50% of the test subjects) is a pitch corresponding to the repetition rate of the vibration pattern, which is exactly 200 Hz. The corresponding repetition period is $\tau = 1/\Delta f = 1/(f_b - f_a)$ (the period of the missing fundamental), indicated in the figure. We also marked the other, much shorter, period in the temporal fine structure of the vibration pattern, which corresponds to the so-called center frequency of the two-tone stimulus $f_c = (f_a + f_b)/2$ $(= 900$ Hz). For another neighboring harmonic pair, such as, for instance, $f_a = 2000$ Hz and $f_b = 2200$ Hz ($n = 10$), the *same* subjective pitch is perceived (with increasing difficulty as $n$ increases—Houtsma, 1970); the vibration pattern has exactly the same envelope as in Fig. 2.20(a), but since the center frequency is now higher (2100 Hz), the vibration

[16] Perhaps the most convincing example of fundamental tracking of complex tones is given by the fact that one is still able to perceive the correct pitch of bass tones from a cheap transistor radio *in spite* of all frequencies below 100–150 Hz being cut off by the inadequate electronic circuitry and speaker!

[17] Unfortunately, these interesting experiments cannot be demonstrated easily, even in a well-equipped physics teaching laboratory!

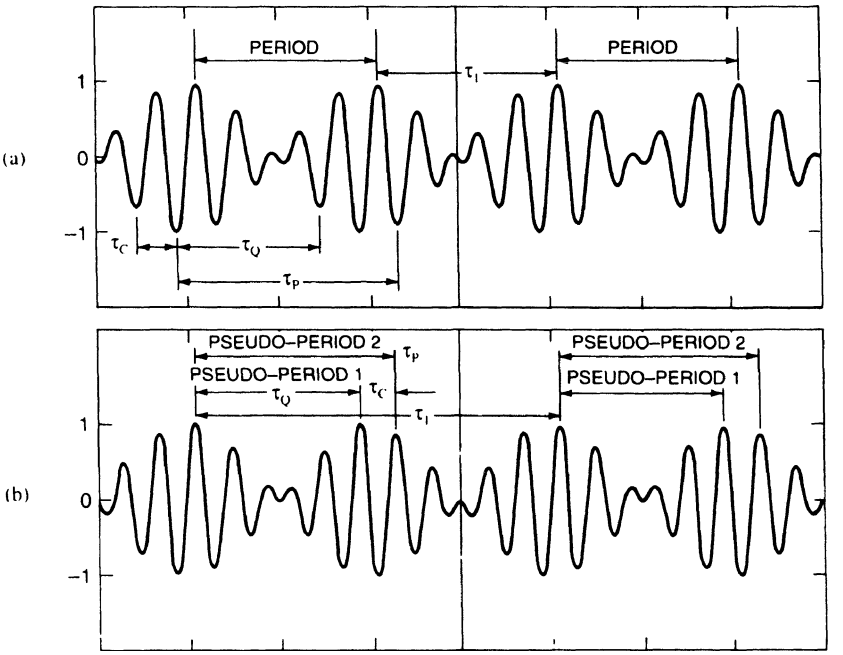



FIGURE 2.20 Vibration pattern of two simultaneous pure tones. (a) The tones are neighboring harmonics ($n$ 4). (b) The tones have the same frequency difference as in (a), but are not neighboring harmonics. $\tau_1$: Exact repetition period of the vibration pattern. $\tau_C$: Period of the center frequency. $\tau_p$, $\tau_q$: Pseudoperiods (see text).

curve has more oscillations (ten, the harmonic order of $f_a$) within one repetition period.

It seems that our hearing system extracts information from the periodic change of the vibration pattern, as it does when (slow) beats of mistuned consonances are perceived (Fig. 2.17). However, an interesting complication arises when the pair $f_a$, $f_b$ does *not* correspond to two neighboring harmonics of some fundamental. Consider a two-tone stimulus in which $f_a = 900$ Hz and $f_b = 1100$ Hz. There is no musical tone of which these could be *neighboring* harmonics; rather, they are the 9th and the 11th harmonic of a tone of fundamental frequency $f_1 = 100$ Hz.[18] Is this the pitch perceived? No! The perception turns out ambiguous: *two possible pitches* can be matched (depending on the context in which the stimulus is presented), corresponding to about 178 Hz or 228 Hz! Figure 2.20(b) depicts the vibration pattern for this case. First, note that the vibration pattern shows an

[18]From now on we will designate the fundamental frequency with $f_1$ rather than $f_0$ (as in relations 2.7 and Fig. 2.19), so that the whole series of harmonics of a complex tone can be designated generically as $f_1, f_2 = 2f_1, f_3 = 3f_1, \ldots \{f_n = nf_1, n \geq 1\}$. In the literature, the fundamental frequency and repetition rate are often designated as *F0*.

envelope with the same modulation period as that in Fig. 2.20(a) (i.e., corresponding to a frequency of 200 Hz). Second, examine very carefully the pattern of peaks and valleys and note that the *exact* fine structure repeats with a period *twice* as long (i.e., corresponding to 100 Hz). The exact repetition rate is thus 100 Hz but neither this one nor the 200 Hz modulation rate is perceived. In fact, it turns out that the two possible pitches heard correspond exactly to the two "pseudoperiods" $\tau_p$ and $\tau_q$ marked on the figure! Moreover, Smoorenburg's experiments show that even in the first example (when the stimulus consists of two neighboring harmonics), additional ambiguous pitches can be heard corresponding to pseudoperiods defined by the time intervals between the central peak and secondary peaks in the subsequent modulation period (Fig. 2.20(a)).

All this is an indication that a far more sophisticated pitch extraction process is at work than the detection of either repetition rate or amplitude modulation rate: either the hearing mechanism is capable of locking on to very detailed, distinct temporal features of the vibration pattern, or a very different signal recognition process is at work, in which the *spatial* pattern of excitation elicited by the two-tone stimulus along the basilar membrane is analyzed in detail by the pitch processor and matched as closely as possible to "familiar" configurations (e.g., the position of resonance regions of neighboring harmonics). Whenever a match is achieved, a pitch sensation is elicited; since more than one "acceptable" match may be possible, ambiguous pitches can result. It can be proven mathematically in many aspects (but not all) that this process leads to the same quantitative results as the above time-cue analysis mechanism. We will elaborate further on this subject in Sect. 4.8 and Appendix II.

Finally, the disparity among individuals concerning the ability to perceive the subjective residue pitches of two-tone stimuli has been invoked by some psychoacousticians to sound a note of caution regarding the interpretation of Smoorenburg's experiments. However, the consistency of the quantitative results (verified by several independent research groups) for those subjects who *do* hear the residue pitch is so remarkable that said disparity may be just the indication of a difference in listening *strategies*, without much consequence for the conclusions drawn from these experiments on the pitch extraction mechanism per se. For an excellent historical review of the most important pitch perception experiments, see Plomp (1976). For a detailed up-to-date discussion of pitch perception and related literature references, see Plack and Oxenham (2005).

## 2.8 Auditory Coding in the Peripheral Nervous System

The discovery of second-order effects in auditory processing, such as the perception of beats of mistuned consonances and fundamental tracking, has had a great impact on the theory of hearing. On the one hand, the perception of beats of mistuned consonances (Sect. 2.6) is an indication that the auditory system somehow does obtain and utilize information on the detailed time structure of the acoustic vibration pattern. On the other hand, fundamental tracking

(Sect. 2.7) could, in principle, imply two alternatives: (1) A mechanism performing a detailed analysis of the temporal pattern of vibration, with the instruction of zeroing in on repetitive features whose rate then leads to a pitch sensation (Fig. 2.20); or (2) A mechanism that analyzes information on the details of the spatial excitation pattern elicited along the basilar membrane, with the instruction to yield a single pitch sensation if that pattern matches at least part of the characteristic excitation elicited by a musical tone. We must anticipate that the second alternative should work best in the region of the lower harmonics (lower $n$ values) where the corresponding spatial excitation maxima are most distinct. Either alternative implies that detailed acoustic information encoded in the periphery must be analyzed at a higher level in the central nervous system. Moreover, both mechanisms may operate simultaneously in a mutually complementary fashion (Sects. 2.9 and 4.8, and Appendix II).

In order to understand the proposed mechanisms we should first describe a few generalities of neural system operation. When locomotion appeared in multicellular species, a fast and more complex information-processing and storage capability became necessary. A *nervous system* evolved together with the sensory and motor machinery to couple the organism to the outside world in real time. At the earliest stage of evolution, the nervous system merely served as a "high-speed" information transmission device in which an environmental, physical or chemical signal converted into an electrical pulse in a sensory detector, is conveyed to a distant part of the organism, triggering a contraction in a muscle fiber. Later, neural networks evolved that could analyze and discriminate among different types of input signals and, depending on the case, send specific orders to different parts of the body. Finally, memory circuits emerged, that could store relevant information acquired during the organism's lifetime to successfully face previously experienced environmental challenges again at a later time.

The basic building block and information-processing unit of a neural system is the *neuron*, a cell with an electrochemically active membrane capable of generating short electric pulses that serve as information carriers. To begin, one usually introduces a "formal" or "ideal" neuron—a cell *model* that bears some salient features of most but by no means all types of neurons, from jellyfish to human cortex. These main features are (Fig. 2.21): (1) A *dendritic tree*, collecting signals that arrive from "upstream" neurons through so-called *synaptic connections*; (2) The cell body or *soma* which contains the nucleus of the cell and controls the neuron's metabolic functions, and which also may receive direct input from presynaptic neurons; and (3) The *axon*, a process that conveys standard-sized electric pulses away from the soma to postsynaptic neurons. The neuron is thus the most basic information collection, integration, and discrimination unit of the nervous system.

The electric potential of an undisturbed neuron with respect to its surrounding medium is constant and always negative (about –70 millivolt (mV)—a pretty large value, considering that its membrane is only about 70 nanometers thick ($7 \times 10^{-8}$ m!). When a *neurotransmitter* is released into a synaptic gap by a presynaptic cell, a local change in the membrane permeability to electrically charged atoms (sodium, potassium, and calcium ions) causes a local change of the electric

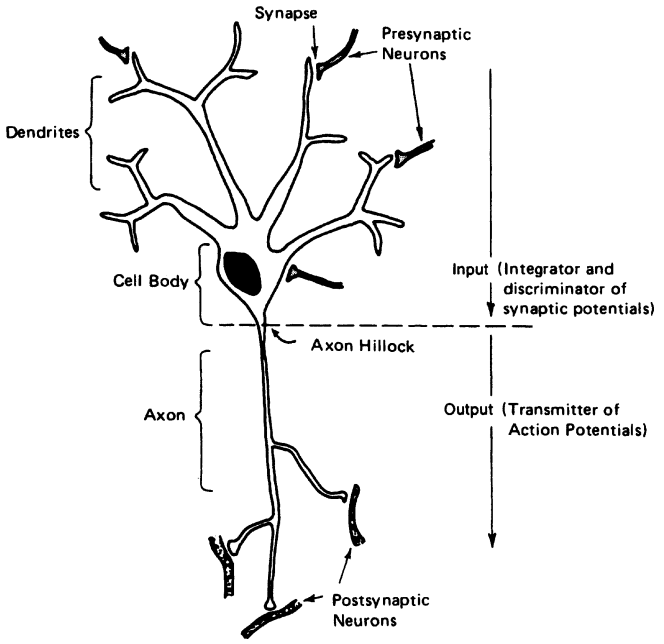



FIGURE 2.21 Sketch of a model neuron.

potential of that postsynaptic neuron, a pulse (or impulse) which then propagates toward the soma. Some synapses (in which the neurotransmitter is dopamine, glycine, or gamma-aminobutyric acid) produce positive pulses, called *excitatory postsynaptic potentials* (EPSP); other synapses (with a neurotransmitter like norepinephrine) produce negative pulses or *inhibitory postsynaptic potentials* (IPSP). EPSPs are of the order of +8 to +10 mV, IPSPs from –5 to –8 mV; they reach a peak in a few milliseconds and decay within about 10 ms. A characteristic time delay (typically less than a millisecond) occurs between the arrival of an impulse at a synapse, the release of the neurotransmitter into the synaptic gap, and the formation of the response in the postsynaptic cell.

All this can be measured with microelectrodes implanted in the cell—a procedure which does not necessarily affect the actual operation of a neuron! The relative amplitude of the signal generated in a postsynaptic dendrite depends on the actual structure and size of the synaptic junction and on the number of neurotransmitter receptors on the postsynaptic cell membrane. As a matter of fact, amplitude and duration of a postsynaptic potential are a measure of the *potency* (or efficacy, efficiency) of the synapse where the impulse was generated, which, as we shall see later (Sect. 4.10), in certain neural networks can change during a learning experience.

Incoming pulses add up more or less linearly as they travel to the cell body. However, there is a limit to this linear behavior in excitable membranes: if the

(always negative) cell potential is depolarized (made less negative) above a certain threshold value (this, too, can be done artificially with a microelectrode), a fundamental instability arises and a positive depolarization pulse is triggered. The threshold is smallest at the axon hillock (Fig. 2.21), and it is precisely there where the positive *action potentials* (AP) are generated. The triggered impulse is independent of the triggering event, and is fired down the axon in a *standardized* digital fashion. AP's are electric pulses of a few milliseconds duration and of considerably larger amplitude than postsynaptic potentials (thus easier to pick up with a microelectrode). Their speed of propagation ranges from about 0.3 m/s in the axons of gray matter in the cortical tissue to about 130 m/s in the myelinated fibers of white matter and efferent (outgoing) nerves.[19] New evidence shows that the AP-generating process can also launch retropropagating pulses back into the dendritic tree; this feedback process may be relevant to the learning process (Sect. 4.9). By implanting a microelectrode in live neural tissue, amplifying the registered pulses (which mainly will come from the much larger APs than EPSPs or IPSPs), and converting them into audio signals, one can listen to characteristic crackle or click sounds in the speaker revealing the function of an individual neural fiber (axon) "under interrogation." For quantitative practical purposes, the electric spikes are registered digitally or graphically as a function of time; their frequency and distribution in time represent a physical expression of neural information at the single neuron level.

The action potential thus represents the fundamental single neuron *output message*. Axons are "wired" to (i.e., make synaptic contacts with) dendrites or cell bodies of other neurons (Fig. 2.21); in the human brain, one given axon may be in synaptic contact with thousands of other neurons. Conversely, one given neuron may be wired to incoming axons from hundreds or thousands of other cells. A single neuron may only impart either excitatory or inhibitory orders to other neurons. After activation, the neuron has a refractory period during which it cannot be re-excited, or during which its firing threshold is increased. When an inhibitory neuron fires a pulse to another inhibitory neuron, it cancels the inhibitory effect of the latter. It is important to note that whether a neuron will fire an output signal is determined by both the spatial and temporal distributions of EPSPs and IPSPs generated within a certain short interval of time by the presynaptic neurons. This is why a neuron is sometimes called the basic "computing unit" of the nervous system. On the other hand, there are many neurons which fire action potentials *spontaneously* at some characteristic rate, without any input signals.

[19]Thus the time it takes a neural impulse (AP) fired by a motoneuron at the top of your brain to reach a muscle moving your toe (myelinated fibers of, say, 1.5 m total length, with two or three intermediate synaptic stages, each of which takes a little less than a millisecond to transmit a signal), would be approximately 15 ms. The time it takes for neural information to be transferred from the left auditory cortex to the right one (involving travel through about 2 cm gray matter and 10 cm white matter, assuming no synapses in between) would be approximately 70 ms, a pretty long time when it comes to, say, speech processing (Sect. 5.7)!

As mentioned above, individual firings usually do not occur equally spaced at regular time intervals. What counts in terms of the actual information being sent is the *fact* that a neuron is firing, or, more importantly, the *average rate* (firings per unit time), or, as we shall discuss in connection with Fig. 2.23, in certain networks it is the actual distribution in time of the pulses. A spontaneously firing neuron may do so at average rates up to several tens of Hertz; for such a neuron, it is the *change* in its spontaneous firing rate (increase or inhibition) that constitutes its neural message.

Electrical impulses are not only generated in neurons. There is another fundamental type of "pseudo-neurons" with excitable membranes, the sensory *receptor cells*, in which electric potentials are triggered by some *physical* agent, like a photon (light particle) in the case of rod and cone cells of the retina or a flexing force acting on the cilia of the hair cells of the basilar membrane (see below). These neurons are the sensory *detectors* of the nervous system and represent the input extremity of the neuronal chain of an animal, feeding information on environment and body through the peripheral afferent pathway system to the information-processing machinery of the central nervous system. On the output extreme of this chain are the special synapses (with acetylcholine as transmitter substance) between axons of so-called motoneurons and the muscular fibers in muscles, glands, and blood vessels; these synapses, when activated, generate the contraction of motile proteins, collectively causing the contraction of these fibers.

There are three spatial domains of information processing in the nervous system. As discussed above, the neuron, represents the fundamental processing unit in the "microscopic domain." At the intermediate or "mesoscopic level", we have assemblies of a limited number of neurons wired to each other to accomplish a limited repertoire of specific tasks. For instance, in the retinal network of the eye, we find groups of neurons working as motion detectors or contrast enhancers; in the sensory receiving areas of the cortex, there are columnar assemblies of hundreds of thousands of neurons representing cooperative units which receive information from a common, limited stimulus region of the retina, basilar membrane, skin, etc. The "macroscopic domain" is, basically, the brain as a whole, which in a human has about $10^{11}$ neurons with $10^{13}$–$10^{14}$ synaptic interconnections.

At the meso- and macroscopic levels, there are two basic modes in which information is represented in the nervous system. One is given by the specific *spatio-temporal distribution of electrical impulses* in the neural network, representing the *transient* or *operating* state of the brain. The other is represented by the *spatial distribution of synapses* and their efficacies (usually called the "synaptic architecture"); this latter mode represents the actual "hardware" or *internal* state of the neural network. At the macroscopic level, the first mode is a horribly complex pattern that varies on a time scale of 0.1–100 s of milliseconds and usually involves millions of neurons even for the simplest information processing tasks. The second mode represents a spatial pattern that is constant or varying very slowly during learning experiences. Note that there is no equivalent to "software" in the neural system—the "programs" or operating instructions are embedded in the changing configuration of the neural hardware (synaptic architecture).

Unfortunately, none of the macroscopic neural patterns, dynamic or static, can be represented in any numerical form, as emphasized in Sect. 1.5. There are also insurmountable difficulties from the experimental point of view.[20] A single microelectrode implanted in a neuron (which, as we said before, does not necessarily impair the function of the cell) will give us readings of individual neural impulses, their frequency, and distribution in time, but it would take thousands of microelectrodes implanted in very close vicinity to each other to obtain a real "spatio-temporal distribution" of impulses in any one processing module of brain tissue. Larger-tip electrodes will provide measurements of electrical signals averaged over tens or hundreds of neurons and will tell us what the average activity of a limited region of neural tissue or nerve fiber is at any given time—but it still would not furnish any details of the exact spatio-temporal distribution of impulses.

Concerning the synaptic architecture, neuron-neuron interconnections can only be observed with a microscope, and although synaptic growth has actually been observed in real time (Hosokawa et al., 1995), only statistical results about synaptic architecture (like the synaptic density) have been obtained, and this in only very limited parts of neural tissue. For instance, one of the first quantitative studies of the density of synapses in brain tissue (Globus et al., 1973) revealed that rats raised in challenging environments have a higher spine count (axon-dendrite interconnections) than rats raised in "boring" environments. This represented a first confirmation of the so-called Hebb hypothesis (Hebb, 1949) concerning the *elementary neural information storage act*: Two originally independent neurons in mutual proximity which, for some externally controlled reason, tend to fire neural pulses simultaneously, will establish (grow, validate, potentiate) a synaptic contact between the two in such a way that, in the future, a pulse in only one of them will trigger a pulse in the other. In Appendix II, we shall show with a neural network model how this elementary mechanism works.

After this primer on the neural system in general terms, we should be in a better position to discuss how the neural system may collect and code information on acoustic vibration patterns. When the acoustic signal of a single pure tone of given frequency arrives at the ear, the basilar membrane oscillations stimulate the hair cells that lie in the resonance region corresponding to that frequency

[20]The problem is not unlike that of wanting to represent in a mathematically tractable way the exact position and velocity of all molecules in a gas: this is impossible from a purely practical point of view because of the sheer number of elements and the fact that there is no continuity of dynamic parameters between any two neighboring elements (see, however, Ashmore, 2008)! Statistical thermodynamics, developed by Ludwig Boltzmann almost 150 years ago, links mathematical *averages* of such molecular quantities with traditional thermodynamic state variables like temperature, pressure, entropy, etc. In a certain sense, computer assisted tomography like fMRI and PET does yield averages of spatial and temporal neural activity, but there is a fundamental difference with thermodynamics: while in a gas there are myriads of different distributions of molecules that all lead to the same dynamic averages, *each constellation* of microscopic neural activity in principle represents a *different* thought, a different image, different information.

(Sect. 2.3). In humans, these sensor units are grouped in a row of about 4000 "inner" hair cells (running along the basilar membrane from base to apex on the side of the modiolus, or center core of the cochlear "snail") and three rows of a total of about 12,000 "outer" hair cells (see Figs. 2.7(a) and (b)). When the stereocilia of a hair cell are deflected into a certain direction, electric impulses are triggered in spiral ganglion neurons that make synaptic contact with the sensor cell (or, in case of an inhibitory synapse, the neuron's spontaneous firing rate could be inhibited). The axons of these neurons form the afferent fibers of the auditory nerve; their action potentials (p. 58) collectively carry digitally encoded information on the motion of the basilar membrane to the central nervous system.

An important feature is the arrangement of the afferent nerve endings. Whereas a single nerve fiber usually contacts only one inner hair cell, thereby receiving messages from an extremely limited region of the basilar membrane, single afferent nerve fibers innervating the outer rows make contact with 10–50 sensor units spread over several millimeters, thus being able to collect information from a much wider resonance frequency domain. There are indications that the response of neurons wired to the inner-row hair cells is excitatory, whereas that from the outer row is inhibitory (Sokolich and Zwislocki, 1974). Inner hair cells respond to the *velocity* of the basilar membrane motion, because the deflection of their cilia is proportional to the velocity of the surrounding endolymphatic fluid (the force on an obstacle immersed in a viscous fluid is proportional to the velocity of the flow). Outer hair cells, in contrast, signal according to *displacement*, probably because their cilia are locked to the tectorial membrane (Fig. 2.7) (the interactive mechanical forces depend on the instantaneous distortion of the cochlear partition); their response saturates at high intensity levels. The fact that about 95% of the afferent fibers in the acoustic nerve terminate on inner hair cells, with 10–50 individual fibers making synaptic contact with each cell, clearly confers to the inner row the role of prime sensory receptor. On the other hand, however, the outer hair cells receive endings of efferent fibers that deliver neural impulses coming from the central nervous system;[21] this fact plus the remarkable motility of the outer hair cells, of fairly recent discovery, places the latter in a special dual role as effectors and receivers (to be discussed in detail in Sect. 3.6). While it is clear that each kind of hair cells plays a very distinct role in acoustic signal transduction, both types must work together: damage to the outer hair cells severely impairs hearing even if the inner row remains fully functional. We will continue discussing this subject in Sect. 3.6; for more detailed up-to-date information, see for instance Gelfand (1990), and Zwicker and Fastl (1999).

It has been found by implanting microelectrodes in acoustically activated cochlear nerve fibers, that a given fiber has a lowest firing threshold for that acoustic frequency $f$ which evokes a maximum oscillation at the place $x$ of the

[21] A small proportion of efferent fibers also act on inner hair cell output, but only indirectly: they synapse on the afferent fibers that are in contact with the inner hair cell.

basilar membrane (Fig. 2.8) innervated by that fiber. This frequency of maximum response is called the neuron's characteristic frequency or "best frequency" (Kiang et al., 1965; see also Sect. 3.6 and Fig. 3.17).

Turning now to the actual distribution in time of individual pulses, measurements (Zwislocki and Sokolich, 1973) have shown that the maximum firing rate is associated with the maximum *velocity* of the basilar membrane when it is moving toward the *scala tympani*; inhibition of the firing rate occurs during motion in the opposite direction, toward the *scala vestibuli*. Moreover, the instantaneous position of the basilar membrane has a (less pronounced) excitatory or inhibitory effect, depending on whether the membrane is momentarily distorted toward the scala tympani, or away from it, respectively. Both effects add up to determine the total response. Figure 2.22 shows a hypothetical time distribution of neural impulses in a nerve fiber of the inner ear connected to the appropriate resonance region of the basilar membrane, when it is excited by a low frequency tone of a trapezoidal vibration pattern shape (after Zwislocki and Sokolich, 1973).

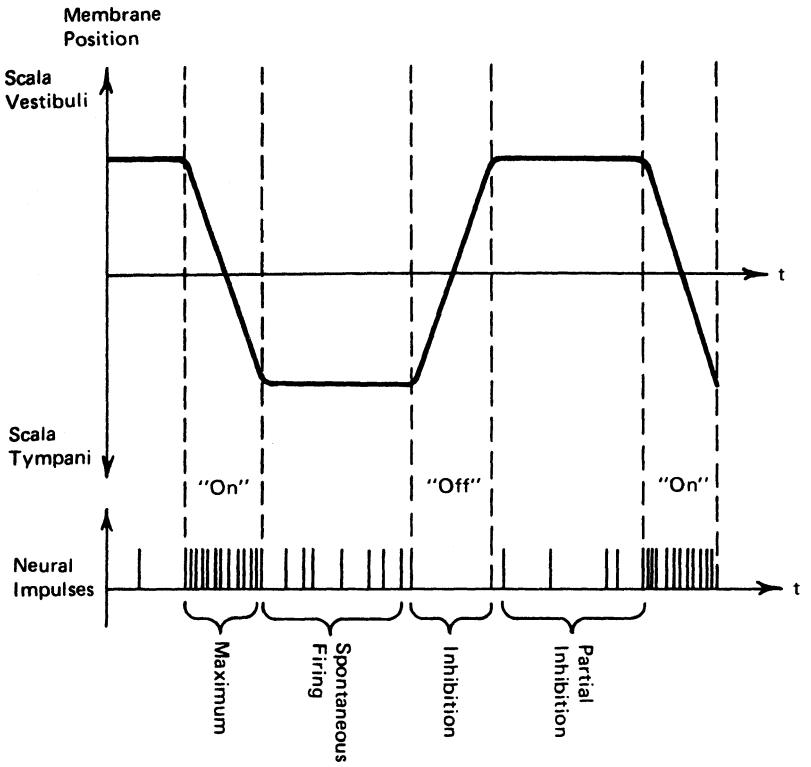



FIGURE 2.22 Sketch of neural impulse patterns in acoustic nerve fibers during the various phases of a trapezoid-shaped vibration pattern of the basilar membrane.

Close inspection of this figure reveals how information on repetition rate (actually, the repetition period) of the original acoustic signal can be coded in the form of "trains" of nerve impulses. Figure 2.22 would correspond to an ideal case of very low frequency. Actually, the acoustic frequencies are usually higher than those of neural firing rates, and the real situation corresponds rather to one in which the "on" and "off" intervals are much less clearly delineated because of their short duration (as compared to the refractory period of a typical neuron) and because of the random character of the impulse distribution. The only statistically important property is that there will be more impulses falling into "on" intervals than into "off" intervals. As a result, for pure tones, the time interval between successive impulses will tend to be an integer multiple of the sound vibration period $\tau$ (Kiang et al., 1965). It is clear that the higher the frequency of the tone, the less well-defined this grouping will be. For frequencies above a few thousand Hertz, it does not work at all. When several fibers receiving stimuli from a narrow region of the basilar membrane are bundled together (as occurs in the auditory nerve), the sum of their impulses (as detected by a macroelectrode that simultaneously makes contact with many fibers at the same time) will appear in synchrony with the auditory stimulus. These collective synchronous nerve signals have been called neural volleys.

## 2.9 Subjective Pitch and the Role of the Central Nervous System

It follows from the preceding section that a given neural fiber of the auditory nerve is capable of carrying two types of information:

1. The simple fact that it is firing (or that its spontaneous rate has been inhibited) tells the auditory system that the basilar membrane has been activated at or near the region innervated by that fiber—the spatial distribution (or "tonotopic" organization) of firing fibers thus encodes the information on *primary pitch*.[22] This process works for the whole range of frequencies but, for pure tone superpositions (as with the harmonics of a complex tone), only applies if their frequencies are separated by more than about a critical band.
2. The actual time distribution of impulses in that fiber may carry information on repetition rate or periodicity and also on details of the vibration pattern itself at the region innervated by the fiber (see below). This only works in the lower frequency range (below approximately 1500 Hz).

[22]The pitch of a pure (single-frequency) tone whose resonance region is innervated by the fiber in question (i.e., whose frequency is equal to the fiber's "best frequency."

There is no doubt that information as to the place of excitation is used by the auditory system at all levels (e.g., see Sect. 4.8). But how does this system utilize the information contained in the time distribution of neural pulses schematically shown in Fig. 2.22?

First, let us return for a moment to the perception of single, pure (sinusoidal) tones. Several arguments point to the fact that the time distribution of neural pulses is *not* utilized in the perception of the pitch of a single pure tone. For instance, theoretical calculations (Siebert, 1970) predict that if primary pitch were mediated by time cues, the DL of frequency resolution (e.g., see Fig. 2.9) would be independent of frequency (which it is not), and in turn should decrease with increasing stimulus amplitude (which it does not).

That time cues are largely ignored in primary pitch perception of pure tones may not come as a surprise. But what about perception of beats of mistuned consonances and periodicity pitch of harmonic complexes? It is difficult to find an explanation of beats of mistuned consonances and other phase-sensitive effects without assuming that, at some stage, a mechanism analyzes the temporal fine structure of the vibration pattern of the stimulus. Indeed, we may invoke the effect shown in Fig. 2.22 to attempt an explanation of how information on the vibration pattern and its variations (second-order beats) might be coded. Consider the superposition of two tones an octave apart. Assume that the resulting vibration pattern is that shown at the bottom of Fig. 2.16. Two resonance regions will arise on the basilar membrane, centered at positions $x_1$ and $x_2$ corresponding to the two component frequencies $f_1$ and $f_2 = 2f_1$ (Fig. 2.8). In the cochlear nerve bundle, we shall have two main foci of activity centered at the fibers with characteristic frequencies $f_1$ and $f_2$, leading to two primary pitch sensations, one octave apart. However, the resonance regions on the basilar membrane are rather broad, with sufficient overlap in the region between $x_1$ and $x_2$ where the points of the membrane will vibrate according to a superposed pattern somehow related to the original motion of the eardrum.[23] Fibers connected to that overlapping region will thus respond with firings that are grouped in "on" intervals of enhanced rate which, say, correspond to the descending (negative slope) portions of the second graph in Fig. 2.16. Note that in this case, the "on" intervals are not of equal duration but instead form an alternating "short—longer—short—longer" sequence. If the two tones were a fifth apart, the vibration pattern of the overlapping region could be that shown in Fig. 2.18, leading to a yet different type of sequence of "on" intervals. Periodicity of this sequence thus would represent the information on repetition rate, whereas structure of the sequence (a sort of "Morse code") would give information on the vibration pattern. Such a fine structure indeed has been identified statistically through electrophysiological measurements.

[23]Traveling waves in the cochlear fluid change their phase relationship and, of course, amplitude as they propagate, thus altering the actual form of the vibration pattern at different points of the basilar membrane.

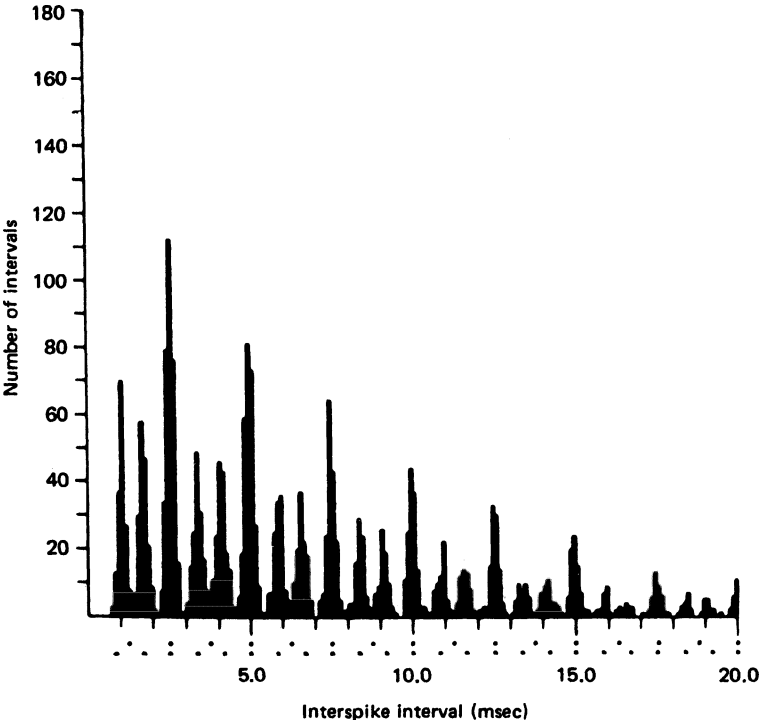



FIGURE 2.23 A histogram showing the number of times (vertical axis) a given interval between neural spikes occurs (horizontal axis), in an auditory nerve fiber stimulated with a two-tone superposition (fifth) with a given mutual phase relationship. Rose et al., 1969. By permission from the authors.

Figure 2.23 is an example (a so-called histogram) of the distribution of time intervals between neural pulses in an auditory nerve fiber (Rose et al., 1969), for a stimulus corresponding to a musical fifth in a given phase relationship. Note the difference in the relative number of times (vertical axis) a given interval between successive pulses (horizontal axis) appears. This represents the (statistical) "Morse code" mentioned above, carrying information on the vibration pattern. The greater the complexity of the original vibration pattern and the higher the frequency of the component tones, the more "blurred" the information conveyed by the pulse sequence will be, that is, the more difficult it is to be interpreted at the higher brain levels. Detailed analysis of the neural pulse time distribution would require the operation at some level of what is called a temporal autocorrelation mechanism (initially proposed by Licklider (1959)), in which a pulse "train" is compared with previous pulse trains, whereby similar repetitive features (such as the periods marked in Fig. 2.20(b)) are enhanced and all others (nonperiodic) are suppressed.

Time cues are also operative in the mechanism responsible for the sensation of spatial (stereo) *sound localization*[24] (e.g., Molino, 1974; Feeney, 1997). This binaural hearing should involve a process called temporal crosscorrelation of the neural signals from both cochleas in which the intra-aural time difference between the signals from both cochleas is determined. There is physiological evidence that such a mechanism exists (in the medial superior olive, Fig. 2.26). A neural model of a crosscorrelator has been proposed by Licklider (1959). In this model, t is assumed that an ascending neuron (Fig. 2.24) can fire only if it is excited simultaneously by both incoming fibers. Because a neural signal propagates with a finite velocity along a fiber (Sect. 2.8), simultaneous arrival at a given ascending neuron ending requires a certain time difference between the original signals in both cochleas. For instance, exact simultaneity (zero time difference) of both cochlear signals would fire the ascending neuron located exactly in the center, because that is the place at which both right and left signals meet. If, however, the original signal is detected first in the right ear, its pulse will travel past the middle point until it meets the delayed pulse from the left ear. It is easy to see that the location $y$ (Fig. 2.24) of the activated ascending neuron will depend on the interaural time delay, which in turn depends on the direction of the incoming sound.

The scheme in Fig. 2.24 is an oversimplified model. Detailed studies or the neuranatomy of the superior olivary complex (Fig. 2.26) reveal a more complicated structure. Binaural information is actually coded through a complex inter-

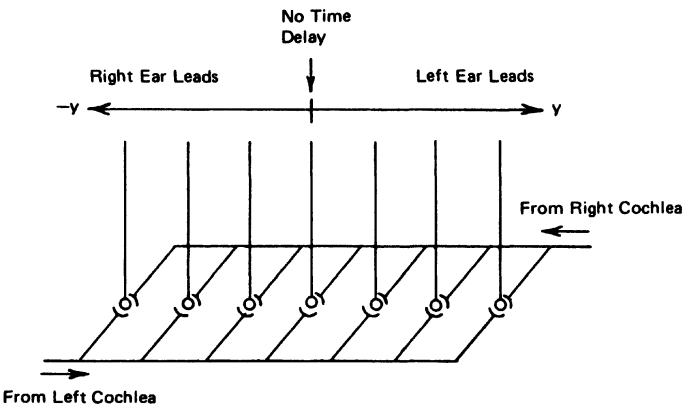



FIGURE 2.24 Simplified model for a neural crosscorrelation mechanism (interaural time-difference detector). After Licklider (1959).

[24]Intensity cues (amplitude difference between the sound waves arriving at the two ears) and spectral cues (timbre difference) also contribute to sound localization, especially at high frequencies and in closed environments.

action of excitatory and inhibitory inputs which are the result of time (phase) and intensity differences between the stimuli reaching both ears (Goldberg and Brown, 1969). Whatever the actual mechanism of the crosscorrelator, its time resolution abilities are astounding: humans can locate sound sources in space on the basis of interaural timing differences of less than 20 microseconds! This is just a tiny fraction of the duration of one action potential. The unusual innervation of inner hair cells, with more than 20 fibers making contact with a single sensor cell, may be required to assure coherent neural information transmission at the required rate (Hudspeth, 1989).

Two tones, a *mistuned interval* apart, fed into separate ears, may "foul up" the crosscorrelator. The gradually shifting phase difference between the two tones (e.g., Fig. 2.16) will be interpreted by this mechanism as a changing difference in time of arrival of the left and right auditory signals, hence signaling to the brain the sensation of a (physically nonexistent) cyclically changing sound direction! This is why two pure tones forming a mistuned consonant interval, presented dichotically with headphones, give the eerie sensation of a sound image that seems to be "rotating inside the head" (p. 49).

Thus far, we have been discussing neural mechanisms whose main purpose is the determination of sound wave *patterns in time*, i.e., the identification and measurement of features like wave peaks and valleys (Fig. 2.20), up-and-down slopes (Figs. 2.16 and 2.22) and differences in times of arrival (Fig. 2.24). In essence, such mechanisms analyze the statistical distribution of neural impulses (see Fig. 2.23) fed via the acoustic nerve into the brain. For the past decades, an unresolved issue has been the question of whether or not such sequence-analysis of neural pulses is a *necessary* hypothesis to explain periodicity pitch perception (e.g., Yost and Watson, 1987). A temporal autocorrelation mechanism with its potential capability of detecting the repetition rate of neural signals could indeed explain many important psychoacoustic features of pitch perception such as beats of mistuned consonances and fundamental tracking, but not all, such as spectral pitch perception (see below). If it is not a time-cue analysis, what is the mechanism that enables us to attach a single pitch to a harmonic tone complex—even if the fundamental is not present in the original stimulus? Why at all do we perceive the pitches corresponding to the frequencies given by relations (2.7a)–(2.7c) when a melody is played with the corresponding harmonic two-tone complexes?

Some early ideas that may lead toward an explanation of these effects (e.g., Terhardt, 1972; Wightman, 1973; Goldstein, 1973) are presented here in a highly simplified way. "Natural" sounds of human and animal acoustic communications contain an important proportion of harmonic tones (vowels, bird song, animal calls). Such tones share a common property. They are made up of a superposition of harmonics, of frequencies $nf_1$, integer multiples of a fundamental $f_1$ (p. 52). These tones elicit a complicated resonance pattern on the basilar membrane, with multiple spatial amplitude peaks, one for each harmonic (Fig. 2.25(a)). In spite of its complexity, this *spatial pattern* does bear some invariant characteristics. One such invariance is the particular distance relationship

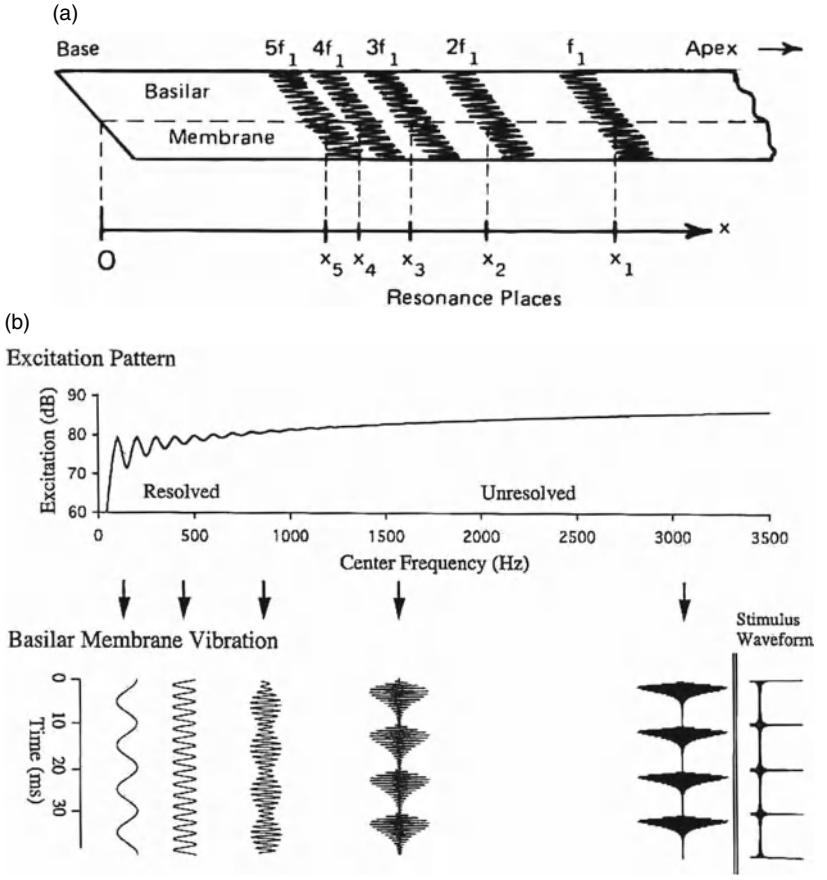



FIGURE 2.25 (a) Sketch of the resonance regions on the basilar membrane elicited by a complex tone. (b) Computer-simulated basilar membrane excitation pattern (top) and local vibration pattern (bottom) (Plack and Oxenham, 2005).

between neighboring resonance maxima.[25] Notice in Fig. 2.25(b) (Plack and Oxenham, 2005), which presents a more realistic model of the excitation pattern of a multiharmonic tone, how the resonance zones run into each other for higher frequencies, and how the local temporal vibration pattern acquires an increasingly superposed and therefore complex characteristic. In the lower harmonic order ($n < 7$–$9$), each harmonic excites a distinct place on the basilar membrane (the pattern that a *space-based mechanism* must recognize), which vibrates with

[25] Another invariant characteristic is the high coherence of the macroscopic time variations of this complicated excitation pattern over the whole spatial domain of the basilar membrane.

simple periodic motion at the harmonic's frequency; phase relationships between harmonics are immaterial in this domain. For higher $n$-values, the resonance regions become unresolved and the membrane oscillation pattern becomes complex (the pattern that a *time-based mechanism* must recognize), depending on the partials' intensities and mutual phase relationships.[26]

We either learn at an early age (Terhardt, 1972, 1974), or we have a built-in mechanism (Wightman, 1973; Goldstein, 1973), to recognize the invariant characteristics of the spatial pattern of excitation by a complex tone as belonging to "one and the same thing." We shall call this mechanism of recognition the *central pitch processor*. The main function of this neural processing unit is to transform the physical excitation pattern along the basilar membrane into a neural activity pattern in such a way that all stimuli with the same periodicity are similarly represented. In this picture, the local temporal pattern of vibration is not considered—it is the envelope along the spatial extension of the basilar membrane that counts. According to the definition of *pragmatic information* (Sect. 1.6), the correspondence between the spatial input pattern and the univocally evoked neural output activity in some very specific regions of the brain (Sect. 5.6) represents the *sensory information* which we call pitch. The pitch perceived is that of the fundamental component $f_1$, which, as mentioned above, is usually the most prominent one (intensity-wise) in environmentally important natural sounds—but needs not to be present in the original tone to be heard.

All of this should work much in analogy with visual pattern recognition. For instance, when you look at the symbol **Ш**, it may not convey any "unique" meaning at all (your interpretation would probably depend upon the spatial orientation of the symbol and the context in which it is shown). But anyone familiar with the Cyrillic alphabet clearly perceives it as just "one thing" (the letter "shch"), no matter where in the visual field, and in what orientation, it is projected.

One usually states that we have built into our central processing system basic "templates" with which to compare the complex structures of spatial excitation pattern of the basilar membrane. Whenever a match is achieved, a unique pitch sensation is elicited. This matching process works even if only a partial section of the excitation pattern is available. If instead of a natural complex sound, we are exposed to one in which some normally expected elements are suppressed (e.g., a missing fundamental), the partially truncated excitation pattern on the basilar membrane fed into the recognition mechanism of the pitch processor may still be eventually matched, within certain limitations. Again, we find many analogies in visual pattern recognition. Observe the nonexisting—but expected—contours in the following letters:

[26] Because of this increasing overlap, the basilar membrane oscillation is treated in mathematical models or equivalent electronic circuits as the action of a series of *filters*, which in frequency space are narrow at the membrane's apical end (Fig. 2.6(b)) (letting through only one harmonic), and wider at the basal end (accepting several neighboring harmonics).

Music

The above-described matching process works even if the alternate harmonic components of a tone are fed into *separate* ears (e.g., Houtsma and Goldstein, 1972). This obviously means that the central pitch processor must be located in some upper stage of the auditory pathway, after the input from both cochleas has been combined. Furthermore, the matching process works even if only *two* neighboring harmonics of a complex tone are presented as seen in Sect. 2.8. In such a case, however, the matching mechanism may commit errors—and lock into one of several "acceptable" positions.

Neither of the older pitch perception theories proposes how the key algorithms are actually carried out by the pitch processor in the nervous system. Yet neuronal networks are known to exist that are perfectly capable of performing the operations of neural pulse summation and multiplication required in the execution of the necessary algorithms. Terhardt's theory (1974) comes closest to proposing a neural wiring scheme. Indeed, its computations are based on a learning matrix, an analog circuit that "learns to respond"[27] to characteristic features of the most frequently occurring input configurations (i.e., to the distance relationships between input excitation maxima caused by a complex tone). We shall return to these theories later, when we discuss explicitly the perception of complex musical tones (Sect. 4.8 and Appendix II), and consonance and dissonance (Sect. 5.2).

Finally, let it be stated that one cannot exclude the possibility that at least partial use is made of the time distribution of neural pulses, in the pitch perception of complex tones. It is hard to believe that the nervous system, always geared to work with such an astounding efficiency, with so many backup systems, is not taking advantage of the handy "Morse code" information (Fig. 2.22) that actually exists in peripheral auditory transmission channels! We already mentioned that one should expect the spatially based pitch mechanism to work for lower harmonic numbers (where neighboring resonance regions are well separated from each other, Fig. 2.25), whereas a temporally based mechanism can only operate in the upper harmonic range, where unresolved harmonics lead to a basilar membrane vibration pattern that approximates that of the acoustic input. Some psychoacoustic experiments indeed seem to demand an explanation via time-cue analysis. For instance, low frequency pure tones of very short duration (two to three actual vibration cycles) can give rise to a

[27]This was accomplished in the laboratory model by appropriately decreasing electrical resistances between transmission lines (the rows and columns of the matrix) that are simultaneously activated (conduce current) by a given, repeatedly presented, input configuration. Today neural networks are numerically modeled with a computer (to be discussed in Appendix II).

clear pitch sensation (Moore, 1973). Or, if an acoustic signal (white noise) is presented to one ear, and the same signal is fed into the other ear delayed by an interval $\tau$ (a few milliseconds), a faint pitch corresponding to a frequency $1/\tau$ is perceived (Bilsen and Goldstein, 1974). Neither of these results could be explained satisfactorily by a "place theory" (spatial cue analysis). For comprehensive comparisons of the "place theory" with the "time theory" of pitch perception, see Lin and Hartmann (1998) and de Cheveigné (2005).

One thing is clear from the preceding discussion. Subjective pitch perception requires that "higher order" operations of pitch extraction be performed in the central nervous system after the input from both cochleas has been combined. For this reason, we conclude this chapter with a summary description of some of the most relevant features of the auditory pathway (Brodal, 1969; Gelfand, 1990). This will also serve as a reference for discussions to come in later chapters. The anatomic exploration of neural pathways and their interconnections is an extremely difficult experimental task. Neurons are cells the processes of which (axons or dendrites) can be many centimeters long; each neuron, especially in cortical tissue, may receive information from thousands of cells while it relays information to hundreds of others. It is nearly impossible to determine the connection pattern microscopically for a given cell. Anatomical exploration of these channels was done in the past with staining techniques in vitro (in brains of cadavers). Today it is possible to explore the functional aspects in living animals with microelectrode recordings, and in human subjects with the noninvasive tomographic techniques fMRI and PET already mentioned briefly in Sect. 1.5.

The noninvasive techniques with external electrodes in electroencephalography, or the small SQUID magnetometers in magneto-encephalography, while only providing information on activity averaged over centimeter-wide areas of the brain, have the advantage of being fast, and thus can be used conveniently for timing estimates of the order of milliseconds. This is very useful for the study of most cortical areas (those located directly under the cranium), but they are impractical for the auditory pathways which are buried deep in the brainstem (magneto-encephalographic recordings can be done through the mouth, which requires some brave volunteers). This problem has been superceded now with tomographic techniques such as the dynamic functional magnetic resonance methods fMRI and positron emission tomography PET. Since what is being imaged in an fMRI is increases in blood flow and oxygenation, the latter yield a time-dependent picture of neural activity (Moonen and Bandettieri, 1999) in millimeter-size regions and on the scale of a few seconds and more. However, corrections for real-time cardiac pulsations are necessary (each image has to be taken at a specific phase of the heart cycle), and for the auditory channel the situation is complicated by the substantial and unavoidable noise of the imaging equipment (cooling system, magnetostriction noise) which often interferes with what is being measured.

Figure 2.26 shows diagrams of the afferent (incoming) auditory pathway from the cochlea to the auditory receiving area of the cerebral cortex inflow chart form. The figure depicts information-transmission channels and relay-processing stations, and bears no scale relationship whatsoever to the actual neuroarchitectonic

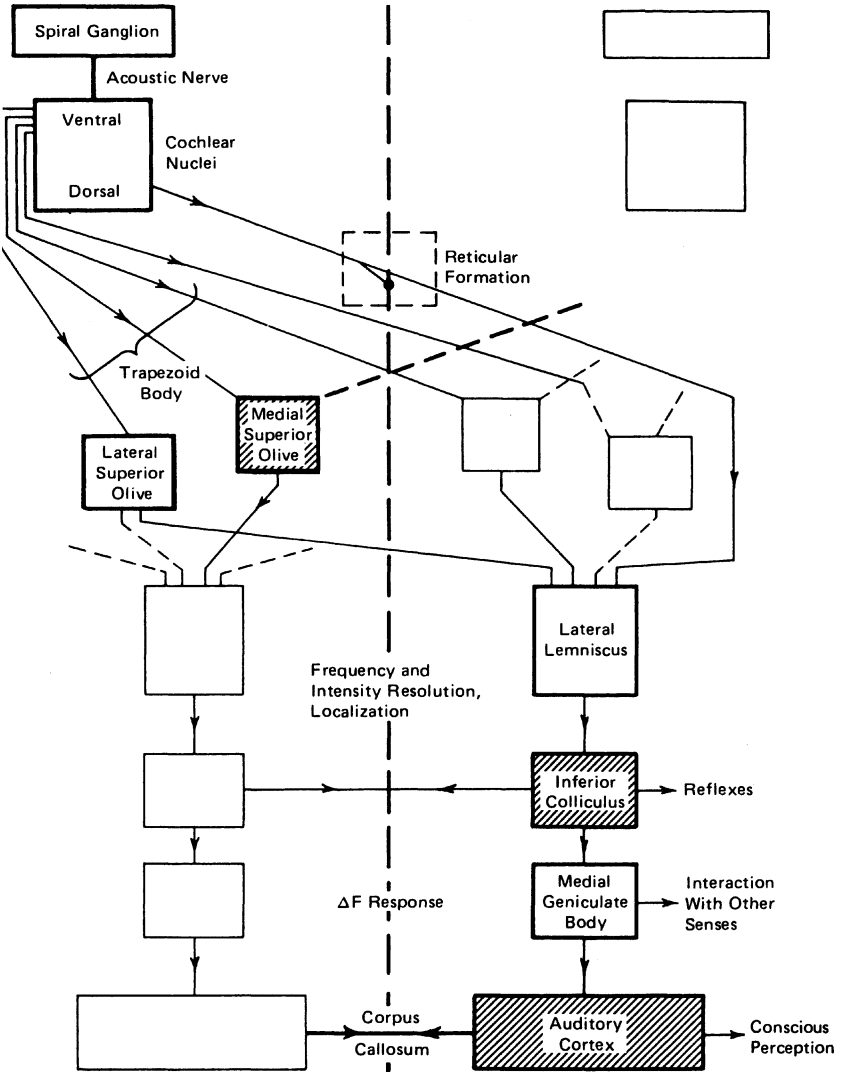



FIGURE 2.26 Flow chart of neural signals in the auditory pathway from one ear through the brain stem to the auditory cortices.

picture. The *spiral ganglion* is a neural network in the cochlea, the first processing stage in this pathway. It is here that neurons contacting inner and outer-row hair cells have a first chance to interact, determining the particular spatio-temporal distribution of activity in the acoustic nerve (the VIII cranial nerve), which conveys this information to the brain. The next processing stage, located in the medulla oblongata, comprises the *cochlear nuclei*, composed of three subdivisions whose elaborate structure is responsible for the first steps of sound resolution and

discrimination tasks. From here, neural information is channeled into three main bundles. One crosses over directly to the opposite contralateral side and enters the *lateral lemniscus*, the main channel through the brain stem (pons). Some fibers terminate in the *reticular formation*, a diffuse network in the brain stem that plays the role of a major cerebral "switchboard."[28] Another bundle (the trapezoid body) sends fibers from the ventral cochlear nucleus to important relay and processing stations, the *lateral and medial superior olives*. Of these, the medial superior olive is the first intra-aural signal-mixing center. This is the place at which a crosscorrelator (Fig. 2.24) may yield the information necessary for sound source localization and the sensation of beats of mistuned consonances (Feeney, 1997). Finally, a third intermediate bundle leads from the ventral cochlear nucleus to the contralateral olivary complex.

The three upper stages (Fig. 2.26) involve the *inferior colliculus*, the *medial geniculate body*, and finally, the primary areas of the *auditory cortex* (Heschl's gyrus in the temporal lobes). Some fibers are connected to the superior colliculus, which is also innervated by visual pathways, which means that there is intrasensory mixing already at a subcortical level. This may contribute to synesthetic effects, such as visual illusions while hearing certain sounds (e.g., seeing colors while listening to music), or acoustic illusions while seeing luminous images (e.g., hearing sounds while watching the polar aurora). Note the interconnections at these various stages with the contralateral pathway and with other sensory pathways and brain centers.

Not shown in Fig. 2.26 is a network of *efferent fibers*, which carries information from the upper stages down to lower ones and terminates in the cochlea. This system plays a role in the control of incoming afferent information. The lower tract of the efferent network, the *olivocochlear bundle*, may participate in an important way in the sharpening process (p. 41). Although there are only about 1600 efferent fibers reaching each cochlea, the larger fibers innervate profusely the outer hair cells and thus can exert a central control on the latter's mechanical (motility) and/or electrical operation (see Sect. 3.6).

Finally, let us point out some generalities that may be useful for later chapters. At the initial stages of the acoustic system (cochlear nucleus), there is a very specific geometric correspondence between activated neural fibers and the spatial position of the source stimulus (resonance regions) on the basilar membrane. As one moves up, however, this tonotopical correspondence is gradually lost (except in an anesthetized state). The number of participating neurons increases dramatically and the neural response becomes increasingly representative of complex features of the sound signal, being more and more influenced by feedback

[28]This structure, which receives raw data from the senses and the body as well as elaborate information from the cortex, consists of several small nuclei in the back of the brainstem, and is responsible for activating or inhibiting cerebral information-processing according to instantaneous needs, controls sleep, arousal, awareness, and even consciousness, and influences many visceral functions. Damage to components of this diffuse network, as happens in a sharp blow to the back of the neck, can lead to irreversible coma.

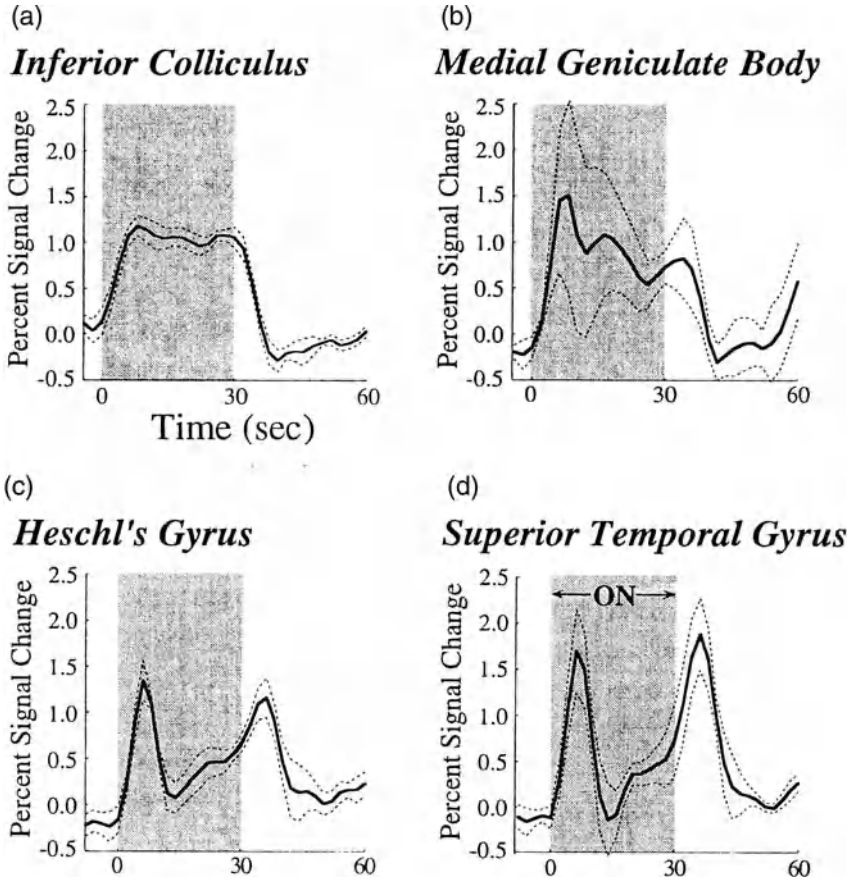



FIGURE 2.27 Neural responses at different afferent levels to a 30second noise burst fed into an ear (Melcher et al., 1999). (a) Inferior colliculus; (b) Medial geniculate body; (c) Heschl's gyrus; (d) Superior temporal gyrus.

information from higher levels on the behavioral state and the performance of the individual. This is borne out clearly in the fMRI responses at different levels of the auditory pathway. For instance, measuring the percent signal changes of neural activity when a 30secondnoise burst is fed into an ear (Fig. 2.27 Melcher et al., 1999), the activity is nearly constant during the noise stimulus in the inferior colliculus (see Fig. 2.26), with specific rise and decay slopes; higher up in the medial geniculate body, there is already a marked change during the (constant) noise period, indicating some influence of feedback information from other brain centers. At the cortical level (Heschl's gyrus and the superior temporal gyrus), the responses are clearly onset and offset signals ("here come the noise!" and "it's over!"). Contralateral (i.e., crossing) channels are "better" information carriers than are ipsilateral channels (to the same side)—if conflicting information is

presented to both ears, the contralateral channel tends to override the information that is carried to a given hemisphere by the ipsilateral channel (Milner et al., 1968).

At the stage of the inferior colliculus, good resolution of frequency, intensity, and direction of sound already exists; so does a selective response to up-down frequency sweeps. Reflexes work, but at this stage, there is no evidence of conscious perception of sound, as ablation experiments have shown. In the medial geniculate body (and probably in the superior colliculus), some pattern recognition capability is already operational. At this stage, information exists on where a given sound source is and where it is going in space and time. The first integration with information from other senses takes place.

The last stage of incoming information processing is performed in the auditory receiving area of Heschl's gyrus. From here on, the information is distributed to other brain centers, where it is analyzed, integrated into the whole cognitive functions of the brain and stored—or discarded as irrelevant. The corpus callosum, a gigantic commissure of about 200 million fibers connecting both cerebral hemispheres (Fig. 2.26), plays a key role in global information-processing, especially in view of the remarkable specialization of the two hemispheres, as already mentioned in Sect. 1.5. We shall return to this topic in more detail in Sect. 5.7.

# 3
# Sound Waves, Acoustic Energy, and the Perception of Loudness

*"First of all, I must ascertain if the instrument has good lungs"*

Johann Sebastian Bach (1685–1750)—his usual joke before pulling all stops in the test of a new organ

In the preceding chapter, we studied simple sound *vibrations* and their subjective effects, without investigating how they actually reach the ear. We referred to experiments in which the sound source (headphone) was placed very close to the eardrum. In this chapter, we shall discuss the process of sound energy *propagation* from a distant source to the listener and analyze how this acoustic energy flux determines the sensation of loudness. We will end the chapter with yet another, more detailed, look at that electromechanical marvel, the cochlea.

## 3.1 Elastic Waves, Force, Energy, and Power

When sound propagates through a medium, the points of the medium vibrate. If there is no sound at all and if there is no other kind of perturbation, each point of the medium[1] will be at rest and remain so until we do something to the medium. The position in space of a given point of the medium when the latter is totally unperturbed is called the *equilibrium position* of that point.

Sound waves are a particular form of so-called *elastic waves.* Whenever we produce a sudden deformation at a given place of a medium (e.g., when we hit a piano string with the hammer or when we suddenly displace air by starting the motion of the reed in a clarinet), elastic forces will cause the points close to the initial deformation to start moving. These points, in turn, will push or pull through elastic forces onto other neighboring points passing on to them the order to start moving, and so on. This "chain reaction" represents an elastic wave propagating away from the region of the initial perturbation. *What* propagates away with this wave is not matter but *energy:* that energy needed to put in motion each point reached by the wave. Sound waves of interest to music are elastic waves in which the points execute motions that are periodic (Sect. 2.1). In its vibration, each point of the medium always remains very, very close to the

[1] A "point" of the medium is meant here in the macroscopic sense, still encompassing billions of molecules!


J.G. Roederer, *The Physics and Psychophysics of Music*,
DOI: 10.1007/978-0-387-09474-8_3, 

equilibrium position. A sound wave propagates with a well-defined speed away from the source in a straight line, until it is absorbed or reflected. The way sound waves propagate, are reflected, and absorbed determines the acoustic qualities of a room or concert hall.

We have mentioned the concepts of force and energy. We must now specify their precise physical meaning. Everybody has an intuitive notion of *force:* the pull or the push we have to apply to change a body's shape, to set an object in motion, to counteract gravity to hold a body in our hand, to slow a motion down, etc. But physics is not satisfied with intuitive concepts. We must give a clear definition of force, as well as the "recipe" of how to measure it. Both definition and recipe must be based on certain experiments whose results are condensed or summarized in the formulation of a physical law.

It is our daily experience that, in order to change the form of a body, we have to do something quite specific to it: we have to "apply a force." Deformation, that is, a change in shape, is not the only possible effect of a force acting on a body. Indeed, it is also our daily experience that in order to alter the motion of a body, we must apply a force. Quite generally, it is found that the acceleration $a$ of a body, representing the rate of change of its velocity caused by a given force $F$ is proportional to the latter. Or, conversely, the force is proportional to the acceleration produced: $F = ma$. This is called Newton's equation. The constant of proportionality $m$ is the *mass* of the body. It represents its "inertia" or "resistance" to a change of motion. If more than one force is acting on a body, the resulting acceleration will be given by the *sum* of all forces. This sum may be zero; in that case, the acting forces are in *equilibrium.*

The unit of force is defined as that force needed to accelerate a body of 1 kg at a rate of 1 $m/s^2$ (increase its speed by one meter per second each second). This unit of force is called the *Newton.* One Newton turns out to be equal to 0.225 pounds force. The pound is a unit of force (weight) still in use in the USA. Since the acceleration of gravity is 9.8 $m/s^2$, the weight of a body of 1 kg mass turns out to be 9.8 Newton (= 2.2 lbs). We can measure a force by measuring the acceleration it imparts to a body of given mass or by equilibrating it (i.e., canceling its effect) with a known force, for instance, the tension in a calibrated spring.[2]

In many physical situations, a given force is applied or "spread" over an extended surface of the body. For instance, in a high-flying aircraft with a pressurized cabin, the air inside exerts a considerable outward-directed force $F$ on each window (and on any other part of the hull), that is proportional to the surface of the window $S$. The relation $p = F/S$ represents the *air pressure* inside the cabin. In general, we define the air pressure as the ratio between the force acting on a surface S that separates the air from vacuum. If, instead of vacuum, we merely have a different pressure $p'$ on the other side of the surface, the force $F$ acting on $S$ will be given by

[2]"Calibrated" means that we have previously determined how much the spring stretches for a given force, for example, a given weight.

$$F = (p - p')S \tag{3.1}$$

All this is very important for music. Sound waves in air are *air pressure oscillations.* Thus, if in relation (3.1), $S$ corresponds to the surface of the eardrum, $p'$ is the (constant) pressure in the middle ear, and $p$ the oscillating pressure in the meatus (Fig. 2.6), $F$ will be the oscillating force acting on the eardrum, responsible for its motion and that of the bone chain in the middle ear.

The pressure is expressed in Newton per square meter, or Pascal. The normal atmospheric pressure at sea level is about 100,000 Newton/m$^2$ (= 1000 Hectopascal). A more familiar unit in the United States is pound per square inch (for instance, the overpressure in auto tires is usually quoted in these units). Converting Newtons into lbs and m$^2$ into sq in, we obtain 1 Newton/m$^2$ = 0.00015 lb/sq in. Normal air pressure at sea level is thus about 15 lb/sq in.

We now turn to the concept of *energy.* Again, we do have some intuitive idea about it—but our intuition may easily fool us in this case. For instance, some people are tempted to say that "it requires a lot of energy to hold a heavy bag for a long time"—yet for the physicist, no energy is involved (except during the act of lifting the bag or putting it down). For the physiologist, on the other hand, a continuous flow of chemical energy to the muscles is necessary to maintain a continued state of contraction of the muscular fibers. To avoid confusion, it is necessary to introduce the concept of energy in a more precise, quantitative way.

The concept of force alone does not suffice for the solution of practical problems in physics. For instance, we need to know for how long, or over what distance, a given constant force has been acting, if we want to determine, say, the final speed acquired by a body accelerated by that force (even the largest force may have only a small *end* effect, if the duration or the path of its action was very short). As a matter of fact, for a material point subjected to a constant force $F$, what really counts to determine a given change of speed, say from a value 0 to $v$, is the product of *force times distance traveled* in the direction of the force. If we call $x$ that distance, it can be shown mathematically, based on Newton's equation, that $Fx = mv$. The product $Fx$ is called *work* and is counted positive if the displacement $x$ is in the same direction as the force $F$. The product $1/2mv^2$ is called the *kinetic energy* of the body of mass $m$. If $Fx$ is positive, we interpret the above relation by saying that the work of the force has increased the kinetic energy of the body, or, equivalently, that "work has been delivered to the system," increasing its kinetic energy from zero to $1/2mv^2$.

Work and kinetic energy are measured in Newton times meter. This unit is called *Joule,* after a British physicist and engineer. A body of 1 kg (unit of mass), moving at a speed of 1 m/s thus has a kinetic energy of 0.5 J. If it moves with twice that velocity, its kinetic energy will be four times larger: 2 J. An average person (70 kg), running at a speed of 3 m/s (6.75 miles/h) has a kinetic energy of 315 J; that of a 2000 kg car cruising at 30 m/s (67.5 miles/h) is 900,000 J.

Energy can appear in forms other than kinetic. Consider a body attached to a spring. We have to supply a given amount of work in order to compress the spring. If we do this very, very slowly, practically no kinetic energy will be involved.

Rather, the supplied work will be converted into *potential energy*; in this case, *elastic* potential energy of the body attached to a compressed spring. Releasing the spring, the body will be accelerated by the force of the expanding spring and the potential energy will be converted into kinetic energy. We may say that potential energy is the *energy of position* of a body, kinetic energy its *energy of motion.*

The sum of potential and kinetic energy of a body is called its total *mechanical energy* (there are many other forms of energy which we will not consider: thermal, chemical, electromagnetic, etc.). There are important cases in which the mechanical energy of a body remains constant. A "musically" important case is the previous example of a body attached to a spring oscillating back and forth under the action of the elastic force of the spring. It can be shown that the resulting vibration about the equilibrium position is *harmonic* (provided the amplitude remains small). When the body is released from a stretched position, its initial kinetic energy is zero. But it possesses an initial elastic potential energy which, as the oscillation starts, is converted into kinetic energy. Whenever the body is passing through its equilibrium position, the elastic potential energy is instantaneously zero while its kinetic energy is maximum. During the harmonic oscillation, there is back-and-forth conversion of potential energy into kinetic, and vice versa.

The total mechanical energy remains constant as long as no "dissipative" forces are present. Friction causes a continuous decrease of total energy and hence a decrease of the amplitude of oscillation. The resulting motion is called a *damped oscillation.* It is extremely important in music. Indeed, many musical instruments involve damped oscillations; a vibrating piano string is a typical example. Other external forces may act in such a way as to gradually increase the mechanical energy. They can be used to compensate dissipative losses and thus maintain an oscillation at constant amplitude. A bowed violin string is a typical example: the forces that appear in the bowing mechanism feed energy into the vibrating string at a rate that is equal to the rate of energy loss through friction and acoustic radiation (Sect. 4.2).

Now we come to a last, but utmost important point concerning energy. Machines (and humans) deliver energy *at a given rate.* Any machine (or human) can perform an almost arbitrarily large amount of work—but it would take a very long time to do so! What really defines the quality or power of a machine is the *rate* at which it can deliver energy (i.e., perform work). This rate, if constant, is given by

$$P = \frac{\text{work done}}{\text{time employed}} = \frac{W}{(t_2 - t_1)} \tag{3.2}$$

$W$ is the work delivered between the times $t_1$ and $t_2$. $P$ is called the mechanical *power.* It is measured in units of J/s, called *Watt* (another British engineer). If you are walking up a staircase, your body is delivering a power of about 300 W; the electrical energy consumed per second by an electric iron is about

1000 W, and the maximum power delivered by a small car engine is 30 kW (1 horsepower = 0.735 kW). A trombone playing fortissimo emits a total acoustic power of about 6 W.

The concept of power is most important for physics of music. Indeed, our ear is not interested at all in the total acoustic energy which reaches the eardrum—rather, it is sensitive to the *rate* at which this energy arrives, that is, the acoustic *power.* This rate is what determines the sensation of *loudness.*

## 3.2 Propagation Speed, Wavelength, and Acoustic Power

After the excursion into the field of pure physics in the preceding section, we are in a better condition to understand the phenomenon of wave propagation. To that effect, we make use of a *model* of the medium. We imagine the latter as made up of small bodies of given mass, linked to each other with compressed springs (representing the elastic forces). Initially, the spring forces are in equilibrium and all points are at rest. Figure 3.1 shows the situation when point $P$ has been suddenly displaced an amount $x_1$ to the right.

Considering the forces shown in Fig. 3.1, we realize that both points $Q$ and $R$—which initially are at rest at their respective equilibrium positions—are subject to a *resultant* force acting toward the right. In other words, according to Newton's equation, they will be accelerated to the right and start a motion into the same direction in which $P$ had originally been displaced. This point $P$, on the other hand, will be on its way back to its equilibrium position, accelerated by a resultant force that acts on it toward the left (Fig. 3.1). A short time later, when points $Q$ and $R$ are on their way toward the right, the compression of the spring between $R$ and $T$ starts increasing, whereas that of the spring between $Q$ and $S$ decreases. It is easy to see that both points $S$ and $T$ will start being subjected to a net force directed to the right that will cause them, in turn, to start moving to the right, while $Q$ and $R$ may be on their way back to the left. This process goes on and on, from point to point—representing a wave propagating away from $P$ toward both sides. The wave "front" is nothing but an order that goes from point to point telling it: "Start moving to your right." The order is given by the compressed springs (their elastic forces). At no time is there any net transport of matter involved. We call this case a *longitudinal wave,* because the displacements of the points are directed parallel to the direction of propagation of the wave. In the real case of

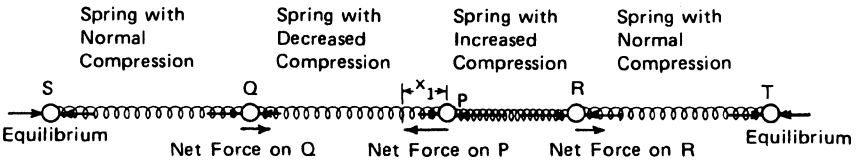



FIGURE 3.1 One-dimensional model of an elastic medium (springs in compression), in which point $P$ has been displaced longitudinally.

a sound wave propagating through air, the concerted action of the spring forces acting on points *P, Q, R,...* roughly corresponds to the air pressure; variations of these forces (e.g., variations of the distance between points) correspond to the *air pressure variations* of the sound wave.

The one-dimensional model of Fig. 3.1 also shows how energy transport is involved in an elastic wave. In the first place, we have to provide work from "outside" to produce the initial displacement $x_1$ of point *P* because we have to modify the lengths of the two springs *PQ, PR.* In other words, we need an energy source. In this case, the initial energy is converted into the form of potential (positional) energy of point *P.* Then, as time goes on, points to the right and to the left of *P* start moving, and the lengths of their springs change. All these processes involve energy both kinetic (motion of the points) and potential (compression or expansion of springs). The energy initially given to *P* is transferred from point to point of the medium, as the wave propagates: we have a *flow* or transport of energy away from the source.

Let us now turn to the case in which the springs in the model are in tension (expanded) instead of being compressed, with neighboring point *pulling* on each other. Physically, this corresponds to a tense violin string. For longitudinal displacements (in the direction along the springs) we obtain a qualitatively similar picture for wave propagation as before, only that all forces shown in Fig. 3.1 are now reversed. But, in addition, we have an entirely new possibility that does not exist for the case of compressed springs: we may displace point *P perpendicularly* to the *x* direction (Fig. 3.2) and obtain a different type of wave. Since all spring forces now pull on the points, according to Fig. 3.2, the resultant force $F_p$ will accelerate *P* down to its equilibrium position *O.* Points *Q* and *R* in turn would be subject to net forces that would accelerate them upward, in a direction essentially perpendicular to *x.* This represents a *transverse* elastic wave, propagating to the right and to the left of *P.* In a transverse wave, the displacements of the points are perpendicular to the direction of propagation. In a medium under tension, like a violin string, *two* modes of elastic wave propagation may occur simultaneously: transverse and longitudinal.

We now turn to the expression for the *speed of propagation* of transverse waves. It can be shown by applying Newton's law to the individual points of the unidimensional model of Fig. 3.2 that, for a string under a tension *T* (in Newton),

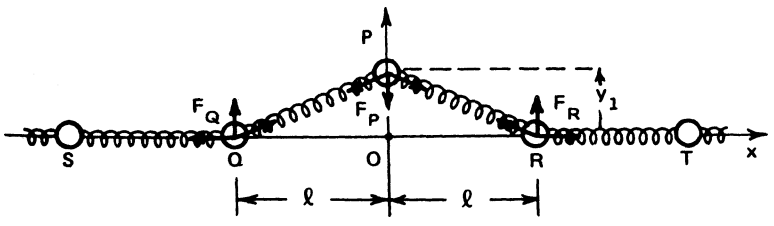



FIGURE 3.2 One-dimensional model of an elastic medium (springs in expansion), in which point *P* has been displaced transversally.

the velocity $V_T$ of transverse elastic waves is given by

$$V_T = \sqrt{\frac{T}{d}} \quad \text{(m/s)} \tag{3.3}$$

$d$ is the "linear density" of the medium, that is, *mass per unit length* (in kg/m). Notice that the tenser a string is, the faster the transverse waves will travel. On the other hand, the denser it is, the slower the waves will propagate.

A physically equivalent relation exists for the propagation speed of longitudinal waves in a medium of density $\delta$ (in kg/m$^3$) and where the pressure is $p$ (in Newton/m$^2$):

$$V_L = \sqrt{\frac{p}{\delta}} \quad \text{(m/s)} \tag{3.4}$$

For an ideal gas, however, the ratio $p/\delta$ turns out to be proportional to the absolute temperature $t_A$, defined in terms of the Celsius or Fahrenheit temperatures $t_C$ and $t_F$ by the simple transformation

$$t_A = 273 + t_C = 273 + 5/9(t_F - 32) \quad \text{(degrees Kelvin).} \tag{3.5}$$

Notice that at the freezing point ($t_C = 0°$C, $t_F = 32°F$), the absolute temperature is $t_A = 273°$. Although ordinary air is not a 100% "ideal gas," it behaves approximately so, and the velocity of sound waves may be expressed as

$$V_L = 20.1\ \sqrt{t_A} \quad \text{(m/s).} \tag{3.6}$$

This turns out to be 331.5 m/s (= 1087 feet/s) at 0°C (32°F) and 344 m (= 1130 feet/s) at 21°C (70°F). The numerical factor in Eq. (3.6) is for air only. In general, its value depends on the *composition* of the medium through which the sound propagates. For pure hydrogen, for instance, it is equal to 74.0. Sound waves thus travel almost four times as fast in hydrogen as in air. This leads to funny acoustic effects if a person speaks or sings after having inhaled hydrogen (a less hazardous experiment is done with helium, for which the numerical constant in (3.6) is approximately 35).

Sound travels fast, but not infinitely fast. This, for instance, leads to small but noticeable *arrival time differences* between sound waves from different instruments in a large orchestra and may cause serious problems of rhythmic synchronization. A pianist who for the first time plays on a very large organ, in which the console is far away from most pipes, initially may become quite confused by the late arrival of the sound, out of synchronism with his fingers. *Reverberation* in a hall is based on superposition of delayed sound waves that have suffered multiple reflections on the walls (Sect. 4.7).

Let us now consider a very long string in which the initial point is set in vibration with simple harmonic motion and continues to vibrate indefinitely compelled

by some external force. After a while, all points of the string are found to vibrate with the same simple harmonic motion. If, at a given instant of time the initial point is, say, at its maximum displacement, its neighbors are not yet quite there, or just had been there, and so on. Figure 3.3 shows the transverse displacements of all points of the string at a *given time.* In other words, this curve is a "snapshot" of the shape of the string during the passage of a sinusoidal transverse wave. The graph in Fig. 3.3 should not be confused with the curve shown in Fig. 2.4, which represents *the time history of only one given point.* The latter indicates a vibration pattern in time, the former a wave pattern in space. The shortest distance between any two points of the string that are vibrating in a parallel way (vibrating in phase, i.e., having identical displacements $y$ at all times), is called the *wavelength.* It is usually designated with the Greek letter $\lambda$. Alternatively, the wavelength can be defined as the minimum space interval after which the spatial wave pattern repeats. Compare this with the definition of the period, which represents the minimum *time* interval after which the vibration pattern of *one given point* repeats (Fig. 2.3(b)).

As time goes on, the snapshot curve seems to move with the speed of the wave to the right (Fig. 3.4)—yet each point of the string only moves up-and-down (for instance consider point $x_1$ in Fig. 3.4). What moves to the right is the configuration, that is, the actual shape or pattern of the string, but not the string itself. In other words, what moves to the right is a quality, for instance the quality of "being at the maximum displacement" (e.g., points $P$, $Q$, $R$ in Fig. 3.4), or the quality of "just passing through equilibrium" (points $S$, $T$, $U$). And of course, what also moves to the right is *energy,* the potential and kinetic energy involved in the up-and-down oscillation of the points of the string.

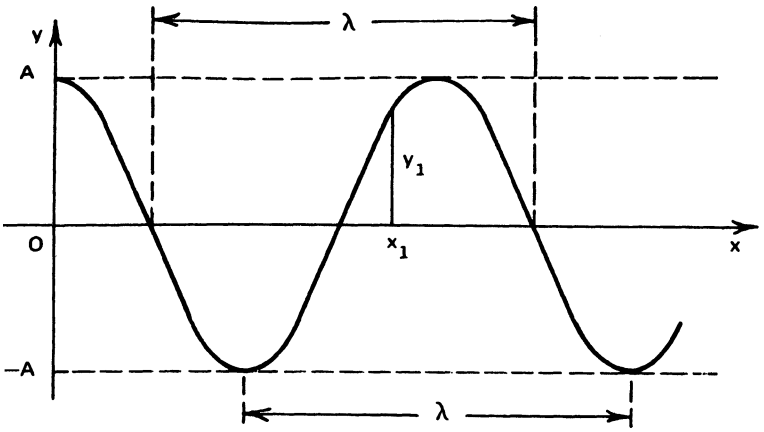



FIGURE 3.3 Snapshot of the displacements $y$ of a string when a single-frequency transverse wave propagates along it in direction $x$.

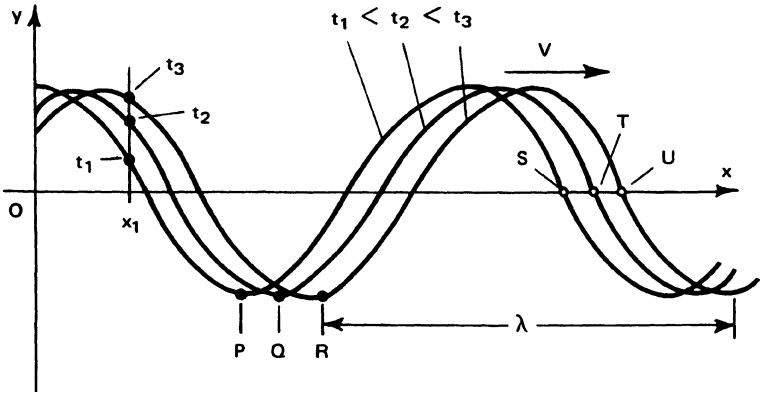



FIGURE 3.4 Three consecutive snapshots of a transverse wave taken at times $t_1$, $t_2$, and $t_3$. The points of the string only move up and down ($y$ direction); what moves to the right along $x$ is the wave profile or pattern (and wave energy).

There is an important relationship between the speed $V$ of a sinusoidal wave, its wavelength $\lambda$, and the frequency $f$ of oscillation of the individual points. Considering Fig. 3.3, we realize that the wave will have moved exactly one wavelength $\lambda$ during the time it takes the initial point (or any other) to make one complete oscillation, that is, during one period $\tau$. We therefore can write for the speed of the wave:

$$V = \frac{\text{distance traveled}}{\text{time employed}} = \frac{\lambda}{\tau}$$

Since the inverse of the period is equal to the frequency $f$ (relation (2.1)), we can also write

$$V = \lambda\, f \tag{3.7}$$

This relation provides the quantitative link between the "space representation" of Fig. 3.3 and the "time representation" of Fig. 2.4. Relation (3.7) enables us to express the wavelength of a transverse wave in a string in terms of the frequency of the oscillation of the individual points and the propagation velocity (3.3):

$$\lambda = \frac{1}{f}\sqrt{\frac{T}{d}} \tag{3.8}$$

It is interesting to note that relations (3.3) and (3.8) can also be applied, to a certain extent, to the basilar membrane, with the tension $T$ replaced by an appro-

priate stiffness parameter. Since the stiffness decreases from base to apex by a factor of about 10,000 (p. 31), according to Eqs. (3.3) and (3.8), the *local* velocity of propagation and wavelength of basilar membrane waves of given frequency will decrease by a factor 100 as they travel toward the apex. The resonance frequency of the basilar membrane, too, is proportional to the square root of the stiffness parameter. Energy considerations show that as a wave propagates, its amplitude will increase (the energy "piles up" because the wave slows down). When the wave reaches the resonance region, the amplitude reaches a maximum and energy dissipation reaches a peak, causing the wave to quickly die down beyond that point. Figure 3.5 shows schematically how the wave elicited by a single-frequency tone propagates along the basilar membrane. We can anticipate that when two or more pure tones are fed into the ear (as in real musical tones), separate "wave packets" as the one shown in the figure will arise defining different resonance regions, one for each frequency component (see Figs. 2.25(a) and (b)). This is, in highly oversimplified terms, how the hydromechanical frequency analysis mechanism works. Remember that Fig. 3.5 represents a transverse wave: Individual points vibrate up-and-down, but the wave pattern (and associated energy) propagates from left to right, with the amplitude of individual oscillations remaining within the wave "envelope." Notice the decrease in wavelength as the wave progresses in the direction toward maximum resonance. All points on the basilar membrane, including those well *outside* the peak resonance region, oscillate with the *same frequency* as the original pure tone. Finally, as we shift the frequency of

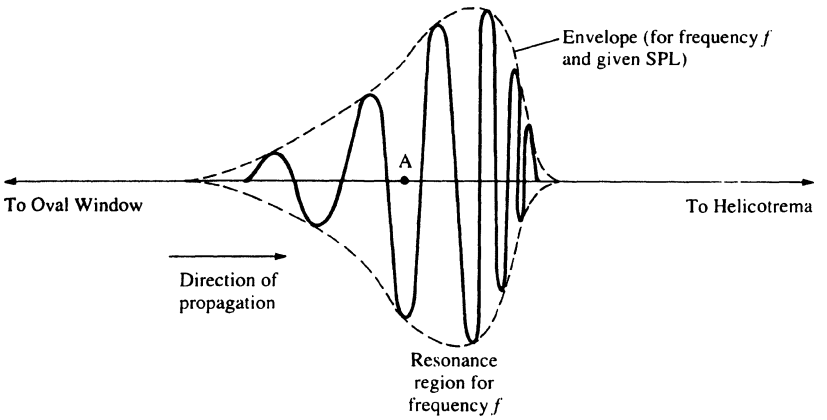



FIGURE 3.5 Sketch of a traveling wave along the basilar membrane, generated by a single-frequency tone. *Full curve*: Snapshot of transverse displacements of the membrane (not in scale!). Picture in your mind this curve traveling within the broken lines toward the right and slowing down as its amplitude dies down on the right. *Broken curve*: Amplitude envelope (which remains fixed unless the frequency and/or amplitude of the stimulus tone change).

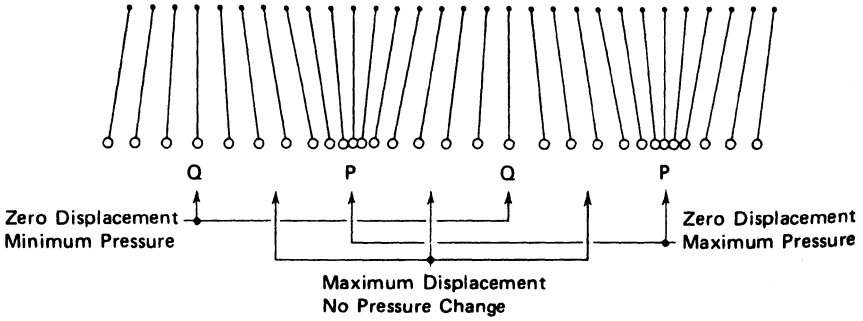



FIGURE 3.6 Longitudinal wave in a one-dimensional medium. To show the actual displacements, each point is depicted as the bob of a pendulum.

this tone up or down, the whole picture of Fig. 3.5 will be displaced toward the base or the apex, respectively.[3]

In the case of *longitudinal waves* such as a sound wave in air, the points vibrate in a direction *parallel* to the direction of propagation, and it is not so easy to picture their real position in visual form. For this reason, sound waves are more conveniently represented as pressure oscillations. Figure 3.6 shows the displacements of the points of a unidimensional model of the medium, when a longitudinal wave is passing through.

Notice that points show their maximum accumulation (i.e., maximum pressure) and maximum rarefaction (i.e., minimum pressure) at the places where their displacement is zero (points *P, Q,* respectively). On the other hand, at places where the displacements are maximum, the pressure variations are zero. This means that the pressure variations of a sound wave are 90° out of phase with the oscillation of the points: the maximum pressure variations (either increase or decrease) occur at places where the displacements of the points are zero; conversely, maximum displacements of the points occur at places where the pressure variations are zero.

A sinusoidal sound wave is one in which the pressure at each point oscillates harmonically about the normal (undisturbed) value (Fig. 3.7). At a point like *A*, all points of the medium have come closest to each other (maximum pressure increase, points *P* in Fig. 3.6); at a point like *B,* they have moved away from each other (maximum pressure decrease, points *Q* in Fig. 3.6). The *average pressure variation* $\Delta p$ is equal to the pressure variation amplitude divided by $\sqrt{2}$ (= 1.41).

[3]This is how the "characteristic frequency" of a neural fiber in the acoustic nerve arises (p. 54): consider a neuron wired to hair cells located at position A of Fig. 3.5. Its response will be related to the amplitude of the local oscillation of the basilar membrane. As the frequency of the test tone gradually rises from a very low value, the entire pattern shown in Fig. 3.5 will move from far right to far left: when the oscillation envelope passes over point A, that neuron's response capability will gradually increase to a maximum (at the characteristic frequency-see lower graph of Fig. 2.25(b)) and then suddenly drop as the pattern moves away to the left.

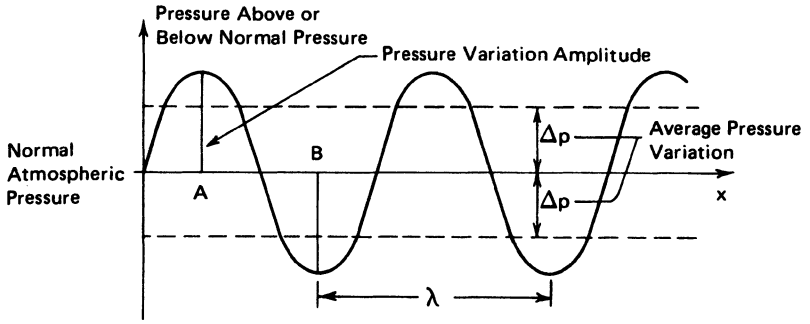



FIGURE 3.7 Pressure variations at a given time $t$ for a single-frequency sound wave propagating along $x$.

Taking into account relations (3.6) and (3.7), we obtain for the wavelength of sinusoidal sound waves in air:

$$\lambda = \frac{20.1}{f}\sqrt{t_A} \quad \text{(in meters)}. \tag{3.9}$$

$t_A$ is the absolute temperature given by Eq. (3.5). Typical values of wavelengths at normal temperature are shown in Fig. 3.8.

Elastic waves can be transmitted from one medium to another—for instance, from air into water, from air into a wall and then again into air, from a string to a wooden plate, and from there to the surrounding air. The nature of the wave may change in each transition (e.g., the transition from a transverse wave in the string and plate to a longitudinal sound wave in the air). However, in each transition, the *frequency remains invariant.* The wavelength, on the other hand, will change according to relation (3.7): $\lambda = V/f$. In this relation, $V$ changes from medium to medium, while $f$ is dictated exclusively by the initial vibration (source).

When an elastic wave hits the boundary between two media, part of it is *reflected* back to the original medium. Some boundaries are almost perfect reflectors (e.g., smooth cement walls for sound waves; the fixed end points of a tense string for transverse waves). This phenomenon is governed by the fact that on the reflecting boundary the points of the medium are compelled to remain at rest, therefore upsetting the balance of elastic forces that "command" the wave propagation. In a reflection, too, the frequency remains unchanged while the direction of propagation is reversed for a perpendicular incidence (or, in general, directed with a reflection angle equal to the incidence angle). Also, the amplitude would remain the same if there were no absorption.

We finally consider the *energy flow* associated with a sound wave. We define it as the amount of total mechanical energy (potential and kinetic, associated with the elastic oscillations of the points of the medium) that is transferred during each second through a surface of unit area (1 $m^2$) perpendicular to the direction of

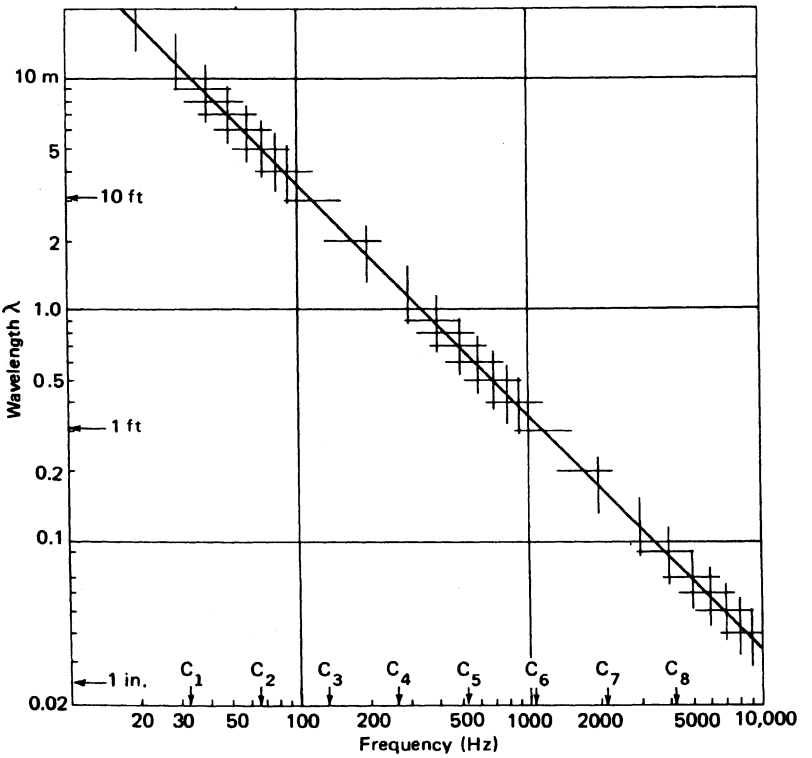



FIGURE 3.8 Wavelength of a sound wave in air at normal temperature, as a function of frequency (logarithmic scale).

propagation (Fig. 3.9). This energy flow is expressed in J per $m^2$ and s, or, taking into account the definition and the units of power (Eq. (3.2)), in $W/m^2$. It is more commonly called the *intensity of* the wave, and designated with the letter *I*. It can be show that there is a relation between the intensity *of* a sinusoidal sound wave and the value of the *average pressure oscillation* associated with the wave (see Fig. 3.7), which we denote with $\Delta p$ (equal to the pressure variation amplitude divided by $\sqrt{2}$):

$$I = \frac{(\Delta p)^2}{V\delta}$$

In this relation, $V$ is the speed of the sound wave (3.6) and 0 is the air density. For normal conditions of temperature and pressure, we have the following numerical relationship:

$$1 = 0.00234 \times (\Delta p)^2 \quad (W/m^2). \tag{3.10}$$

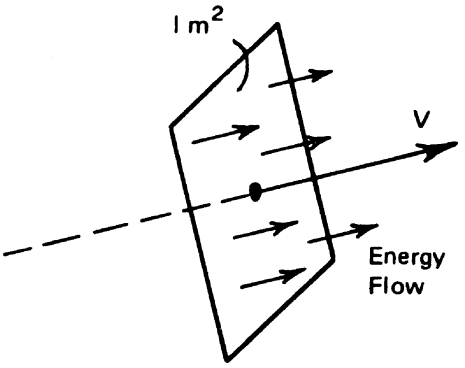



FIGURE 3.9 Energy flow through a unit area perpendicular to the flow direction.

$\Delta p$ must be expressed in Newton/m$^2$. As we shall see in Sect. 3.4, the faintest pure sound that can be heard at a frequency of 1000 Hz has an intensity of only $10^{-12}$ W/m$^2$. According to relation (3.10), this represents an average pressure variation of only $2.0 \times 10^{-10}$ Newton/m$^2$, that is, only $2.0 \times 10^{-10}$ that of the normal atmospheric pressure! This gives an idea of how sensitive the ear is.

A given sound source (a musical instrument or a loudspeaker) emits sound waves into all directions. In general, the amount of energy emitted per second depends on the particular direction considered. Let $I_1$ be the intensity of the wave at point $A_1$ propagating along the direction shown in Figure 3.10. This means that an amount of energy $I_1 a_1$ flows through the surface $a_1$ during each second. If we assume that no energy is lost on its way, this same amount of energy will flow each second through surface $a_2$ at point $A_2$. Therefore,

$$I_1 a_1 = I_2 a_2$$

Since the areas of the surfaces $a_1$ and $a_2$ are proportional to the squares of their respective distances $r_1$ and $r_2$ to the source, the intensity of a sound wave varies inversely proportional to the square of the distance to the source:

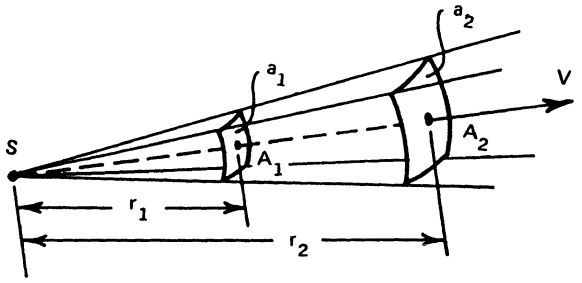



FIGURE 3.10 Radial sound energy flow.

$$\frac{I_1}{I_2} = \left(\frac{r_2}{r_1}\right)^2 \tag{3.11}$$

This law is no longer true if we take into account sound reflections and absorption.

If we imagine the whole sound source encircled by a spherical surface, the total amount of energy flowing each second through that surface is called the *acoustic power output* of the source. It represents the rate at which the source emits energy into *all* directions in form of sound waves. Its value is given in W. Typical instruments radiate between 0.01 W (clarinet) up to 6.4 W (trombone playing fortissimo).

## 3.3 Superposition of Waves; Standing Waves

In the absence of reflecting walls, sound waves travel in straight lines away from the source. As shown in the previous section, their intensity decreases rapidly, proportional to $1/r^2$, where $r$ is the distance to the source. If we have more than one source, the waves emitted by each one will propagate individually as if no other wave would exist, and the resultant effect at one given point of the medium (for instance in the auditory canal) will be a pressure oscillation that is simply given by the algebraic sum of the pressure oscillations of the individual waves.[4] In other words, *sound waves superpose linearly.*[5] This even happens to the traveling waves of the basilar membrane for low intensity sounds, giving rise to two resonance regions independent of each other (two envelopes of the type shown in Fig. 3.5).

Let us consider the superposition of two pure sound waves of frequencies $f_1$ and $f_2$ and, according to relation (3.7), of wavelengths $\lambda_1 = V/f_1$, $\lambda_2 = V/f_2$, traveling in the *same* direction. In order to obtain a snapshot of the resulting pressure variations, we just have to add the values of the individual pressure variations, as caused by each wave separately at each point $x$ along the direction of propagation. Since the velocity of sound waves does not depend on frequency (nor on the vibration pattern as a whole), all points of the medium will repeat exactly the same complex vibration pattern—only subject to a different timing. The energy flow—that is, the intensity of the superposition of two (or more) waves traveling in the same direction with random phases—is simply the sum of the energy flow contributions from the individual components:

$$I = I_1 + I_2 + I_3 + \cdots \tag{3.12}$$

[4]Note carefully that what is added here are pressure variations, and not absolute values of the pressure!

[5]Not true for extremely loud (powerful) sound waves like those from an explosion.

A particularly important case is that of two sinusoidal waves of the same frequency and the same amplitude *traveling in opposite directions.* This, for instance, happens when a sinusoidal wave is reflected at a given point (without absorption) and then travels back, superposing itself with the incoming wave. Let us first consider transverse waves in a string (Fig 3.11). Adding the contributions from each component, we obtain another sinusoidal wave of the same frequency but of different amplitude. The striking fact, however, is that this resultant wave *does not propagate at all!* It remains anchored at certain points $N_1, N_2, N_3, \ldots$, called *nodes,* that do not vibrate. Points between nodes vibrate with different amplitudes, depending on their position. In particular, points $A_1, A_2, A_3, \ldots$ (midway between nodes), called *antinodes*, vibrate with maximum amplitude (twice that of each component wave). Figure 3.12 shows the successive shapes of a string when two sinusoidal waves of the same amplitude travel in opposite directions. This is called a *standing wave*. The wave profile changes in amplitude periodically, but does not move, neither to the right nor to the left. At one time ($t_1$), the string shows a pattern of maximum deformation; at another ($t_5$), it has no deformation at all. As we shall see in the next chapter, standing waves play a key role in music, especially in the sound generation mechanisms of musical instruments.

In a standing wave, *there is no net propagation of energy* either. The whole string almost acts as one elastic vibrating spring: at one given time (e.g., $t_5$ in Fig. 3.12), all points are passing through their equilibrium position, and the energy of the whole string is in kinetic form (energy of motion). At another instant (e.g., $t_1$ in Fig. 3.12), all points are at their maximum displacement and the energy is all potential. In other words, in a standing wave all points oscillate in phase. Note carefully that this does not happen with a propagating wave. In Fig. 3.3, for instance, at one given instant of time, there are points which have maximum displacement (only potential energy) as well as points with zero displacement (only kinetic energy), or points in any intermediate situation (both forms energy). Furthermore, in a propagating wave, all points have the same amplitude; what varies are the times at which a maximum displacement is attained (the points are out of phase).

A careful inspection of Fig. 3.12 reveals that the distance $l_N$ between two neighboring nodes, $N_1$, $N_2$, or the distance $l_A$ between two antinodes $A_1$, $A_2$ is exactly one-half of a wavelength $\lambda$:

$$l_N = l_A = \lambda/2 \tag{3.13}$$

On the other hand, the distance $l_{NA}$ between a node $N_1$ and an antinode $A_1$ is a quarter wavelength:

$$I_{NA} = \lambda/4 \tag{3.14}$$

Standing waves can also be longitudinal. They arise when two sound waves of the same frequency and pressure variation amplitude travel in opposite directions. This happens, for instance, when a sound wave travels along a pipe and is reflected

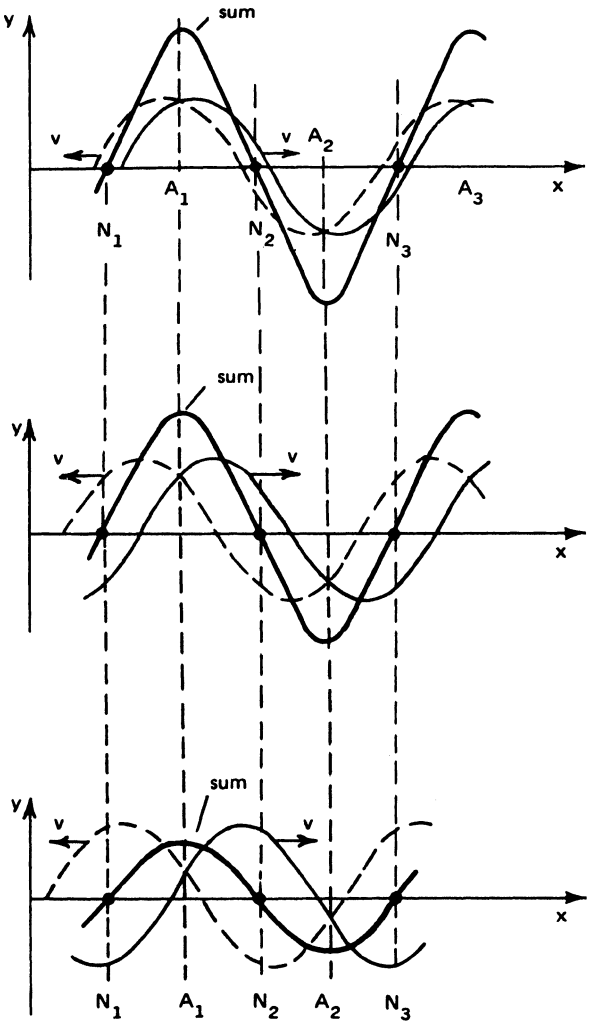



FIGURE 3.11 Superposition of two transverse waves of the same amplitude and frequency, moving in mutually *opposite* directions $+V$ and $-V$. The resulting wave pattern does not propagate: it remains "anchored" at the nodes $N$, changing periodically only its amplitude.

at the other end; standing waves also arise from reflections on the walls in rooms and halls. They have the same properties as transverse standing waves, and the discussion given above applies here too. There is, however, an important additional remark to be made. As pointed out in the previous section, sound waves are most conveniently described by pressure oscillations. We showed there that points with maximum pressure variation have zero longitudinal displacement (Fig. 3.6), whereas places with zero pressure variations correspond to points with maximum

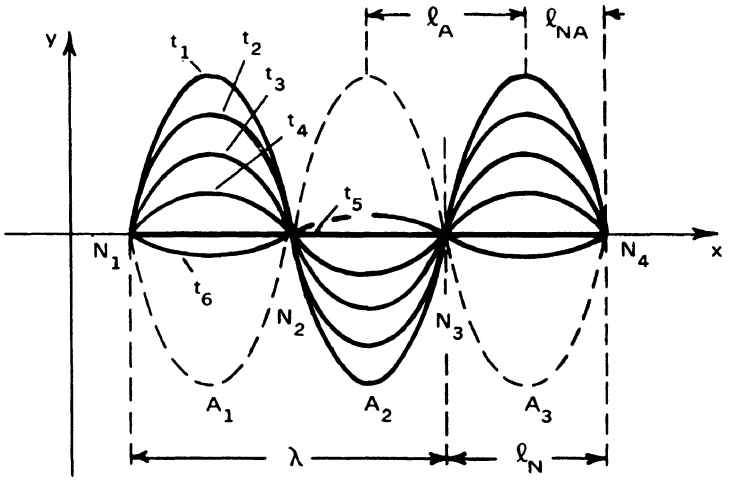



FIGURE 3.12 Successive shapes of a string in standing wave oscillation.

displacement. We can translate this to the case of a standing sound wave: *Pressure nodes* (i.e., points whose pressure variations are permanently zero) are *vibration antinodes* (points which oscillate with maximum amplitude), whereas *pressure antinodes* (points at which the pressure oscillates with maximum amplitude) are *vibration nodes* (points which remain permanently at rest).

## 3.4 Intensity, Sound Intensity Level, and Loudness

In Sect. 2.3, we stated that, for a pure sound, the amplitude of the eardrum oscillations leads to the sensation of loudness. This amplitude is directly related to the average pressure variation $\Delta p$ of the incoming sound wave and, hence, to the acoustic energy flow or intensity $I$ reaching the ear (relation (3.10)). We start here by investigating the range of intensities $I$ of pure sound waves to which the ear is sensitive. There are two limits of sensitivity to a tone of given frequency: (1) A lower limit or *threshold of hearing* representing the minimum just audible intensity; (2) An *upper limit of hearing* beyond which physiological pain is evoked, eventually leading to physical damage of the hearing mechanism. One finds that these two limits vary from individual to individual and depend on the particular frequency under consideration. In general, for a tone of about 1000 Hz (a pitch between the notes $B_5$ and $C_6$), the interval between limits is largest. The enormous range of intensities encompassed between the two limits of hearing is startling. Indeed, for a tone of 1000 Hz, it is found that the average threshold intensity lies near $10^{-12}$ W/m$^2$, whereas the limit of pain lies at about 1 W/m$^2$. This represents a ratio of intensities of one trillion to one, to which the ear is sensitive! Table 3.1 shows the relations between sound intensity and musical loudness sensation, for

a 1000 Hz tone.[6] At 1000 Hz, the intensity range of *musical* interest extends from about $10^{-9}$ to $10^{-2}$ W/m$^2$. This still represents a variation of a factor of 10 million!

Because of this tremendous range, the unit of W/m$^2$ is impractical There is another reason why it is impractical. The difference limen (DL) or just noticeable difference of a given stimulus (Sect. 1.4) is usually a good physical "gauge" to take into account when it comes to choosing an appropriate unit for the corresponding physical magnitude. Experiments show that the DL *in tone intensity* is roughly proportional to the intensity of the tone. This proportionality thus suggests that the appropriate unit should gradually increase, as the intensity of the tone we want to describe increases. This, of course, would lead to an awful complication unless we introduce different magnitude which is an appropriate function of the intensity. This new magnitude should accomplish three simultaneous objectives: (1) A "compression" of the whole audible intensity scale into a much smaller range of values. (2) The use of relative values (for instance, relative to the threshold of hearing) rather than absolute ones. (3) The introduction of a more convenient unit whose value closely represents the minimum perceptible change of sound intensity.

The introduction of the new quantity is done in the following way: Note in Table 3.1 that what seems more characteristically related to the loudness effect is the *exponent* to which the number 10 is raised when we quote the value of the sound intensity (left column); –12 for the threshold of hearing; –9 for a *ppp* sound; –7 for piano; –5 for forte;, –3 for forte-fortissimo, and 0 for limit of pain

TABLE 3.1 Comparison of sound intensity with musical loudness sensation.

| Intensity (Watt/m$^2$) | Loudness |
|---|---|
| 1 | Limit of pain |
| $10^{-3}$ | *fff* |
| $10^{-4}$ | *ff* |
| $10^{-5}$ | *f* |
| $10^{-6}$ | *mf* |
| $10^{-7}$ | *p* |
| $10^{-8}$ | *pp* |
| $10^{-9}$ | *ppp* |
| $10^{-12}$ | Threshold of hearing |

[6]The music notation really does not represent an absolute measure of loudness. Musicians for instance will argue that we are perfectly able to perceive fortissimos and pianissimos in music played on a radio with its volume control turned down to a whisper. What happens in such a case is that we use cues other than intensity-particularly if we know the piece-to make subjective judgments of "relative" loudness. On other hand, systematic experiments (Patterson, 1974) have revealed that the interpretation of the musical loudness notation in an actual dynamic context is highly dependent on the instrument and on the pitch range covered.

($10^0 = 1$). This strongly suggests that we should use what in mathematics is called a *logarithmic function* to represent intensity.

The *decimal logarithm* of a given number is the exponent to which 10 has to be raised in order to yield that number. For instance, the logarithm of 100 is 2 because $10^2 = 100$; the logarithm of 10,000 is 4 because $10^4 = 10{,}000$; the logarithm of 1 is zero, because $10^0 = 1$; and the logarithm of 0.000001 is $-6$ because $10^{-6} = 0.000001$. These relations are written symbolically: log 100 = 2; log 10,000 = 4; log 1 = 0; log 0.000001 = $-6$. For any number intermediate between integer powers of ten, the logarithm can be found with a calculator.

An important property is that the logarithm of the *product* of two numbers is the *sum* of the logarithms of the individual numbers. For instance, the logarithm of the number $10^4$ times $10^3$ is 4 *plus* 3 (i.e., 7), because $10^4 \times 10^3 = 10^{4+3} = 10^7$. In general, for any two numbers $a$ and $b$, we have the relation $\log(a \times b) = \log a + \log b$. For the logarithm of a division $a/b$, we have, instead, $\log (a/b) = \log a - \log b$.

Decimal logarithms indeed can be used to define a more appropriate magnitude to describe sound intensity. First of all, we adopt the hearing threshold (at 1000 Hz) of $10^{-2}$ W/m$^2$ as our reference intensity $I_0$. Then we introduce the quantity

$$IL = 10 \times \log \frac{I}{I_0} \tag{3.15}$$

This is called the *sound intensity level.* The unit of $IL$ is called the *decibel,* denoted "db." For the hearing threshold, $I/I_0 = 1$ and $IL = 0$ db. For the upper limit of hearing, $I/I_0 = 10^{12}$ and $IL = 10 \times \log 10^{12} = 120$ db. A typical "forte" sound (Table 3.1) has a sound intensity level of 70 db; *ppp* corresponds to 30 db.

It is important to note that when a quantity is expressed in decibels, a *relative* measure is given, with respect to some reference value (e.g., the hearing threshold in the definition of $IL$). Whenever the intensity 1 is multiplied by a factor of 10, one just *adds* 10 db to the value of $IL$; when the intensity is multiplied by 100, one must add 20 db, etc. Likewise, when the intensity is divided by a factor of 100, one must *subtract* 20 db from $IL$.Table 3.2 gives some useful relationships.

TABLE 3.2 Comparison of changes in sound intensity level $IL$ (in db) with changes in intensity.

| Change in $IL$ | What happens to the Intensity |
|---|---|
| add (subtract) 1 db | Multiply (divide) by 1.26 |
| +(−) 3 db | ×(÷) 2 |
| +(−) 10 db | ×(÷) 10 |
| +(−) 20 db | ×(÷) 100 |
| +(−) 60 db | ×(÷) 1,000,000 |

We may use relation (3.10) to express the intensity in terms of the much more easily measurable average pressure variation $\Delta p$. We find that the minimum threshold $I_0$ at 1000 Hz roughly corresponds to an average pressure variation $\Delta p_0 = 2 \times 10^{-5}$ Newton/m$^2$ (20 Micropascal). Since according to relation (3.10) $I$ is proportional to the *square* of $\Delta p$, we have

$$\log \frac{I}{I_0} = \log \left( \frac{\Delta p}{\Delta p_0} \right)^2 = 2 \log \frac{\Delta p}{\Delta p_0}$$

Hence, one may introduce the quantity

$$SPL = 20 \log \frac{\Delta p}{\Delta p_0} \tag{3.16}$$

called the *sound pressure level* (*SPL*). For a traveling wave, the numerical values of Eqs. (3.15) and (3.16) are identical, and *IL* and *SPL* represent one and the same thing. For *standing* waves, however, there is no energy flow at all (Sect. 3.3) and the intensity $I$ as used in Eq. (3.15) cannot be defined; hence *IL* loses its meaning. Yet, the concept of average pressure variation $\Delta p$ at a given point in space (e.g., inside an organ pipe) still remains meaningful and so does the sound pressure level. This is why relation (3.16) is used more frequently than Eq. (3.15). Note carefully that the definitions of *IL* and *SPL* do not involve at all the frequency of the sound wave. Although we did make reference to a 1000 Hz tone, nothing prevents us from defining *IL* and *SPL* through relations (3.15) and (3.16), respectively, for any arbitrary frequency. What does depend on frequency, and very strongly, are the subjective limits of hearing (e.g., 10 and $\Delta p_0$) and, in general, the *subjective* sensation of loudness, as we shall see further below.

Funny things seem to happen with the sound intensity level, or sound pressure level, when we superpose two sounds of the same frequency (and phase). Just consider Table 3.2. Adding two tones of the same intensity, which according to Eq. (3.12) means doubling the intensity, adds a mere 3 db to the sound level of the original sound, whatever the actual value of the *IL* might have been. Superposing *ten* equal tones (in phase) only increases by 10 db the resulting *IL*. To raise the *IL* of a given tone by 1 db, we must multiply its intensity by 1.26, meaning that we must add a tone whose intensity is 0.26 (about 1/4) that of the original tone.

The minimum change in *SPL* required to give rise to a detectable change in the loudness sensation (DL in sound level) is roughly constant and of the order of 0.2–0.4 db in the musically relevant range of pitch and loudness. The unit of *IL* or *SPL*, the decibel, is thus indeed of "reasonable size"—close to the DL.

There is an alternative way of looking at the DL of intensity or sound level. Instead of asking how much the intensity of *one* given tone must be changed to give a just noticeable effect, we may pose the totally equivalent question: What is the minimum intensity $I_2$ that a *second* tone of the same frequency and phase must have, to be noticed in presence of the first one (whose intensity $I_1$ is kept

constant)? That minimum intensity $I_2$ is called the *threshold of masking.* The original tone of constant intensity $I_1$ is called the "masking tone," the additional tone is the "masked tone." Masking plays a key role in music. In this paragraph, we only mention the masking of tones of frequency (and phase) identical to that of the masking tone; further below, we shall discuss masking at different frequencies. The relation between the masking level *ML* (*IL* of the masked tone at threshold), the DL of sound level, and the *IL* of the masking tone can be found from their defining expressions (based on Eq. (3.15)):

$$ML = 10\log\left(\frac{I_2}{I_0}\right); \quad DL = 10\log\left(\frac{I_1 + I_2}{I_1}\right); \quad IL = 10\log\left(\frac{I_1}{I_0}\right)$$

So far, we have been dealing with the physical quantities *IL* and *SPL.* Now we must examine the psychological magnitude *loudness,* associated with a given *SPL.* In Sects. 1.4 and 2.3, we mentioned the ability of individuals to establish an order for the "strength" of two sensations of the same kind, pointing out that complications arise when absolute quantitative comparisons are to be made. In the case of loudness, judgments of whether two pure tones sound equally loud show fairly low dispersion among different individuals. But judgments as to "how much" louder one tone is than another require previous conditioning or training and yield results that fluctuate greatly from individual to individual.

Tones of the same *SPL* but with different frequency, in general, are judged as having different loudness. The *SPL* is thus not a good measure of loudness, if we intercompare tones of different frequency. Experiments have been performed to establish *curves of equal loudness*,[7] taking the *SPL* at 1000 Hz as a reference quantity. These are shown in Fig. 3.13 (Fletcher and Munson, 1933). Starting from the vertical axis centered at 1000 Hz toward both sides of lower and higher frequencies, respectively, are drawn curves that correspond to *SPL*'s of tones that are judged "equally loud" as the reference tone of 1000 Hz. Note, for instance, that while a *SPL* of 50 db (intensity of $10^{-7}$ W/m$^2$) at 1000 Hz is considered "piano," the same *SPL* is barely audible at 60 Hz. In other words, to produce a given loudness sensation at low frequencies, say a "forte" sound, a much higher intensity (energy flow) is needed than at 1000 Hz. This is why bass tones seem to "fade away" much before the trebles, when we gradually move away from a fixed sound source. Or why we have to pay so much more for a hi-fi set, particularly the speakers, if we want well-balanced basses.

The lowest curve in Fig. 3.13 represents the threshold of hearing for different frequencies. Again, it shows how the sensitivity of the ear decreases considerably toward low (and also toward very high) frequencies. Maximum sensitivity is achieved at about 3000 Hz. The shape of this threshold curve is influenced by the acoustic properties of the auditory canal (meatus) and by the mechanical

[7] Obtained through "loudness matching" experiments conducted in a way quite similar to pitch matching experiments.

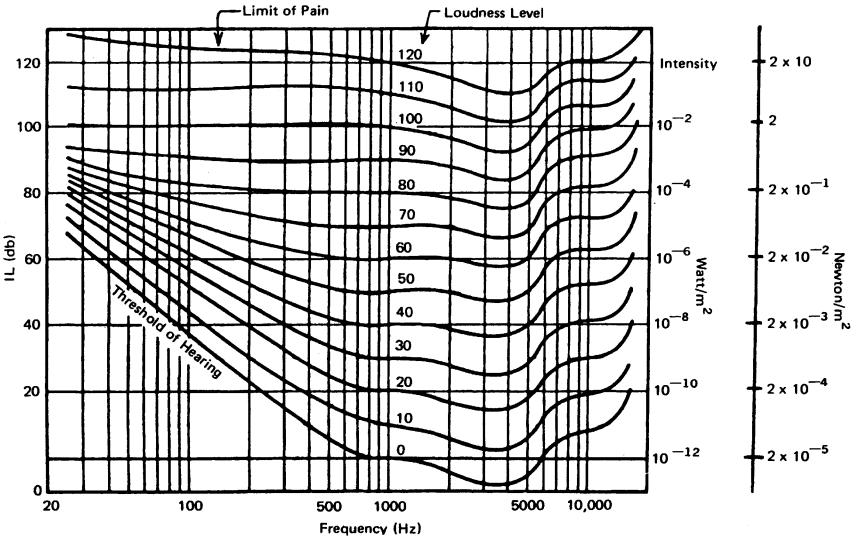



FIGURE 3.13 Curves of equal loudness (Fletcher and Munson, 1933) in a sound intensity level (*IL*) and frequency diagram. The corresponding scales of sound intensity (Watt/m$^2$) and average pressure variation (Newton/m$^2$) are also shown. Reprinted by permission from the *Journal of the Acoustical Society of America.*

properties of the bone chain in the middle ear; note carefully that all curves shown in the figure represent averages from a large population of experimental subjects. Finally, it should be emphasized that the curves in Fig. 3.13 are valid only for *single, continuously sounding pure tones.* We shall discuss further below what happens to the loudness sensation if a tone is of short duration (less than a second). Later studies (Molino, 1973) show that equal loudness contours appear to depend on the frequency of the reference tone (which was 1000 Hz in Fig. 3.13).

Now comes a sometimes confusing issue. A new quantity is introduced, called the *loudness level,* which we designate *LL.* It is defined in the following way: The *LL* of a tone of frequency *f* is given by the *SPL* of a tone of 1000 Hz that is judged to be equally loud. This means that the curves of Fig. 3.13 are curves of constant loudness level. The unit of *LL* is called the *phon.* Fig. 3.12 can be used to find the *LL* of a tone of given *SPL*, at any frequency *f.* For instance, consider a tone of 70 db *SPL* ($I = 10^{-5}$ W/m$^2$) at 80 Hz. We see that the curve that passes through that point intersects the 1000 Hz line at 50 db. The *LL* of that tone is thus equal to 50 phons. In general, the numbers shown along the 1000 Hz line represent the *LL* in phon of the corresponding constant loudness curves.

Note very carefully that *LL* still is a *physical* magnitude, rather than a psychophysical one (in spite of the name). It represents those intensities or *SPL*'s that sound equally loud, but it does not pretend to represent the percept loudness in an absolute manner: a tone whose *LL* is twice as large simply does not sound twice as loud! Many studies have been made to determine a subjective scale of loudness.

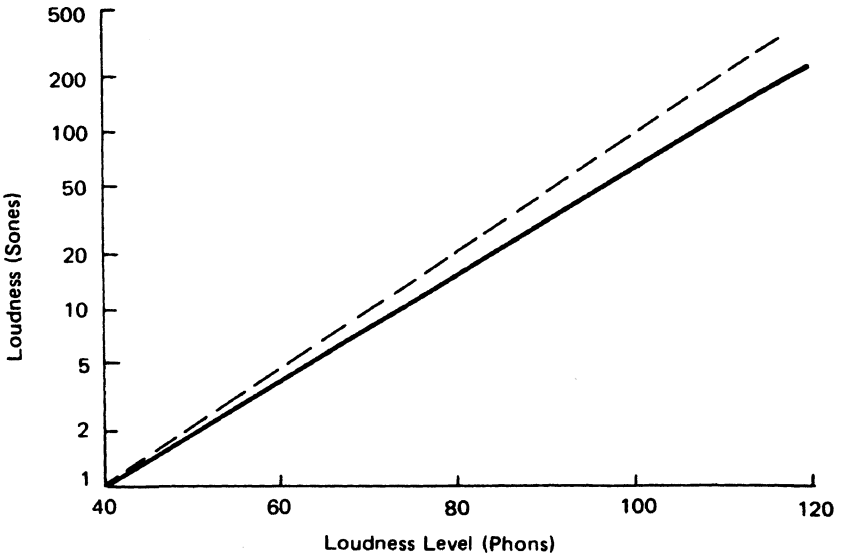



FIGURE 3.14 *Solid line*: Experimental relation between the psychological magnitude loudness and the physical magnitude loudness level (after Stevens, 1955). *Broken line*: Power law relationship (3.17) (Stevens, 1970).

Figure 3.14 (solid line) is the result (Stevens, 1955), relating the "subjective loudness" $L$ with the *loudness level LL,* in the range of musical interest. The quantity $L$ describing subjective loudness is expressed in units called *sones.* Notice that the relationship is not linear (the loudness scale in Fig. 3.14 is what one calls a logarithmic scale). It is such that increasing $LL$ by 10 phons, the loudness $L$ is merely *doubled.* This, for instance, means that *ten* instruments playing a given note at the same $LL$ are judged to sound only *twice* as loud as one of the instruments playing alone!

It has been shown that the relation between $L$ and the wave intensity $I$ or the average pressure variation $\Delta p$, can be described approximately by the simple function (Stevens, 1970):

$$L = C_1\sqrt[3]{I} = C_2\sqrt[3]{(\Delta p)^2} \tag{3.17}$$

where $C_1$ and $C_2$ are parameters that depend on frequency. This yields the broken line shown in Fig. 3.14, which lies well within the statistical fluctuation of the actual measurements (not shown). Notice that the logarithmic relationship (as in Eqs. (3.15) and (3.16)) is all but gone. Yet, an appreciable "compression" of the subjective loudness scale still remains: in order to vary $L$ between 1 and 200, the intensity $I$ must change by a factor of eight million.

When we superpose two or more tones of the same frequency (and randomly mixed phases), the resulting tone has an intensity (energy flow) which is the sum

of the intensities of the component tones: $I = I_1 + I_2 + I_3 + \cdots$ (3.12). Since in this case, the individual tones cannot be discriminated from each other, this total intensity determines the resulting loudness through relation (3.17). $L$ obviously will not be equal to the sum of the loudnesses of the individual tones. Different situations arise when the component waves have *different frequencies.* We may distinguish three regimes:

(1) If the frequencies of the component tones fall all *within the critical band* of the center frequency (Sect. 2.4), the resulting loudness is still directly related to the total intensity (energy flow), sum of the individual intensities:

$$L = C_1 \sqrt[3]{I_1 + I_2 + I_3 + \cdots} \tag{3.18}$$

This property actually leads to a more precise determination (Zwicker, Flottorp, and Stevens, 1957) of the critical band than that given in Sect. 2.4.

(2) When the frequency spread of the multitone stimulus *exceeds the critical band,* the resulting subjective loudness is *greater* than that obtained by simple intensity summation (3.18), increasing with increasing frequency difference and tending toward a value that is given by the sum of individual loudness contributions from adjacent critical bands (Zwicker and Scharf, 1965):

$$L = L_1 + L_2 + L_3 + \cdots \tag{3.19}$$

Masking effects must be taken into account if the individual loudnesses $L_1$, $L_2$, etc., differ considerably among each other. The limit of loudness integration (3.19) is never achieved in practice.

(3) When the frequency difference between individual tones is large, the situation becomes complicated. First of all, difficulties arise with the concept of total loudness. People tend to focus on only *one* of the component tones (e.g., the loudest, or that of highest pitch) and assign the sensation of total loudness to just that single component:[8]

$$L = \text{maximum of } (L_1 + L_2 + L_3 + \cdots) \tag{3.20}$$

[8] A well-known situation, similar to this case, arises with the sensation of pain. If you are pinched in two places very near to each other, the pain may be "twice" that of a single pinch (equivalent to case (1) above). But when the places are far apart, you have difficulty in sensing out what one may call "total pain" (case (3)). Actually, you will tend to focus on the one giving the greater pain sensation.

All this has very important consequences for music. For instance, two organ pipes of the same kind and the same pitch sound only 1.3 times louder than one pipe alone (Table 3.2). When their pitch is a semitone or a whole tone apart, their loudness still will be roughly 1.3 times that of one single pipe (a semitone or whole tone falls within the critical band, Fig. 2.13). But two tones that are a major third apart will sound louder than the previous combination. These facts have been well known for centuries to organ builders and composers as well. Since there is no possibility for a manual loudness control of individual organ tones as in string instruments or woodwinds, loudness on the organ can be altered only by changing the number of simultaneously sounding pipes. But since, according to the previous paragraph, loudness summation is more effective when the component tones differ in frequency appreciably (relation (3.19)), stops sounding one (4′ stop), two (2′), or more octaves above the written note (and also below (16′)) are mainly used for that purpose.[9] On the other hand, without adding or subtracting stops, loudness may also be controlled through the number of notes simultaneously played. Each new voice entering in a fugue increases the subjective loudness of the piece, as does each additional note in a chord. Some organists play the final chords of a Bach fugue pulling additional stops. This is absurd—Bach himself programmed the desired loudness increase by simply writing in more notes than the number of voices used throughout the fugue![10]

We have examined the summation of loudness of two or more superposed tones, but we have not yet discussed what happens to the threshold of hearing of a tone when it is sounded in the presence of another. If the frequencies of both tones coincide, this threshold is given by the masking level discussed earlier (p. 97). If their frequencies differ, we still can determine a masking level, defined as the minimum intensity level which the masked tone must exceed in order to be "singled out" and heard individually in presence of the masking tone. The intensity threshold of isolated pure tones (bottom curve in Fig. 3.13) changes appreciably, that is, increases, if other tones are simultaneously present. The most familiar experience of masking is that of not being able to follow a conversation in presence of a lot of background noise. The *masking level ML* of a pure tone of frequency *f* in presence of another pure tone of fixed characteristics (frequency 415 Hz and intensity level *IL)* is shown in Fig. 3.15 (Egan and Hake, 1950). The level *ML* to which the masked tone has to be raised above the normal threshold of hearing (as given by the bottom curve of Fig. 3.13) is represented for different *IL* values of the masking tone. The regions close to $f_0$ (dotted portions) must be

[9]Their addition, of course, also will contribute to a change in timbre (Chapter 4).

[10]Many baroque organs had a stop called "Zimbelstern," sounding very high-pitched miniature cymbals or bells mounted on a rotating star at the top of the organ case; this stop was pulled to reinforce the loudness of a final chord (without in any way interfering with its harmony: frequencies above about 5000 Hz do not contribute to, or interfere with, the periodicity pitch sensation in the normal musical range).

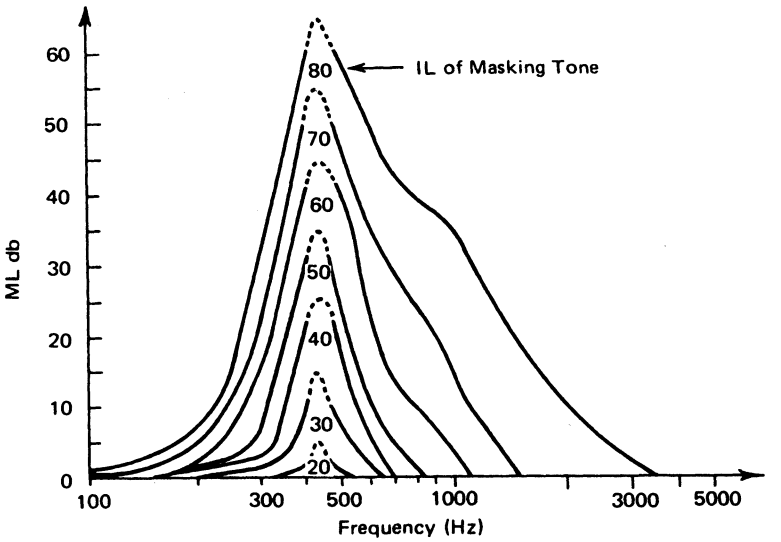



FIGURE 3.15 Masking level corresponding to a pure tone of 415 Hz, for various sound level values (Egan and Hake, 1950) of the masking tone. Reprinted by permission from the *Journal of the Acoustical Society of America*.

extrapolated to the value deduced from the DL of loudness (p. 96); beat phenomena play a role there, which has nothing to do with masking per se. At higher *IL,* additional complications arise due to the appearance of aural harmonics at frequencies $2f_0$, $3f_0$, etc. (Sect. 2.5). Notice the asymmetry of the curves at higher *IL* (caused by these aural harmonics): a tone of given frequency $f_0$ masks more efficiently the higher frequencies than the lower ones. Masking, on the other hand, causes a small change in pitch of a single-frequency masked tone due to the asymmetric distribution of basilar membrane excitation by the masking tone (e.g., Fig. 3.5).

Masking plays an important role in polyphonic music, particularly in orchestration. On many occasions in musical scores, the participation of a given instrument such as an oboe or a bassoon may be totally irrelevant if it plays at the same time when the brasses are blasting a fortissimo. Likewise, the addition of flute stops or other soft string-like stops to a *tutti* of diapasons, mixtures, and reeds in the organ is completely irrelevant from the point of view of loudness.

Finally, we must mention the effect of the *duration* of a tone on the loudness sensation (called *temporal integration).* First of all, there is a *time threshold—a* minimum duration that a given pure sound must have in order to yield a tone sensation at all. This minimum duration is about 10–15 ms, or, at least, two to three oscillation periods if the frequency is below 50 Hz; tones lasting less than

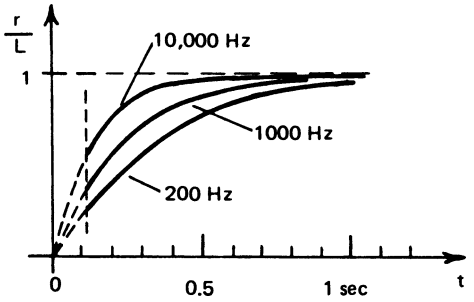



FIGURE 3.16 Relative loudness of pure tones of short duration *t*. *r*/*L*: ratio of actually perceived loudness *r* to the loudness *L* of a steady tone of the same frequency and amplitude.

this are perceived as "clicks," not as "tones."[11] Sounds lasting longer than 15 ms (or two to three periods, whichever is longer) can be individualized as tones of given pitch and loudness, but the subjective loudness does depend on tone duration (Plomp and Bouman, 1959). The shorter the tone pulse, the lower its loudness, if the physical tone intensity (energy flow) is kept constant (e.g., Richards, 1977). Figure 3.16 shows a sketch of the relative loudness decrease, or loudness attenuation, as a function of tone duration, for different frequencies. Notice that the final response value is reached sooner for higher frequencies; after about half a second, the loudness reaches a constant value that depends only on intensity (relationship shown in Fig. 3.14). Masking (Fig. 3.15), too, has significant and rather complicated time-dependent characteristics for short-duration masking tones; the reader is referred to Zwicker and Fastl (1999) for details. In general, for short tones it is found that, in first approximation, the loudness sensation is not related to the instantaneous power flow (intensity of the sound wave), but to the total acoustic *energy* delivered by the tone to the ear (intensity times duration). Actually, there are indications that in this case the loudness sensation is related to the total number of *neural impulses* that have been transmitted in association with the short tone (Zwislocki, 1969) (see the next section).

For long exposure times, an effect called *adaptation* sets in, consisting of a *decrease* in subjective loudness when a tone of constant intensity is being heard during several minutes.[12] Although there is great disparity among subjects, loudness matching experiments reveal some common features (Scharf, 1983). Adaptation increases with the frequency of pure tones of the same *SPL* (higher frequency tones are "turned down" faster); for tones of the same frequency, it decreases when

[11] Quite generally, there is a physical principle which states that the frequency of a vibration (hence also the associated pitch) cannot be defined more accurately than the inverse of the total duration of the vibration.

[12] This should not be confused with acoustic fatigue, a psychological process through which our brain is able to ignore a continuous but otherwise irrelevant sound.

the *SPL* increases (there is no adaptation effect for *SPL*s higher than about 40 db). In general, for musically relevant frequencies and *SPL*s, the subjective loudness of a tone levels off after about 100 s, and then remains constant.

Loudness attenuation for short tones has very important implications for music performance. If we want to play a "staccato" passage on the piano at a given loudness, we must hit the keys much harder than if we were to play the same notes "legato" at the same loudness.[13] This effect is much more pronounced with non-decaying tones, such as organ sounds. Indeed, the organist has a considerable control of the subjective loudness of one given note by giving it the right duration; phrasing in organ playing is the art of dynamic control by giving the appropriate duration to a note (it obviously can work only for short tones—see Fig. 3.16). On the other hand, loudness adaptation also plays a role in music: to combat it, trills have been invented! In contrast, the use of the "pedal point" (constant-sounding bass note) is a proof of the above-stated fact that adaptation is unimportant at low frequencies. Given the importance of these time effects for music, much more research on attenuation and adaptation is warranted.

## 3.5 The Loudness Perception Mechanism and Related Processes

What physical, physiological, or neural processes are responsible for the difference between the limited subjective scale of loudness and the huge range of detectable intensities of the original sound (Table 3.1 and relation (3.17))? In the case of primary pitch perception (Sect. 2.3), we already had encountered such a "logarithmic compression": whereas the audible frequency scale ranges from about 20 Hz to about 16,000 Hz, this only comprises nine octaves of pitch. In that case, the compression is mainly caused by the mechanical resonance properties of the cochlear partition: the curve in Fig. 2.8 indeed represents a roughly logarithmic relationship between the position $x$ of the resonance region on the basilar membrane, and frequency $f$ of a pure tone.

In the case of the loudness detection mechanism, the compression is partly neural, partly mechanical. When a pure sound is present, the primary neurons connected to sensor cells located in the center of the region of maximum resonance amplitude (i.e., fibers with the same characteristic frequency, p. 62) increase their firing rate above the spontaneous level. This increase is a monotonic function of the stimulus amplitude, but not a linear one (e.g., Sachs and Abbas, 1974). Indeed, when the latter increases by a factor of, say, 100 (a 40 db increase in *SPL*, relation (3.16)), the firing rate is found to increase only by a factor between 3 and 4.

[13] For tones of musical instruments complications arise: during tone buildup which may last several tenths of a second, a natural change in intensity and spectrum occurs at the source. Also, reverberation effects are important (Sect. 4.7).

Another element contributing to loudness compression is related to the following: At higher *SPL*'s a primary neuron's firing rate *saturates* at a level that is only a few times that of spontaneous firing. Any further increase in intensity would not greatly alter this firing rate; neurons simply cannot transmit pulses at a rate faster than the saturation value (determined by the refractory time after each pulse, Sect. 2.8). Individual auditory nerve fibers with similar characteristic frequencies have widely differing firing thresholds (the *SPL* needed to increase the firing rate above spontaneous). The stimulus level increase required to achieve saturation, too, may vary from 20 db to 40 db or more for different fibers. As a consequence, the ensemble of auditory neurons servicing one given area of the basilar membrane may indeed cover the wide dynamic range of the acoustic system through an appropriate division of labor. In a systematic study of the individual characteristics of auditory neurons, Liberman (1978) identified three groups: fibers with high spontaneous firing rates (greater than 20 impulses per second) and having the lowest thresholds; fibers with very low spontaneous rates (less than 0.5/s) and high threshold values spread over a wide range (up to 50–60 db); and a group with intermediate spontaneous rates and thresholds. Each group of fibers has distinct cellular characteristics (diameter and architecture of synaptic contact with inner hair cells). The interesting fact is that each inner hair cell receives fibers from all three groups. This gives a chance to *one* individual sensor cell to deliver its output over a wide dynamic range—and it may be one more reason why so many afferent fibers contact each inner hair cell; see p. 61. According to these results, sound intensity is encoded both in *firing rate* and the *type* of the acoustic nerve fiber carrying the information.

The previous statement must be complemented, though. As the intensity of a pure tone increases, the amplitude of the traveling wave on the basilar membrane increases everywhere, not just in the peak resonance region. This gives a chance to neurons whose characteristic frequency is *different* from that of the incoming sound wave (i.e., wired to hair cells that lie near, but not at, the resonance place) to increase their firing rate when their (off-resonance) thresholds have been surpassed (see Fig. 3.17). In summary, an increase in intensity leads to an increase of the total number of transmitted impulses—either because the *firing rate of each neuron* has increased or because the *total number of activated neurons* has increased. This latter effect depends mainly on the shape of the membrane oscillation amplitude distribution—a purely mechanical property.

The relation between subjective loudness and total firing rate qualitatively explains the main properties of loudness summation (Sect. 3.4). For simultaneous tones of frequencies differing more than a critical band, the grand total of transmitted neural impulses is roughly equal to the sum of the pulse rates evoked by each component separately; hence, the total loudness will tend to be the sum of the *loudnesses* of each tone (relation (3.19)). In contrast, for tones whose frequencies lie within a critical band, with resonance regions on the basilar membrane overlapping substantially, the total number of pulses will be controlled approximately by the amplitude of local basilar membrane vibrations, related to the sum of the original stimulus *intensities* (relation (3.18))

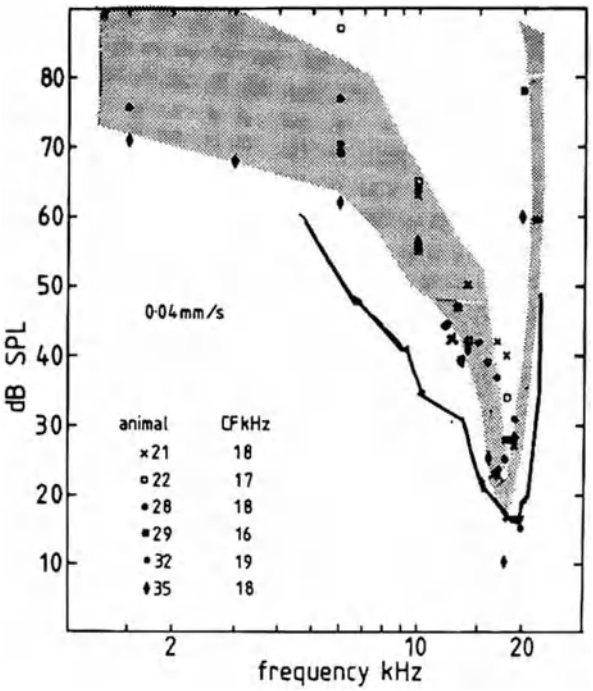



FIGURE 3.17 Comparison of basilar membrane (solid curve) and neural tuning. The gray area corresponds to the tuning curves of ten acoustic nerve fibers wired to the same region of the basilar membrane. By permission, Johnstone et al. (1983).

The dependence of loudness sensation (and masking thresholds) with tone duration (Fig. 3.16) points to a *time-dependent buildup* of the acoustic signal processing operation. Only after several tenths of seconds does the neural mechanism reach a "steady state" of tone processing. It is important to emphasize again that (quite fortunately for music) during this buildup, the pitch sensation is stable and unique from the beginning (except for the first tens of milliseconds).

There is, however, a slight dependence of pitch with loudness, for a tone of constant frequency, mentioned in Sect. 1.2. For tones above about 2000 Hz, pitch increases when loudness increases and vice versa; below 1000 Hz, the opposite happens (e.g., Walliser, 1969). Pitch matching experiments show that an increase of *SPL* from 40 to 80 db causes the pitch of a 5000 Hz tone to increase 2%, or that of a 150 Hz tone to decrease by 1.5% (Terhardt et al., 1982). Listening to a loud final chord in a reverberant environment (e.g., as played on a large organ in a cathedral), many listeners perceive a bothersome overall pitch decrease during the decay phase of the sound. This effect is probably caused by the asymmetry of the distribution of excitation along the basilar membrane (see Fig. 3.5) and by the nonlinear neural response due to saturation—a change in intensity

causes a shift of the central point of excitation (even if the frequency remains constant), leading to a change in primary pitch sensation (Hartmann, 1996). Another result of this asymmetry may be the small, but rather relevant, effect in which the pitch of a pure tone of fixed frequency is found to shift slightly when another tone of different frequency is superposed (e.g., Walliser, 1969; Terhardt and Fastl, 1971). This effect may have some relevant consequences for musical intonation (Sect. 5.4).[14] Note that these pitch changes cannot be explained by a time-cue mechanism!

Here we come again to the question, formulated on p. 41, concerning the *primary* or spectral pitch detection mechanism—if a pure tone of given intensity and frequency sets the basilar membrane into resonant oscillation covering a finite spatial range $\Delta x$, and primary pitch is encoded in the form of spatial position $x$ of the activated fibers, how come only one sensation of pitch is produced? An equivalent mechanism operates in the visual system, too, being responsible for important effects of contrast enhancement (Ratliff, 1972). In the visual system, this "sharpening" process is carried out by a neural network (partly in the retina) whose function is to concentrate or "funnel" the activity into a limited number of neurons surrounded by a region of neural "quietude" or inhibition, thus enhancing contrast. Until the 1970 s, it was believed that an equivalent neural mechanism existed in the acoustic system for pitch sharpening. Yet, recent studies of the cochlea have revealed astounding electromechanical properties of the outer hair cells that are responsible for a nonlinear feedback process that amplifies and sharpens auditory tuning, prior to conversion into neural signals. This process will be the subject of the next section.

## 3.6 Music from the Ears: Otoacoustic Emissions and Cochlear Mechanics

Music *from* the ears?? Well. . .not quite, but almost!

The understanding of the hearing mechanism has progressed through three main epochs. The first was dominated by von Helmholtz' (1877) view that the basilar membrane acted as a spectral analyzer by mechanically sustaining externally driven standing waves (the "little resonators in the ear" picture). The second epoch, from the 1940 s to the 1970 s, was dominated by von Békésy's (1960) experimental results showing that the incoming sound elicits a spatially-confined, hydromechanical traveling wave with the location of maximum amplitude determined by the frequency of the incoming signal (the "waving flag" description, Fig. 3.5). The present epoch began in the seventies with a wide array

[14] We should point out that there is also a shift in pitch when the pressure in the cochlear fluid changes (e.g., pitch shifts perceived during yawning), or when its chemical composition changes (e.g., drug injections into the cerebrospinal fluid).

of experiments and theoretical studies demonstrating that von Békésy's traveling waves are locally amplified by an *electromechanical* process in which the outer hair cells, because of their (hitherto unexpected) motility, function as both mechanical force sensors and mechanical feedback elements (for a comprehensive review, see Dallos (1992)). This cycle-by-cycle amplification works best at *low* signal levels, which accounts for the high sensitivity and dynamic range of the ear. The feedback process may enter into self-oscillation, or resonate after the external stimulus has ceased, and thus lead to cochlear vibrations in the acoustic domain that can be picked up as weak tones by a very sensitive microphone tightly plugged into the external ear canal (Kemp, 1978).

Although of no direct consequence for music, these *otoacoustic emissions* are manifestations of a process which may explain some of the astonishing, in part mutually conflicting, capabilities of the acoustic system:

(1) To detect a sound which displaces the sensor elements by only fractions of a nanometer ($10^{-9}$ m);
(2) To be sensitive to a dynamic range of intensities of at least a billion to one;
(3) To respond to time-varying features of the order of fractions of a microsecond; and
(4) To resolve the frequency of a tone to an accuracy far greater than what the purely mechanical constituents of the basilar membrane can handle.

The new evidence gained at the cellular level of hearing reveals the cochlea as "an evolutionary triumph of miniaturization . . . the most complex mechanical apparatus in the human body, with over a million essential moving parts . . . an acoustic amplifier and frequency analyzer compacted into the volume of a child's marble" (Hudspeth, 1985, 1989).

Von Békésy's measurements of basilar membrane motion were carried out in well-preserved cochleas of dead animals, revealing broad resonance regions for single-frequency tones. When it became possible to record neural impulses from individual live auditory nerve fibers (see Sect. 2.8), it emerged that the latter were much more sharply tuned[15] to a characteristic frequency than what the relatively poor frequency analysis of the basilar membrane would suggest, indicating the action of a sharpening process which, at that time, was thought to be mediated by the neural system, as in vision. More recent measurements using the Mössbauer effect and laser techniques on *live* animals however revealed a considerably sharper tuning of the basilar membrane itself. Fig. 3.17 shows a comparison of basilar membrane and neural tuning (Johnstone et al., 1983). The solid curve represents the *SPL* of the incoming sound wave of a *given frequency* for which the

[15] "Sharp tuning" means that the band of frequencies to which the neuron responds is very narrow; "shallow tuning" would mean that responses occur for a wide range of frequencies.

basilar membrane responds with a given *peak velocity* (0.04 mm/s in the figure) at a *given point* where the radioactive source is located (this is called an isovelocity tuning curve). The deep minimum at about 18 KHz indicates that, at the point in question, the basilar membrane is most responsive to signals of that frequency (only 20 db *SPL* are needed to produce that velocity). One octave lower, i.e., at 9 KHz, one needs a 15 db stronger signal to elicit the same velocity response. A neural tuning curve, on the other hand (the gray area in Fig. 3.17 defines the limits of ten normal tuning curves), represents the *minimum SPL* of a signal of given frequency to which a fiber connected to that point responds (e.g., increases its firing rate above threshold). There is far more similarity between basilar membrane and neural tuning than could be anticipated from the earlier measurements on dead animal cochleas. Microelectrode recordings from inner hair cells show that their tuning curves are very similar to those of the afferent nerve fibers; this clearly indicates that the sharpening process in live cochleas must take place somewhere in the cochlear partition, most likely in the subtectorial space (Fig. 2.7(a)).

Greatly improved Mössbauer and new laser techniques now allow far more precise measurements of the motion of the basilar membrane in live animals (e.g., Johnstone et al., 1986; Ruggero and Rich, 1991; Ruggero et al., 1997). The results show that under real conditions, the tuning curves of the basilar membrane are indeed very similar to those of the inner hair cells or neural fibers associated with the same place on the membrane. Moreover, the measurements show that the sharpness of the tuning curve of the basilar membrane increases dramatically as the *SPL* of the stimulus decreases toward threshold. Indeed, for a given single-frequency tone, an auditory nerve fiber wired to the resonance region increases its response with increasing amplitude of the stimulus tone; but the increase of the response *per input increase* (i.e., the output vs input slope in db/db) is about five times higher at low stimulus levels than at high intensities. Such a "non-linear" response is responsible for both the sharpening of the tuning curves and the compression of the response at high intensity, further contributing to the astounding dynamic range of the ear mentioned in Section 3.5.

Theoretical studies show that the *energy supply* for the amplification mechanism cannot be accounted for by the original sound input (e.g., de Boer, 1983). The sensitivity, sharp frequency selectivity, and nonlinear response properties of the basilar membrane must come from an active mechanical feedback action of the outer hair cells. A convincing fact is that these cells are capable of *changing shape*—mainly their length—at audio frequency rates under electrical stimulation (Kachar et al., 1986). These shape changes affect the traveling wave-induced local deformation of the cochlear partition feeding back energy so that the resulting aggregate wave is amplified. The motility of the outer hair cells is produced by the concerted action of contractile prestin molecules (a protein specifically named after the musical tempo *presto*) found in the cells' lateral membranes; their action speed is so incredibly high as to rule out electromechanical mechanisms similar to those found in normal muscle fibers. Recent measurements of force generation

by the stereocilia (Sect. 2.8, p. 31) of the outer hair cells (more precisely, by coupled bundles of stereocilia) reveal the possible existence of a second mechanism to boost the mechanical oscillation (Kennedy et al., 2005).

While it seems certain that the activation of the cochlear feedback mechanism is locally triggered by the mechanical activation of the outer hair cells themselves, there are strong indications that this process can be modulated by neural commands from the central nervous system via the efferent fibers of the olivocochlear bundle (p. 73). Given the saturation of outer hair cell response at high stimulation levels, this feedback action is limited to low input levels, within 40 db above threshold. It is now clear that to satisfy the amazing demands for speed in the auditory system, signal detection and amplification must be preferentially handled by processes occurring *within* one cell: the acoustic apparatus cannot afford the "leisurely pace" of the nervous system that works on a time scale of several milliseconds or more!

In summary, current thinking (Dallos, 1992; Allen and Neely, 1992; Camalet et al., 2000) about the cochlear amplifier that boosts the mechanical response of the basilar membrane is as follows, for a *single-frequency* tone:

1. The sinusoidal oscillation of the "piston" at the round window (Fig. 2.6(b)) elicits a traveling wave on the basilar membrane that has a maximum amplitude at a frequency-dependent position along the membrane.
2. The local up-and-down oscillation is picked up by the stereocilia of the outer hair cells, a process that triggers an electrical signal in each cell which causes motor protein molecules in its membrane to contract, and/or their stereocilia bundles to flex in synchronism with the applied stimulus. Neural signals from the efferent olivocochlear bundle may influence this process according to commands from the brain.
3. The inner hair cells pick up the mechanical oscillations of the cochlear fluid and transduce them into neural signals in the contacting fibers.
4. This in-phase mechanical reaction of the outer hair cells reinforces (hydrodynamically, or by "rattling" on the tectorial membrane) the local oscillation of the basilar membrane.
5. If some upper limit is reached, the cellular reaction levels off, or saturates; if not. . . go back to step (2)!

Question: Why does this sharpening process work only in the region of maximum resonance of the basilar membrane for that input frequency? After all, away from the place of maximum amplitude of the traveling wave, there still is oscillation, albeit of increasingly lower amplitude (Figs. 2.25(b) and Fig. 3.5). In other words, why is not the *entire* vibration pattern along the basilar membrane amplified? A second tuning mechanism must be operating! Indeed, there are experimental indications that the *stereocilia themselves* are tuned to the *local* resonance frequency and therefore respond efficiently only when the frequency of the passing wave matches the local resonance frequency (e.g., Kennedy et al., 2005);

this happens *only* around the resonance maximum corresponding to the input frequency (Fig. 3.5).[16]

Another puzzling question relates to thermal noise effects. A stereocilium pivots at its base and triggers an excitatory electrical response in the host cell when it is deflected in one specific direction (and an inhibitory one in the opposite direction, while orthogonal responses are ineffective—this explains the phase-specific firing of an auditory nerve fiber shown in Fig. 2.22). What is amazing is that the threshold of hearing occurs for a cilium deflection of only 0.003 degrees! It can be shown that such small displacements should occur even in absence of *any* acoustic stimulation, due to the so-called Brownian motion in the endolymphatic fluid (tiny parcels of fluid jittering around because of thermal equilibrium fluctuations). But the ensuing random noise is mitigated by the fact that stereocilia are linked together in the bundles by extremely fine molecular filaments that make them move as a bundle for concerted action! A pretty universal physical mechanism called *self-tuned criticality* is most likely at work (Camalet et al., 2000).

Finally, let us return to the "music from the ears." As mentioned above, otoacoustic emissions are very weak sounds that can be detected in a tightly closed ear canal; their level is generally far below the threshold of hearing. They are of two basic types: spontaneous emissions and emissions evoked by external sound stimuli. Figure 3.18 shows an example of a spontaneous otoacoustic emission spectrum (Zwicker and Fastl, 1999) (zero *SPL* corresponds to threshold of hearing). Spontaneous emissions vary greatly from individual to individual. About 50% of normal-hearing subjects exhibit one or several such emissions, and most animals have them; nonexistence does *not* indicate an abnormality in hearing (tinnitus, or "ringing of the ears," is unrelated to these emissions in most cases). Their frequency range extends from about 400 to 4000 Hz, although most occur between 1000 and 2000 Hz; for a given subject, their intensity can vary appreciably within a day or even be intermittent on a short time scale, but the frequencies at which they occur are very stable for each cochlea. Evoked otoacoustic emissions are more difficult to measure; they require a sound transmitter in the probe, and more stringent conditions for the analysis of microphone recordings. They can be divided into two classes: emissions that occur simultaneously with a continuous probe tone, and delayed responses to short sound impulses. There is little doubt that otoacoustic emissions are a by-product of the above-described nonlinear feedback process mediated by the outer hair cells; why they occur when they do, and why they appear at selected frequencies, is not known; we just can note that active feedback circuits generally do have the capability of self-excitation or poststimulus resonance. Otoacoustic emissions play no role in music, but they are

[16]Electrophysiological investigations show that the frequencies to which hair cells are most sensitive are inversely related to the lengths of their hair bundles-and it so happens that these lengths increase several times along the basilar membrane from base (high frequency end) to apex (low frequency end). All this is seen in cochleas of lower vertebrates; it is still unclear to what extent it also applies to higher mammals and humans.

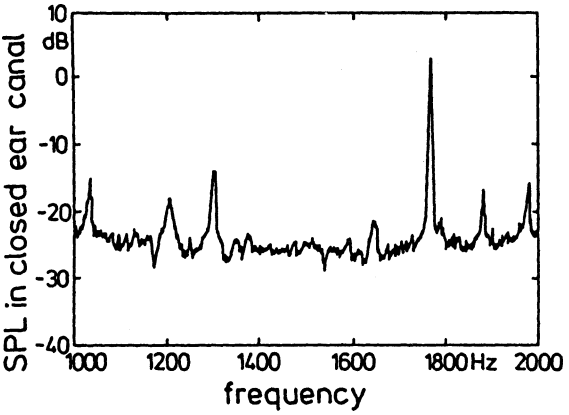



FIGURE 3.18 Example of a frequency spectrum of otoacoustic emissions (Zwicker and Fastl, 1999) recorded by a sensitive microphone in the closed outer ear canal. Zero *SPL* corresponds to threshold of hearing.

an important clinical diagnostic tool; we refer the interested reader to the technical/clinical literature for further information (e.g., Zwicker and Fastl (1999), Gelfand (1990), Ashmore (2008), and references therein.)

# 4
# Generation of Musical Sounds, Complex Tones, and the Perception of Timbre

*". . . it is not surprising that it was possible to put a man on the Moon before the acoustics of a traditional instrument like the piano has been thoroughly explained"*

*A. Askenfelt, in The Acoustics of the Piano,*
Publ. of the Royal Swedish Academy of Music, 1990.

In the preceding two chapters, the two principal tonal attributes pitch and loudness have been analyzed mainly on the basis of pure, single-frequency tones. These are not, however, the tones that play an active role in music. Music is made up of *complex* tones, each one of which consists of a superposition of pure tones blended together in a certain relationship so as to appear to our brain as unanalyzed wholes. A third fundamental tonal attribute thus emerges: Tone quality, or *timbre*, related to the kind of mixture of pure sounds, or harmonic components, in a complex tone (Sect. 1.2).

Most musical instruments generate sound waves by means of vibrating strings or air columns. In Chap. 1, we called these the primary vibrating elements. The energy needed to sustain their vibration is supplied by an excitation mechanism and the final acoustic energy output in many instruments is controlled by a resonator. The room or concert hall in which the musical instrument is being played may be considered as a natural "extension" of the instrument itself, playing a substantial role in shaping the actual sound that reaches the ears of the listener.

In this chapter, we shall discuss how real musical tones are actually produced in musical instruments, how they are made up as superpositions of pure tones, how they interact with the environment in the rooms or halls, and how all this leads to the perception of timbre and instrument recognition.[1] The chapter ends with a brief review of cognitive brain functions relevant to the perception of complex tones.

[1]For a detailed physical and mathematical study of musical instruments with abundant illustrations and literature references, see the treatise by Fletcher and Rossing (1998).

J.G. Roederer, *The Physics and Psychophysics of Music*,
DOI: 10.1007/978-0-387-09474-8_4, © Springer Science+Business Media, LLC 2008

## 4.1 Standing Waves in a String

We consider the case of a tense string, anchored at the fixed points $P$ and $Q$ (Fig. 4.1), of length $L$ and mass per unit length $d$, stretched with a given force $T$ that can be changed ad libitum, say, by changing the mass $m$ of the body suspended from the string as shown in the figure. We now pluck or hit the string at a given point. Two transverse elastic wave pulses will propagate to the left and to the right, away from the region of initial perturbation in a manner discussed in Sect. 3.2, with a velocity given by relation (3.3). These wave pulses, when reaching the fixed anchor points $P$ and $Q$ will be reflected; a positive or "upward" pulse will come back as a negative or "downward" pulse, and vice versa. After a certain time (extremely short, in view of the high speed of the waves in a tense string), there will be waves simultaneously traveling back and forth in both directions along the string. In other words, we will have elastic wave energy "trapped" in the string between $P$ and $Q$, these two points always remaining at rest. If there were no losses, this situation would remain so forever and the string would continue to vibrate indefinitely. However, friction and leaks through $P$ and $Q$ will eventually dissipate the stored energy and the waves will decay.[2] For the time being, we shall ignore this dissipation.

In view of the discussion in Sect. 3.3, we realize that the above picture of waves traveling back and forth along a string strongly resembles the situation arising in a standing wave. Indeed, it can be shown mathematically that *standing waves are the only possible stable form of vibration for a string with fixed ends, with the anchor points P and Q playing the role of nodes.*

This has a very important consequence. Among all imaginable forms of standing waves, only those are possible for which nodes occur at $P$ and $Q$. In other words, only those sinusoidal standing waves are permitted that "fit" an *integer number of times* between $P$ and $Q$ (Fig. 4.2), that is, for which the length of the string $L$ is an integer multiple of the distance between nodes $l_N$ given by relation (3.13). Taking into account this relation, we obtain the condition $L = n\, l_N = n\, \lambda/2$,

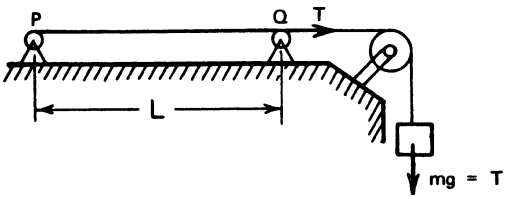



FIGURE 4.1 Demonstration device ("sonometer") for the study of standing waves on a string of length $L$ under controllable tension $T$.

[2] It is this energy "leak rate" through the fixed points (mainly the bridge) that is transformed into sound power in a string instrument's resonance body.

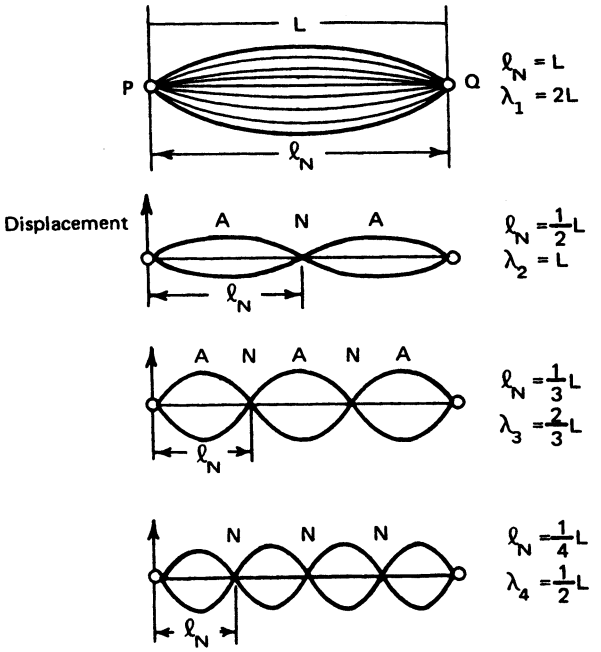



FIGURE 4.2 Standing wave modes in a vibrating string.

where $n$ is any integer number 1, 2, 3. This tells us that only the following wavelengths are permitted (Fig. 4.2):

$$\lambda_n = \frac{2L}{n} \quad n = 1, 2, 3, \ldots \tag{4.1}$$

Using relation (3.8), we find that a string can only vibrate with the following frequencies:

$$f_n = \frac{1}{\lambda_n}\sqrt{\frac{T}{d}} = \frac{n}{2L}\sqrt{\frac{T}{d}} = n\, f_1 \tag{4.2}$$

The lowest possible frequency is obtained for $n = 1$:

$$f_1 = \frac{1}{2L}\sqrt{\frac{T}{d}} \tag{4.3}$$

This is called the *fundamental frequency* of the string. Note in Eq. (4.2) that all other possible frequencies are integer multiples of the fundamental frequency. They are called the upper *harmonics* of $f_1$ (Sect. 2.7). Notice in particular that the first harmonic ($n = 1$) is identical to the fundamental frequency; the second harmonic $f_2$ is the upper octave of $f_1$; the third harmonic is the twelfth (a fifth

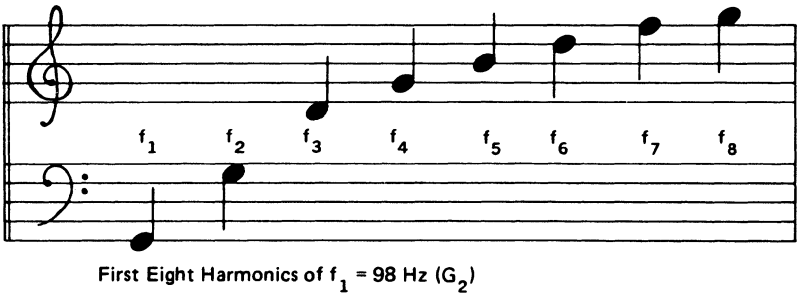



FIGURE 4.3 First eight harmonics of $f_1 = 98$ Hz ($G_2$).

above the octave); the fourth harmonic the fifteenth (double octave); etc. (Fig. 4.3). The upper harmonics are also called *overtones.*[3]

Relation (4.3) tells us that the fundamental frequency of oscillation of a string is proportional to the square root of the tension, that it is inversely proportional to its length and to the square root of its mass per unit length. This explains many characteristic features of piano strings. For the upper part of the keyboard, strings become shorter and shorter (higher fundamental frequency $f_1$; if we have to tune a given string a little sharper, we have to increase the tension (higher $f_1$) and conversely; in the low pitch region, in order to save space and maximize power output, instead of increasing the length of a string, its mass per unit length $d$ is increased (lower $f_1$) by surrounding it with a coil of additional wire. In the violin, where we have only four strings of nearly equal length, each one must have different tension and/or mass in order to bear a different basic pitch. In order to vary the fundamental frequency $f_1$ of a given string, one changes its vibrating length $L$ by pressing the string against the fingerboard, thus introducing a node at the point of contact.

The appearance in a natural way of discrete frequencies related to a fundamental frequency fixed by the conditions of the system, with all other frequencies "prohibited," receives the name of "quantization" and plays a fundamental role everywhere in physics. The discrete set of possible forms of vibration of a physical system is called the *mode* of vibration of the system. The fundamental, the octave, the twelfth, etc., are the first, second, third, etc., modes of vibration of a tense string; their frequencies are given by relation (4.2). In this relation, only quantities depending on the string appear; *the vibration modes are thus a permanent characteristic of the particular physical system.* In which of the possible modes will a given string *actually* vibrate? This is determined by how the vibrations are initiated, that is, by the primary excitation mechanism. Because of

[3]More precisely, *overtones* are the higher frequency components of a complex vibration, *regardless* of whether their frequencies are integer multiples of the fundamental frequency or not, that is regardless of whether relation (4.2) applies.

the capability of linear superposition of waves, *many different modes may occur simultaneously without bothering each other.* In this section, we focus our attention on how a string *can* vibrate. In the next one, we shall examine the question of how a string actually *will* vibrate.

Let us consider the case of a vibrating string in which only *one* mode is excited. This can be accomplished easily in the laboratory by letting an alternating electrical current of a given frequency flow through a tense metal string, spanned through the gap of a strong permanent magnet. Magnetic forces on the current in the string will drive a transverse vibration at the frequency of the current. Whenever this frequency is near one of the string's harmonics, a large standing wave is produced; one can visually observe the nodes and antinodes (as sketched in Fig. 4.2), and clearly hear the sound produced (provided the string is mounted on a resonating box). The use of a stroboscope (a light source that flashes at a given, controllable frequency) enables one to "freeze" the shape of the string, or to observe it in "slow motion."

A more accessible and widely known experiment can be done with a piano. Press slowly, and then hold down, the key of a bass note, say $G_2$ (Fig. 4.3), so that no sound is produced (the hammer does not strike) but the damper remains lifted off that string. Then hit hard and staccato the key of the first upper octave ($G_3$). After that sound has stopped, you very clearly hear the $G_2$ string vibrating one octave higher; it has been excited (by resonance) in its second harmonic mode, $G_3$! Now repeat the same experiment hitting the twelfth ($D_4$) while holding down $G_2$: you hear the $G_2$ string vibrating in $D_4$. Continue with $G_4$, $B_4$, and so on. As a test, hit $A_3$ or $F_3$ while holding down $G_2$—there will be *no* effect: the $G_2$ string remains silent. The reason is that $A_3$ or $F_3$ are not upper harmonics of $G_2$, and the $G_2$ string simply cannot sustain stable vibrations at those frequencies.

Relation (4.2) really is only an approximate one, particularly for the higher order modes. The reason is that the speed of a transverse wave in a string *does* depend slightly on the frequency (or wavelength) of the wave (this is called *dispersion*) and expressions (3.3) and (3.8) are not entirely correct. Indeed, the wave velocity $V_T$ is slightly larger than the value given by relation (3.3). This deviation increases with increasing distortion of the string, that is, it becomes more important for smaller wavelengths and larger amplitudes. The result is that the frequencies of higher vibration modes of a piano string are slightly sharper than the values given by Eq. (4.2).[4] In general, when the frequencies of the higher modes of vibration of a system are not integer multiples of the fundamental frequency, we call these modes *inharmonic.* Vibrating solid bodies other than strings—for example, xylophone bars, bells, or chimes—have many inharmonic vibration modes, whose frequencies are not at all integer multiples of the fundamental frequency. In most of what follows, we shall assume that, for simplicity, the overtones of a vibrating

[4]This does not greatly affect the pitch of the ensuing complex tone (Appendix II); but it does affect tuning of the piano in the treble and bass register (when tuning is done by octaves).

string do coincide with the upper harmonics, and that relation (4.2) is exactly true. Thus, we will often use indistinctly the terms upper harmonics, modes, or overtones, although physically they are different concepts in the case of real strings.

## 4.2 Generation of Complex Standing Vibrations in String Instruments

There are two fundamental ways of exciting the vibration of a tense string: (1) A one-time energy supply by the action of striking (piano) or plucking (harpsichord, guitar); and (2) A continuous energy supply by the action of bowing (violin family). In both cases, the resulting effect is a *superposition of many vibration modes activated simultaneously.* In other words, individual musical sounds generated naturally by strings contain many different frequencies at the same time—those of the vibrating system's harmonics. 4 Figure 4.4 shows how this may arise in practice: adding the first and, say, the third harmonics, one obtains a resulting superposition that at a given instant of time may look as depicted in this figure. Each mode behaves independently, and the instantaneous shape of the string is given by the superposition (sum) of the individual displacements, as dictated by each component separately. It is possible to use the previously mentioned experimental setup of a vibrating string in a magnetic field with an electric current flowing through it, this time using the combined output of *two* sinusoidal voltage generators whose frequencies are set equal to two of the string's harmonics, respectively. With the use of a stroboscope, it is possible to clearly visualize the instantaneous shape of the string when it is vibrating in two modes at the same time. The relative proportion with which each overtone intervenes in the resulting vibration determines to a great extent (Secs. 1.2 and 4.8) the particular character, quality, or timbre of the generated tone. The pitch of a string's complex tone is determined by the fundamental frequency (4.3), as has been anticipated in Section 2.7.

A simple experiment with a piano convincingly shows that a string can indeed vibrate in more than one mode at the same time. Press down slowly, and hold, a given key (say $G_2$) (Fig. 4.3) so as to lift the damper off the string. Then hit hard and staccato *simultaneously* $D_4$, $G_4$, $B_4$. After their sound disappears, it is possible to clearly hear the $G_2$ string vibrate in all three modes simultaneously—

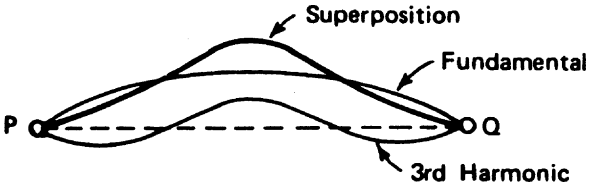



FIGURE 4.4 Superposition of two standing waves (fundamental and third harmonic).

we have one string alone sound a full G major triad! What happens is that three modes (third, fourth, and fifth harmonics) have been excited at roughly similar amplitudes (by resonance). A more drastic experiment is the following: Hold the $G_2$ key down—and hit with your right underarm all black and white keys of two or more octaves above $G_3$—after the initial burst of noise has decayed, the $G_2$ string vibrates beautifully in the dominant seventh chord $G_3$, $D_4$, $G_4$, $B_4$, $D_5$, $F_5$, $G_5$, . . . (Fig. 4.3). None of the other sounded notes could either excite or entertain a stable vibration on the $G_2$ string.

Whereas the above experiment shows that a given piano string *can* vibrate simultaneously in different modes at the same time, the following experiment shows that a piano string, sounded normally, actually *does* vibrate in many harmonic modes. Pick again a bass note, say, $G_2$. But this time, press the $G_3$ key down slowly without sounding it, and keep it down. Then sound a loud, staccato $G_2$. The $G_3$ string starts vibrating in its own fundamental mode. The reason is that this mode has been excited (through resonance) by the *second* harmonic of the vibrating $G_2$ string. If instead of $G_2$ one had sounded $A_2$, the $G_3$ string would have remained silent. Then repeat the same experiment several times, successively pressing silently the keys of $D_4$, $G_4$, $B_4$, $D_5$, . . ., etc. Each one of them will be excited by the corresponding upper harmonic mode of the $G_2$ string.[5]

Many vibration modes appear together when a string is set into vibration. What determines *which* ones and *how much* of each? This is initially controlled by the particular way the string is set into vibration, that is, by the primary excitation mechanism. Depending on how and where we hit, pluck, or bow the string, different mixtures of overtones and, hence, different qualities of the ensuing sound will be obtained. We may explain this on the basis of the following examples. Assume that we give a string the initial form shown in Fig. 4.5(a) (this would be rather difficult to accomplish in practice, though). Since the shape more or less conforms to the fundamental mode (Fig. 4.2), the string will indeed start vibrating in that mode when it is released. If, now the initial form is that shown in Fig. 4.5(b), the string would vibrate in the third mode when released (Fig. 4.2). But what will happen if the initial shape has the far more realistic form of Fig. 4.6, which is obtained when we pluck the string at the midpoint $A$ between $P$ and $Q$? To find out, let us superpose, that is, add linearly, the cases of Figs. 4.5(a) and (b). We obtain the shape shown in Fig 4.7(a) which resembles rather well that of initial configuration of a plucked string (Fig. 4.6). We thus anticipate that the fundamental mode and at least the third harmonic should be simultaneously present in the vibration of a string plucked at the midpoint. We can greatly improve the approximation to the shape of Fig. 4.6 by adding more higher harmonics in appropriate

[5]The main objective of the *pedal* mechanism of the piano is based on this phenomenon: pressing the pedal lifts *all* dampers, and the strings are let free to vibrate by resonance. When one given note is sounded, all those strings will be induced to vibrate that belong to the series of harmonics of that note.

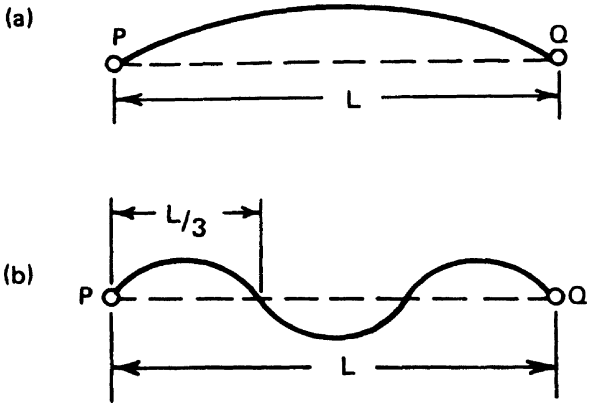



FIGURE 4.5 Initial shape (deformation) of a string needed to cause it to vibrate in the fundamental mode (a) or the third harmonic (b).

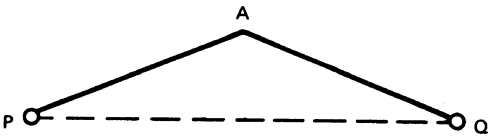



FIGURE 4.6 Initial form given to a string when it is plucked at point A.

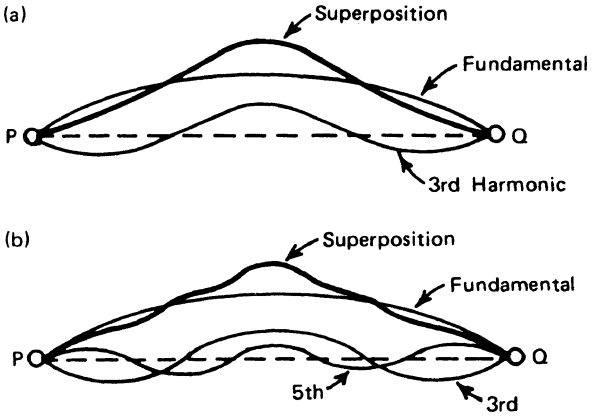



FIGURE 4.7 Superposition of two (a) and three (b) harmonics selected so as to approximate the triangular shape shown in Figure 4.6.

proportions (Fig. 4.7(b)). One can iron out the remaining wiggles shown in this figure by just adding more and more higher harmonics in appropriate proportion until the wanted shape is almost exactly reproduced.

A remarkable fact is that there is no guesswork involved in all this: it can be accomplished in rigorously mathematical form! In fact, it can be shown that *any arbitrary initial shape of a string can be reproduced to an arbitrary degree of accuracy by a certain superposition of geometrical shapes corresponding to the string's harmonic vibration modes* (standing waves). It is this "mathematical" superposition of shapes, in particular, the proportion of their amplitudes and phases, which defines the physical superposition of harmonics with which the string will actually vibrate when it is released from its initial configuration. In other words, each one of the component standing waves, which when added together make up the initial form of the string (e.g., Fig. 4.7(b)), proceeds vibrating in its own way with its own characteristic frequency and amplitude once the string is released. As time goes on, the instantaneous form of the string changes periodically in a complicated way; but every time a fundamental period $\tau = 1/f$ has elapsed, all component modes will find themselves in the same relationship as at the beginning, and the string will have the same shape it had initially. It is very important to remark here that the initial configuration of the string determines not only the amplitudes of the harmonic vibration modes but also their phases (relative timings). The point at which the string is plucked will determine the particular proportion of upper harmonics, that is, the initial timbre of the sound emitted (Sect. 1.2). If we pluck at the center, we will have the situation shown in Fig. 4.7(b), and only *odd* harmonics will appear. On the other hand, the closer to the fixed extremes we pluck, the richer the proportion of upper harmonics will be. In general, all those harmonics that have a node at the plucking point will be suppressed (e.g., all even harmonics in the example of Fig. 4.6), whereas those having an antinode there will be enhanced. This effect is most efficiently exploited by the harp player to control the timbre of the instrument's sound.

In a string that is set in vibration by plucking, we have a situation in which the primary excitation mechanism initially gives a certain *potential energy* to the system, by deforming the string. After release, this initial energy is periodically converted back and forth into kinetic energy of vibration (Sect. 3.1). On the other hand, when a string is set into vibration by striking, a certain amount of *kinetic energy* is initially provided by the striking mechanism (e.g., the hammer in the piano), setting the points of the initially undeformed string in motion. This initial energy is then periodically converted into potential energy of deformation. It can be shown mathematically that *from the knowledge of the initial velocities of the points of the struck string, the ensuing superposition of harmonics can be deduced.* Thus, a string hit at the midpoint will oscillate principally with the fundamental frequency, plus a mixture of decreasing intensities of the odd harmonics. The closer to the end points $P$ or $Q$ a string is struck, the richer in upper harmonics the tone will be. As it happened with the plucked string, harmonics whose nodes are at or near the striking point will be excluded, and those having an antinode there will be enhanced. In the more realistic situation of a piano hammer striking

the string, theory and careful measurements (Hall and Askenfelt, 1988) show that the *duration* of the hammer-string contact influences the mix of upper harmonic modes in a significant way: the longer the contact, the poorer in upper harmonics will be the string vibration (modes with periods shorter than the contact duration will be excluded).

When a string is set into vibration by plucking or striking, we observe the vibration to decay away quite rapidly. This is caused by the action of dissipative forces: elastic friction inside the string and, most importantly, forces that set into small vibratory motion whatever is holding the string in place at its end points. Only part of this energy loss is actually converted into sound wave energy. A freely vibrating string mounted on a rigid, heavy frame produces only a faint sound: most of the vibration energy disappears forever in form of frictional energy (heat). The conversion into sound wave energy can be greatly increased by mounting the string on a board of special elastic properties, called a *resonator* (sound board in the piano, the body of a violin). In that case, the end points of the string are allowed to vibrate a tiny bit (so little that, as compared to the rest of the string vibrations, these end points technically still function as nodes), and the energy of the string can be gradually converted into vibration energy of the board. Owing to the usually quite large surface of this board, this energy is then converted with greater efficiency into sound wave energy. The resulting sound is much louder than in the case of a rigidly mounted string—but *it decays much faster,* because of the considerably increased *rate* at which the available amount of string energy is spent (power, Sect. 3.1).

Let us examine the process of vibration decay in more detail. For simplicity, we consider a string that has been set into free vibration in its fundamental mode only. We focus our attention on the gradually decreasing amplitude of the oscillation of the string, say, at an antinodal point. Measurements show that, for a given string, damped oscillations with larger amplitude decay at a faster rate than those with smaller amplitude. The resulting motion is shown in Fig. 4.8. Notice the slope of the envelope curve, which decreases as the amplitude decreases. This is called an *exponential* decay of amplitude. Most important (and quite fortunate for music!), the frequency of a slowly damped oscillation remains constant.

This is roughly the way a string behaves when it vibrates freely in one given mode after it is struck or plucked. If it is mounted on a rigid board, the energy loss will be relatively small and so will the amplitude damping (Fig. 4.9(a)). If, instead, it is mounted on a sound board, it will give away energy at a larger rate by setting the board and the surrounding air into oscillation. The oscillations will thus decay faster (Fig. 4.9(b)).

A characteristic quantity is the so-called *decay half-time.* This is the time interval after which the amplitude of the oscillations has been reduced to one-half the initial value (Fig. 4.8). The remarkable fact of an exponential decay is that this half-time is always the same throughout the decay: it takes the same time to halve the amplitude, no matter what the actual value of the latter is. The decay half-time is thus a characteristic constant of a damped oscillation. The typical decay half-time of a piano string is about 0.4 s.

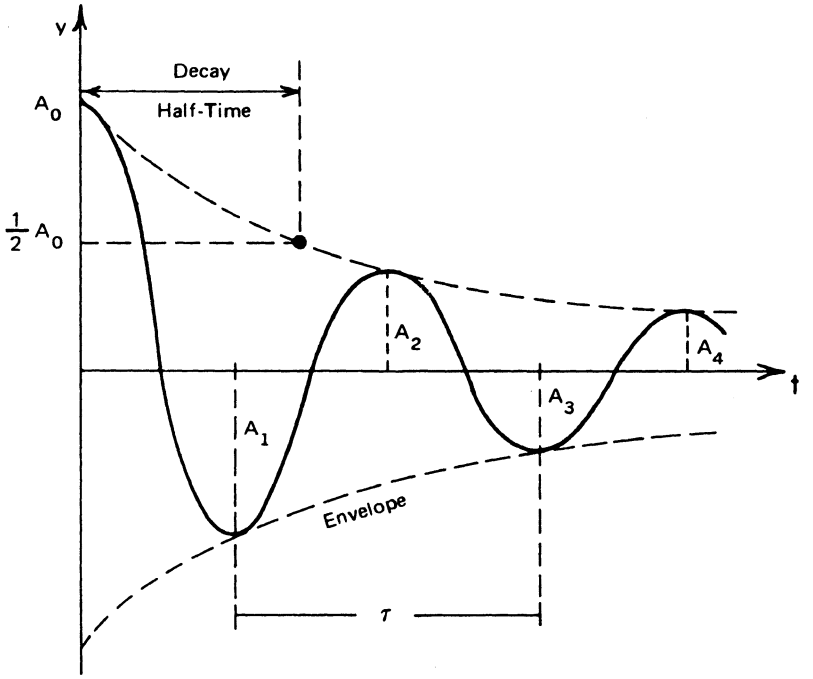



FIGURE 4.8 Graph of a damped harmonic oscillation.

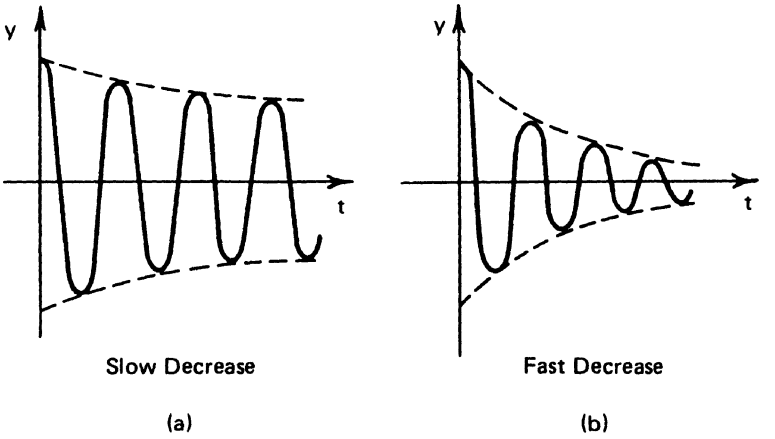



FIGURE 4.9 Slow and fast decays of a damped oscillation.

When a string vibrates in several modes at the same time, the situation is more complex. However, we still find that *each mode* decays exponentially, only that the decay half-time will be different for different modes. The resulting complex sound thus not only decreases in loudness, but also its timbre will gradually change.

In the piano strings, the upper frequency modes decay somewhat faster than the lower harmonics; in a vibrating bell, the lower harmonics continue to sound long after the upper ones have decayed. Otherwise, the overall behavior of a freely vibrating string is exclusively determined by the way in which the vibration has been initially excited (hit or plucked string).

There has been a long-standing dispute between pianists and physicists about what is called "touch" in piano playing. Pianists pay great attention to *the way* a piano key is depressed and contend that this influences the resulting tone far beyond just determining its loudness. The physicist responds that since the hammer is on a *free* flight totally detached from the player during the last portion of its motion, the resulting tone can depend on only *one* parameter: the *speed* with which the hammer strikes the string. Therefore, in the case of a *single* tone, piano touch is nothing but loudness with a timbre that is irrevocably coupled to that loudness[6] and the ensuing decay. All a player can really do is control the final velocity of the hammer; tone quality cannot be changed independently of loudness, and "beautiful" or "bad" touch cannot exist for single tones, says the physicist. The touch that undoubtedly does exist when a musical piece is played is related to other psychoacoustic effects such as subtle tone duration control, small variations of loudness from tone to tone, lifting the melody above the accompaniment, loudness and timing differences of the notes of a chord, even the percussive component given by the "thump" sound of the keys as they hit the stop rail (Askenfelt and Jansson, 1990). There is some hope, though, for the pianists participating in the "touch dispute". Recent measurements (Askenfelt and Jansson, 1990) have revealed that the detailed motion of the freely flying hammer as a *rotating and oscillating elastic body* can be slightly different for different types of touch (more precisely, for different player-controlled accelerations of the hammer prior to its release)! This could lead to a touch-related rubbing motion against the string during contact—but it has not yet been shown that this effect actually does influence in any measurable way the excitation of the string.

What can we do in order to avoid the damping of a string vibration? Obviously, we must compensate the energy loss by somehow supplying extra energy to our vibrating system at a rate equal to the dissipated power. If the supplied power *exceeds* the energy loss rate by a certain amount, the amplitude will gradually

[6]There is a somewhat complicated physical reason for loudness-timbre coupling. As mentioned earlier (p. 122) the duration of the hammer-string contact influences the relative proportion of upper harmonic modes, with a longer contact leading to fewer upper modes. The duration of contact, in turn, depends on the stiffness of the felt on the hammer's head: a softer hammer stays longer in contact with the string than a harder one (for equal impact speed). But there is a remarkable fact (Hall and Askenfelt, 1988): the *effective* stiffness of a given hammer depends on the impact velocity with which the hammer hits the string, with a greater effective stiffness for higher impact velocities and vice versa (this is called a nonlinear behavior of stiffness). As a consequence of all this, hitting a piano key harder will not only increase the amplitude of the string oscillation (louder tone) but shorten the contact time and thus *automatically* increase the proportion of upper harmonics (brighter timbre).

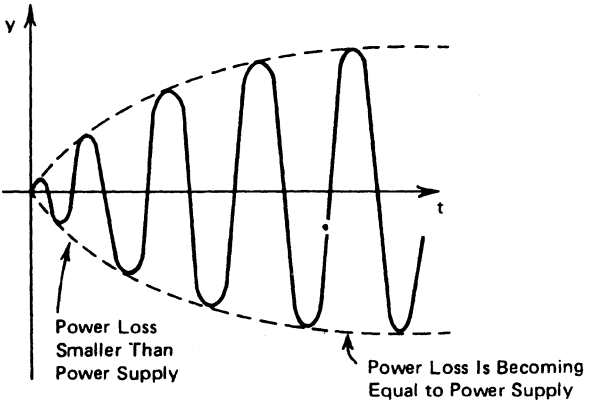



FIGURE 4.10 Buildup of a harmonic oscillation driven at constant power supply.

build up. But this buildup would not go on indefinitely; while the power supply remains constant, the power dissipation will increase as the amplitude increases, and a regime will eventually be attained in which the dissipated power has become equal to the supplied power (Fig. 4.10). This happens during the tone buildup of any instrument of continuous sounding capability (bowed violin string, flute, organ pipe, etc.). In such a case, each harmonic is found to build up independently, as if there were an individual power supply mechanism for each mode. The larger this power supply, the larger, of course, will be the final intensity level.

The bowing mechanism is a good example of how string oscillations can be sustained in a constant regime. The physical problem is mathematically complicated and can be treated only after making several simplifying assumptions (Friedlander, 1953; Keller, 1953). Here, we can only give a qualitative description of the theory. The interaction between the bow and the string is produced by forces of friction. Quite generally, we distinguish two types of frictional interaction. The first is so-called *static* friction that arises when there is *no displacement* between the interacting bodies. This happens when the string "sticks" to the bow, thus moving with the same speed as the latter (or, in a more familiar example, when you try to push a heavy table while it "sticks" to the floor). The second type is *dynamic* friction, arising when the two interacting bodies (their contact surfaces) are *sliding* against each other. This happens when the string "snaps back" and moves in the opposite direction as that of the bow (and happens while you continue to push a table *after* it had started moving). Dynamic friction is weaker than static friction; both mechanisms are controlled by the force, perpendicular to the surface of contact, with which one body presses against the other (and vice versa). In the case of a bowed string, this perpendicular force is called the *bowing pressure—a* dreadful name in the ears of a physicist, because it is *not* a pressure, just a force.

In Appendix I, we discuss in more detail an idealized situation. The main physical conclusions are as follows: (1) T*he amplitude of the vibration of a bowed string (loudness of the tone) is controlled solely by the bow velocity, but in order*

*to maintain constant the nature or type of the string motion (timbre of the tone), one must keep the bowing pressure proportional to the bowing speed.* This is well known by string instrument players, who increase both bow speed $b$ and bowing pressure $P$ at the same time to produce an increase in loudness *without* a change of timbre, or who increase $b$ and decrease $P$, to produce an increase in loudness *with* change in timbre. (2) *A bowed string always has an instantaneous shape that is made up of sections of straight lines* (e.g., Fletcher and Rossing, 1998); this result was verified experimentally long ago. A study of the energy balance in the bowing mechanism reveals that most of the energy given to the string by the bow during the "sticking" portions of the motion is spent in the form of frictional heat (work of the dynamic friction force) during the slipping phases. Only a small fraction is actually converted into sound energy![7]

As in the case of a plucked or struck string, the particular mixture of harmonic modes of vibration will depend on the position of the bowing point. Bowing close to the bridge (*sul ponticello*) will enhance the upper harmonics and make the tone "brighter"; bowing closer to the string's midpoint (*sul tasto*), reduce the intensity of upper harmonics considerably, and the sound is "softer."

In the previous discussion, we have tacitly assumed that the bow is being displaced exactly perpendicular to the string. If it has a small component of parallel motion, *longitudinal* vibration modes of the string can be excited. Their frequency is much higher than the fundamental frequency of the transverse modes; these longitudinal oscillations are responsible for the squeaky sounds heard in beginners' play.

## 4.3 Sound Vibration Spectra and Resonance

When a string vibrates in a series of different modes at the same time, the generated sound waves are complex, too. Each harmonic component of the original string vibration contributes its own share to the resulting sound wave, of frequency equal to that of the corresponding mode and of intensity and phase that are related to the intensity and phase of the latter through the intervening transformation processes. The result is a superposition of sound waves blended together into one complex wave, of fundamental frequency $f_1$ (equal to the fundamental frequency of the vibrating element (4.3)) and with a series of upper harmonics of frequencies $2f_1$, $3f_1$, $4f_1$, etc. The resulting vibration is periodic, repeating with a period equal to $\tau_1 = 1/f_1$. In other words, the fundamental frequency $f_1$ also represents the repetition rate of the resulting complex vibration (Sect. 2.7). The shape of the resulting curve depends on *what* harmonics are present, on *how* much there is of each (i.e., on their relative amplitudes), and on their *relative timing* (i.e., their relative phases).

[7]For a detailed discussion of the bowing of strings, see Schelleng (1973) and Cremer (1984).

And here, we come to a mathematical theorem that had an absolutely smashing impact on practically every branch of physics—including physics of music. In short, the theorem states that *any periodic vibration,* however complicated, *can be represented as the superposition of pure harmonic vibrations,* whose fundamental frequency is given by the repetition rate of the periodic vibration. But that is not all: this theorem also provides the mathematical "recipes" for the numerical determination of the amplitudes and phases of the upper harmonic components! It is called Fourier's theorem, named after a famous 19th century French mathematician. The determination of the harmonic components of a given complex periodic motion is called *Fourier analysis;* the determination of the resultant complex periodic motion from a given set of harmonic components is called *Fourier synthesis.* Similarly, given a complex tone, the process of finding the harmonic components is called *frequency analysis.* Conversely, given a group of harmonic components, the operation of blending them together to form one complex tone is called *sound synthesis*.

Let us discuss one example of Fourier analysis. Of course, we cannot give here the whole mathematical operation that is needed to obtain the numerical results; we will have to accept these and, rather, concentrate on their physical interpretation. We pick the periodic motion shown in Fig. 4.11 corresponding to a sawtooth wave. $\tau$ is the period, $f_1 = 1/\tau$ the repetition rate or fundamental frequency. This type of vibration can be generated electronically. To some extent it represents in an idealized form the motion of a bowed string. Figure 4.12 shows how this periodic motion can be made up of pure, harmonic vibrations. Of course, many wiggles still remain in the curve corresponding to the sum. But this is because, for better clarity, we have stopped adding after only the sixth harmonic. Adding more and more upper harmonics (of amplitude and phase obtained by the method of Fourier analysis) will iron out these wiggles, and a real sawtooth shape will be approached more and more closely. Observe attentively the relative amplitudes and the relative timing (phase) of the harmonic components, and how positive and negative portions add up to give the resultant curve. With a trained eye and intuition, it is possible to estimate qualitatively the principal harmonic components of a periodic motion of almost any arbitrary shape.

We now have to find a way to physically characterize a given complex sound. In principle, we must specify *three distinct series of quantities:* the frequencies of the harmonic components, the pressure variation amplitudes or intensities of the

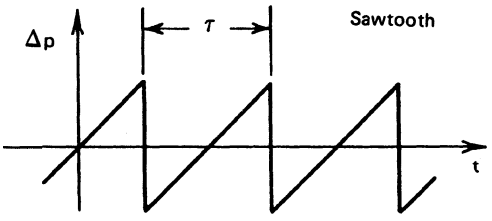



FIGURE 4.11 "Sawtooth" pressure oscillation.

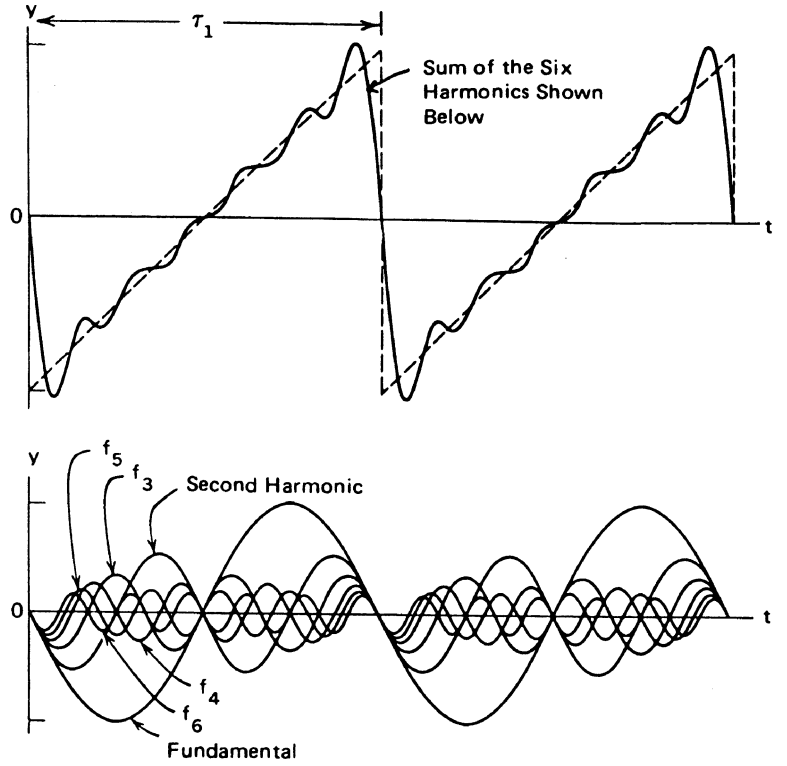



FIGURE 4.12 Fourier analysis (up to the sixth harmonic) of a sawtooth wave.

components, and their phases or relative timings (e.g., Fig. 2.5). In practice, however, it is customary to specify only the fundamental frequency $f_1$, and the intensities of the harmonic components, because, first, it is understood that all upper frequencies are just integer multiples of the fundamental frequency $f_1$, and second, the phases of the components play only a secondary role in timbre perception, particularly for the first (and most important) half-dozen or so harmonics (Sect. 4.8).

The sequence of intensity values $I_1$, $I_2$, $I_3$, ... of the harmonic components of a complex tone represents what is called the *power spectrum* of the tone. Two complex tones of the same pitch and loudness but different spectrum sound differently, that is, have a different tone quality. The difference in spectrum gives us an important cue to distinguish between tones from different instruments—but other cues, particularly tone attack and decay, are also needed for instrument identification (Sect. 4.8). The fact that a multiplicity of physical parameters $(I_1, I_2, I_3, \ldots)$ is related to timbre indicates that the latter is a *multidimensional* psychophysical magnitude.

Tone spectra can be represented graphically by plotting for each harmonic frequency (horizontal axis) the intensity with which that harmonic component intervenes (vertical axis) (e.g., Fig. 4.16). Quite often, values of *IL* (3.15) or

*SPL* (3.16) are used instead of intensities to represent a spectrum. Also, intensity or *IL* values *relative* to that of the fundamental, or relative to the total intensity $I = I_1 + I_2 + I_3 + \ldots$ may sometimes be used. There are many books in which sound spectra of actual musical instruments are reproduced. A word of caution is necessary though. As we shall see (Sect. 4.8), from a psychophysical point of view, a conventional harmonic (Fourier) representation of a tone spectrum makes not much sense beyond about the seventh harmonic, because in that range neighboring components start falling within a critical band. Since this is the elementary acoustic information collection and integration unit of the ear (Sect. 2.4), the auditory system could not resolve the individual intensities of these upper harmonics (see also Fig. 2.25(b)). A psychophysically more meaningful representation of tone spectra is obtained by listing the integrated intensity values *per critical band* (frequency intervals of roughly 1/3 octave extension).

Only steady tones can be "resolved" into a superposition of harmonics of discrete frequencies that are integer multiples of the fundamental. When a vibration pattern is *changing* in time, this is not possible anymore. However, a sort of "expanded version" of Fourier analysis is applicable. It can be shown mathematically that a time-dependent tone leads to a *continuous spectrum* in which all frequencies are represented, with a given intensity for each infinitesimal frequency interval. If the tone is *slowly* time-dependent, discrete frequencies (those of the harmonics) will still be represented with highest intensity (spectral peaks), but if the time change is appreciable from cycle to cycle, the discrete character will disappear and the spectrum will tend to become a continuous curve covering the full frequency range (even if the tone was originally pure). This fact leads to another important observation concerning "high fidelity" equipment (see also p. 46). We noted in Section 1.2, and will come back to this in Section 4.8, that the *transients,* that is, rapid time variations of the vibration pattern of a tone play a determining role in the perception of quality or timbre. Thus, in order to reproduce the transients of a given tone correctly, the recording and reproducing systems must leave the tone spectrum undistorted *over the full frequency range.* Our ear does not need the frequency components lying much over about 5000 Hz of a steady tone, but the reproducing system needs them to give us a correct version of the rapidly *changing* portions of a tone!

The sound spectrum of a string instrument is not at all equal to that of the vibrations of the strings. The reason lies in the frequency-dependent efficiency of the *resonator* (sound board of a piano, body of the violin), whose main function is to extract energy from the vibrating string at an enhanced rate and convert it more efficiently into sound wave power. As mentioned earlier, the string vibrations are converted into vibrations of the resonator in a process in which the fixed end points of the string (particularly the one situated on the bridge) are allowed to vibrate a tiny bit. This remnant vibration is so small that it does not invalidate the fact that, from the string's point of view, these points are still vibration nodes. In spite of being so small, these vibrations *do* involve an appreciable

energy transfer.[8] The explanation is found in the very definition of work (Sect. 3.1): although the displacement of the string's end points is extremely small, the *forces* applied on them are large (of the order of the tension of the string), so that the *product* force times displacement (work) may be quite appreciable. Owing to the large surface of a typical resonator, conversion of its vibration energy into sound wave energy is very efficient—thousands or even millions of times more efficient than the direct conversion of a vibrating string energy into sound.

Like a string, the complex elastic structure of a piano soundboard or body of a violin has preferred modes of oscillation. In this case, however, there is no such simple integer multiple relationship between the associated frequencies as given in relation (4.2). Moreover, there are so many modes with nearly overlapping frequencies that one obtains a whole continuum of preferred vibration frequencies, rather than discrete values.[9] Let us briefly discuss how these vibration modes arise. To that effect, instead of considering the violin body as a whole, we examine the vibration of just one of the plates of the body. To find out the possible vibration modes, it is necessary to excite the plate with a sinusoidal, single-frequency mechanical vibrator, at a given point of the plate (e.g., with which the bridge is normally in contact). Elastic waves propagate in two dimensions away from the excitation point and are reflected at the edges of the plate. The only stable modes of vibration are standing waves compatible with the particular boundary conditions of the plate. This process is very difficult to treat mathematically. In the laboratory, however, it is possible to make the vibrations of the plate visible through a laser technique called holography (Reinicke and Cremer, 1970). The simplest mode of oscillation (called "ring mode") is one in which the central region of the plate moves nearly sinusoidally up and down, with the boundary acting as a nodal line. The ring modes of the violin plates determine the "tap tone," the sound that is evoked by tapping the body. Figs. 4.13 and 4.14 (Jansson et al., 1970) depict holograms of four successive vibration modes of the top plate (with *f* holes and sound post, but without fingerboard) and back plate of a violin, respectively. Each one of the dark curves represents a contour of equal deformation amplitude. The amplitude difference between neighboring fringes is about $2\ 10^{-5}$ cm. In the *assembled* instrument, the vibration modes of the top plate (Fig. 4.13) remain nearly unchanged, but new vibration modes appear in the low frequency range.

The vibration response of a resonator to a given signal of *fixed* amplitude (either from a mechanical vibrator, or from a vibrating string mounted on that resonator) depends strongly on the frequency of the primary oscillations. For that reason, a soundboard reacts differently to vibrations of different frequency. Some

[8]The function of the *mute,* when it is applied to the bridge of a string instrument, is to decrease this energy transfer for the higher frequency components, thus altering the quality of the resulting tone.

[9]Neither does a *real* string, of *finite* thickness, have sharp, discrete modes of oscillation.

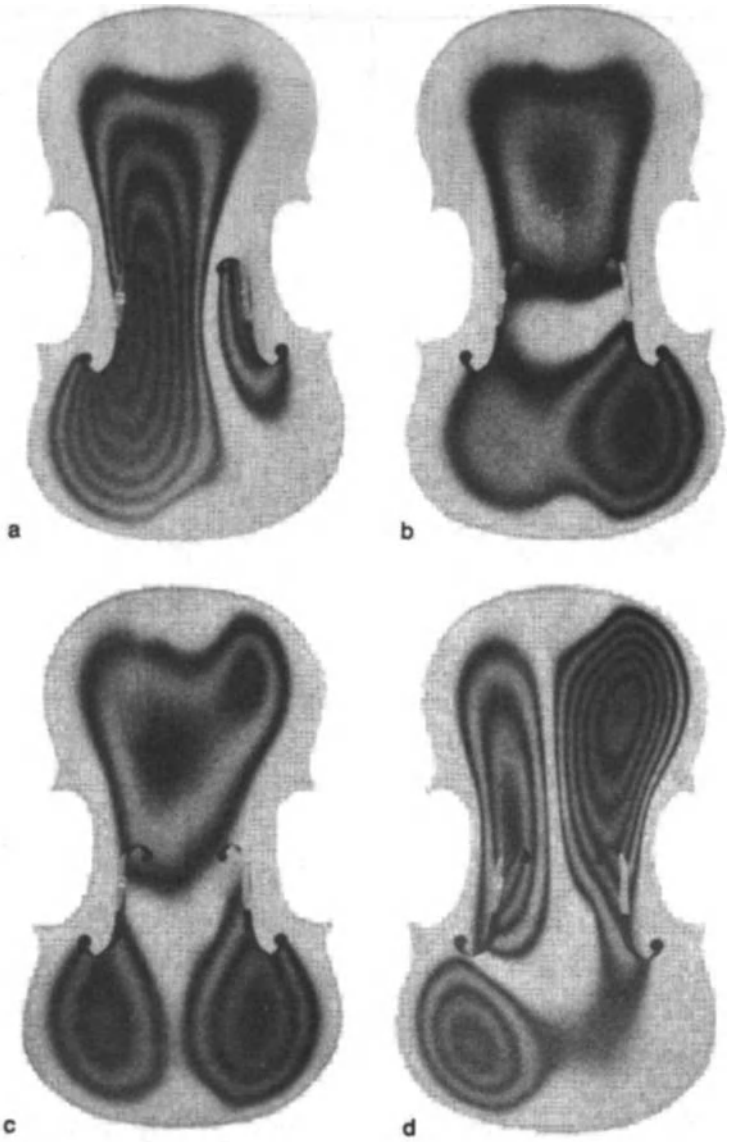



FIGURE 4.13 Holograms depicting the first four vibration modes of the top plate of a violin (with *f*-holes and mounted sound post, without fingerboard). Each one of the dark curves represents a contour of equal deformation amplitude. (a) 540 Hz; (b) 775 Hz; (c) 800 Hz; (d) 980 Hz. Reprinted with permission from Jansson et al. (1970).

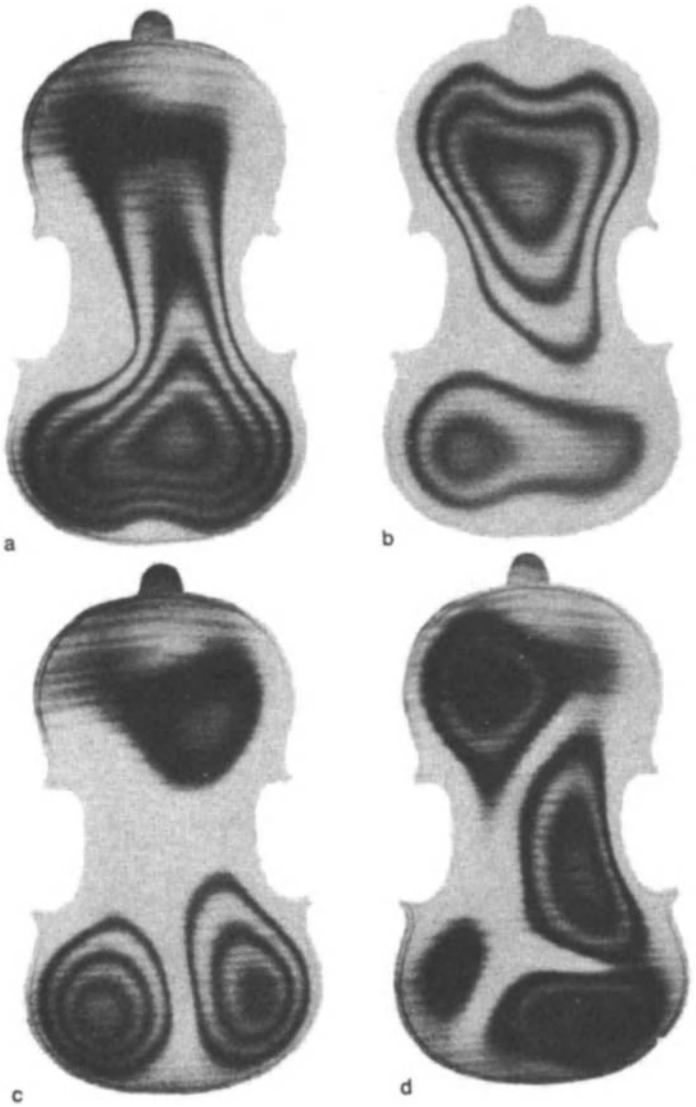



FIGURE 4.14 Same as in Figure. 4.13 for the back plate of a violin. (a) 740 Hz; (b) 820 Hz; (c) 960 Hz; (d) 1110 Hz. Reprinted with permission from Jansson et al. (1970).

frequencies will be preferentially enhanced, whereas others may not be amplified at all. A frequency for which the energy conversion is especially efficient is called a *resonance frequency* of the resonator. A resonator may have many different resonance frequencies; they may be well defined (sharp resonance) or spread over a broad range of frequencies. The graph obtained by plotting the output signal (for instance, as measured by the intensity of the emerging sound wave) as a

function of the frequency of a sinusoidal input vibration of constant Amplitude, is called a *resonance curve,* or response curve.

Usually, the intensity of the output signal $I$ is represented in relation to some given reference signal $I_{\text{ref}}$, and expressed in *decibels* (db) (Sec. 3.4):

$$R = 10 \log \frac{I}{I_{\text{ref}}} \tag{4.4}$$

where $R$ is the value of the response function. The dependence of $R$ with frequency gives the above-mentioned resonance curve. Figure 4.15 is an example corresponding to the plate of a violin (Hutchins and Fielding, 1968). The first sharp rise marked in Fig. 4.15 corresponds to the tap tone.[10] The response curve of an *assembled* violin shows a first resonance peak in the 280–300 Hz range (near the D string pitch) corresponding to the first vibration mode of the *air* enclosed by the body. The next resonance, usually about a fifth above the air resonance, is called the *main wood resonance.* Beyond 1000 Hz, the multiple resonance peaks of an assembled instrument are considerably less pronounced than those of a free plate, shown in Fig. 4.15. The response curve of a piano soundboard is even more complicated; this complication, however, ensures a relatively even amplification over a wide range of frequencies.

Figure 4.15 represents the response curve of a resonator to a single, *harmonic* vibration of given frequency *f.* What happens if it is excited by a string vibrating with a complex spectrum of harmonics, of frequencies $f_1$, $f_2 = 2f_1$, $f_3 =$

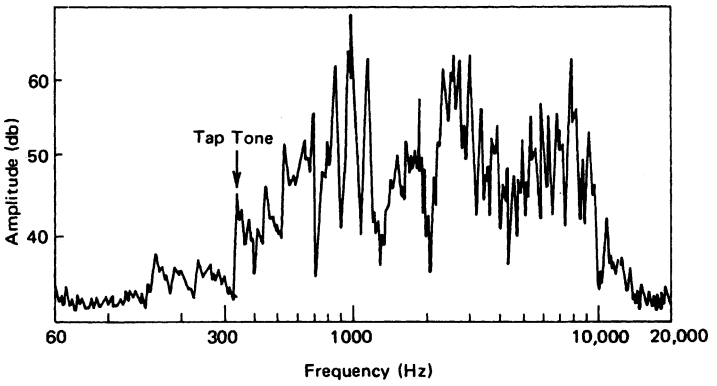



FIGURE 4.15 Resonance curve of a violin plate (Hutchins and Fielding, 1968). Permission from *Physics Today* is acknowledged.

[10]The position (in frequency) and the shape of this particular resonance peak is of capital importance for the quality of a string instrument (Hutchins and Fielding, 1968). See also p. 157.

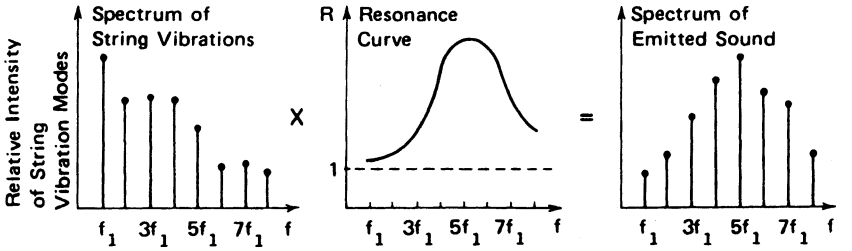



FIGURE 4.16 Effect of a soundboard with a hypothetical resonance curve on the spectrum of a complex string vibration.

$3f_1$ ..., etc., and intensities $I_1$, $I_2$, $I_3$, ...? Each harmonic component will be converted independently, as prescribed by the value of the response function $R$ corresponding to its own frequency. The timbre of the resulting sound is thus governed by both, the original string vibration spectrum *and* the response curve of the resonator.

As an example, consider the hypothetical spectrum of a vibrating string shown in Fig. 4.16. This string is mounted on a hypothetical soundboard, the response curve of which is also shown. The output sound spectrum is given in the right graph (relative values). The fundamental is considerably reduced; instead, the fifth harmonic appears enhanced above all others. According to Fig. 4.16, in this example, more power would be extracted from the fifth harmonic than from any other. If the string had originally been plucked or struck, this harmonic would decay faster than the others because its energy reservoir would be depleted faster. This leads to a time-dependent change in spectrum or tone quality as the sound dies away. If, on the other hand, the string were to be bowed, the energy loss in each mode would automatically be compensated for by the bowing mechanism, the resulting tone quality remaining constant in time.

Finally, we come to a point most important for music. The response curve of a resonator is an immutable characteristic of a musical instrument. If, for instance, it has a resonance region around, say, 1000 Hz, it will enhance all upper harmonics whose frequencies fall near 1000 Hz, no matter *what* note is being sounded (provided, of course, that its fundamental frequency lies below 1000 Hz) and no matter what the original string vibration spectrum was. A broad resonance region that enhances the harmonics lying in a fixed frequency range is called a *formant.* A musical instrument (its resonator) may have several formants. It is believed that formants, that is, the enhancements of harmonics in certain fixed, characteristic, frequency intervals are used by the auditory system as a most important "signature" of a complex tone in the process of *identification* of a musical instrument (Sect. 4.9). One of the reasons in favor of this hypothesis is the fact that formants are an invariable characteristic common to all tones of a given instrument (whose fundamental frequency lies at least one octave below the formant peak),

whereas the spectrum of individual tones may vary considerably from one note to another.

## 4.4 Standing Longitudinal Waves in an Idealized Air Column

Let us consider a long, very thin cylinder, open at both ends Fig. 4.17. The air inside can be considered as a unidimensional elastic medium (Sect. 3.2) through which longitudinal waves can propagate. At any point *inside* the cylinder, the pressure is allowed to momentarily build up, decrease, or oscillate considerably with respect to the normal atmospheric pressure outside—the rigid walls and the inertia of the remaining air column hold the necessary balance to the forces (3.1) that arise because of the pressure difference. But at the open end points *P* and *Q,* no large pressure variations are allowed even during the shortest interval of time, because nothing is there to balance the arising pressure differences. These points thus must play the role of *pressure* nodes, and any sound wave caused by a perturbation inside the pipe and propagating along it will be reflected at either open end. We hence have a situation formally analogous to the vibrating string, discussed in

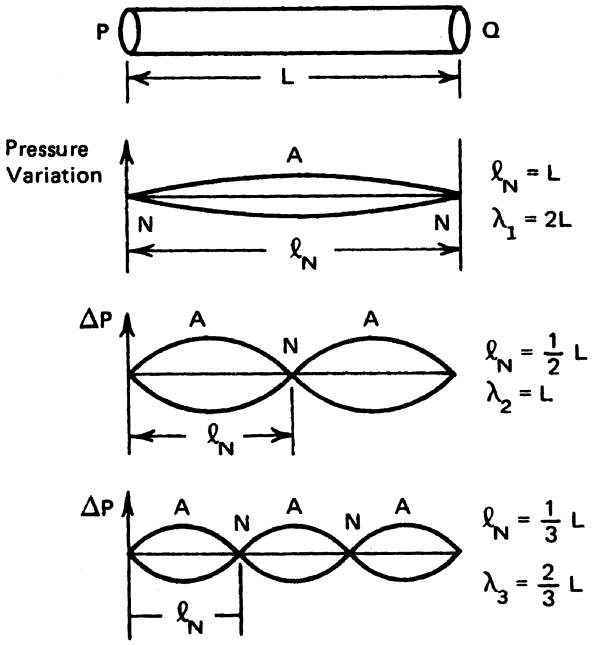



FIGURE 4.17 Standing wave modes (pressure variations) in an idealized cylindrical pipe, open at both ends.

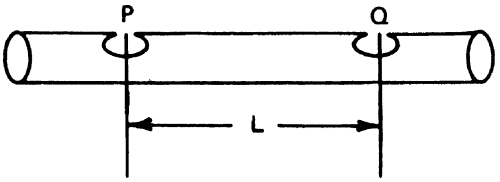



FIGURE 4.18 Elementary pipe with two holes.

Section 4.1. Sound waves generated in the open pipe remain trapped inside and *the only possible stable vibration modes are standing longitudinal waves with pressure nodes at the open ends P and Q* (Fig. 4.17). Notice that in view of our discussion in Section 3.3 on p. 93, the open end points are *antinodes of displacement,* that is, points with maximum vibration amplitude.

The open air column does not necessarily have to be physically defined in the manner shown in Fig 4.17. For instance, there is an open air column comprised between points *P* and *Q* of the pipe shown in Figure 4.18. Indeed, since there are holes at *P* and *Q,* the air pressure at these points must remain constant and equal to the external pressure. *P* and *Q* thus play the role of open ends of the enclosed air column. Figure 4.18 corresponds to the case of an idealized flute, where *P* is the mouth hole and *Q* the first open finger hole.

In a real open pipe of finite diameter, *the pressure nodes do not occur exactly at the open end,* but at a short distance further out ("end correction," p. 145). The relations given below are thus only first approximations.

From Fig. 4.17 and relation (3.6), we obtain the frequencies of the vibration modes of an open cylindrical pipe:

$$f_n = \frac{n}{2L} 20.1\sqrt{t_A} = nf_1 \quad n = 1, 2, 3, \ldots \tag{4.5}$$

$f_1$ is the fundamental frequency

$$f_1 = \frac{10.05}{L}\sqrt{t_A} \tag{4.6}$$

Remember that $t_A$ is the *absolute* temperature of the air in the pipe, given by Eq. (3.5). *L* in Eqs. (4.5) and (4.6) must be expressed in *meters.* Taking into account that the wavelength $\lambda_1$ of the fundamental tone is related to the length *L* of the tube by $\lambda_1 = 2L$ (Fig. 4.17) and inspecting Fig. 3.8, one may obtain an idea of typical lengths of open-flue organ pipes, flutes, and recorders as a function of frequency. An increase in frequency (pitch) requires a decrease in length. Relation (4.6) also shows the effect of air temperature on the fundamental pitch of a vibrating cylindrical air column. An increase in temperature causes an increase in frequency (sharper tone). Thus flutes and flue organ pipes must be tuned at the temperature at which they are expected to be played. Fortunately, the fundamental frequency (4.6) is controlled by the absolute temperature $t_A$,

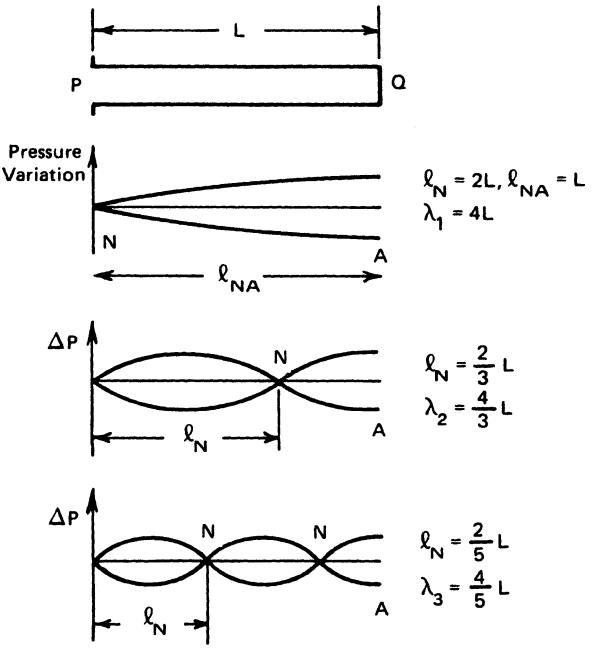



FIGURE 4.19 Standing wave modes (pressure variations) in an idealized cylindrical pipe, closed at one end.

appearing under a square root. Both facts make the influence of temperature variations on pitch a rather weak one, but enough to be concerned with—as flautists and organists well know.

We now turn to the case of a narrow cylinder closed at one end (Fig. 4.19). We realize that, whereas at the open end *P* the pressure must remain constant and equal to that of the outside air (pressure node), at the stopped end *Q* the inside pressure can build up or decrease without restriction. Indeed, a *pressure antinode* is formed at *Q*. This is more easily understood by considering the actual vibratory motion of the points of the medium. Quite obviously, there must be a *vibration node* for all air molecules near *Q:* they are prevented from longitudinal back-and-forth oscillation by the cover of the pipe. According to the discussion in Sect. 3.3, such a vibration node corresponds to a pressure antinode.

Figure 4.19 shows how the standing wave modes "fit" into a stopped pipe in such a way as to always have a pressure node at the open end and a pressure antinode at the stopped end. For the fundamental frequency, we find the relation

$$f_1 = \frac{20.1}{4L}\sqrt{t_A} = \frac{5.03}{L}\sqrt{t_A} \tag{4.7}$$

($L$ is in meters, $t_A$ is the absolute temperature (3.5)). This is exactly *one-half* the fundamental frequency (4.6) of an open pipe of the same length. In other words,

an idealized stopped cylindrical pipe sounds an octave below the pitch of a similar pipe open at both ends.

With respect to the higher modes of a stopped cylindrical pipe, an inspection of Fig. 4.19 (and conversion of wavelength to frequency) reveals that only *odd* multiples of the fundamental frequency $f_1$ (4.7) are allowed:

$$f_1; \quad f_3 = 3f_1; \quad f_5 = 5f_1; \quad \ldots \tag{4.8}$$

The frequencies $2f_1, 4f_1, 6f_1, \ldots$ are forbidden—their modes cannot be stably sustained in an ideally thin stopped cylindrical pipe. In other words, *the overtones of a stopped pipe are odd harmonics of its fundamental.*

The clarinet is perhaps the most familiar example of an instrument behaving very nearly like a stopped cylindrical pipe. The mouthpiece with the reed behaves as the closed end, the bell, or the first open finger hole defining the open end. The fundamental pitch of a note played on the clarinet indeed lies one octave below the note corresponding to the same air column length, played on a flute.

Organs include several stopped pipe ranks. One of the reasons is money and space savings: open bass pipes are very long (according to relation (4.6) an open $C_1$ pipe is 5.3 m high). The same pipe, if stopped, only needs to be 2.65 m long. Of course, there is more than money at stake: a stopped pipe yields a different quality of sound than an open one of the same fundamental frequency.

Our last case under discussion here is a (very narrow) conical pipe, stopped (closed) at the tip *P* (Fig. 4.20). The determination of the modes of vibration requires a rather complex mathematical analysis. The results can be summarized in a few words: an idealized narrow conical pipe stopped at the tip has the same vibration modes as an open pipe of the same length: relations (4.5) and (4.6) apply. A *truncated* (narrow) cone (Fig. 4.21), closed at the end *P,* has a series

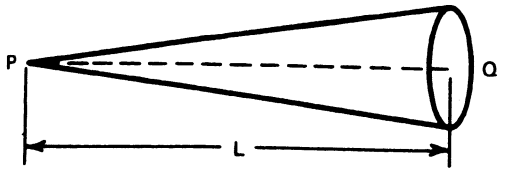



FIGURE 4.20 Conical pipe.

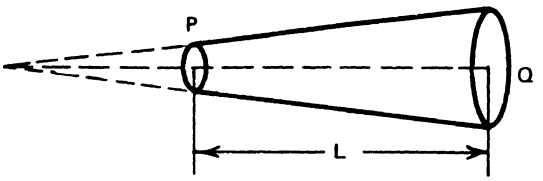



FIGURE 4.21 Truncated cone pipe closed at P.

of vibration modes that do not bear the integer number relationship: in the lower frequency range (near the fundamental), they correspond nearly to the modes of an open pipe of the same length *L,* but for higher frequencies they approach those of a closed cylindrical pipe of length *L.* In other words, the vibration modes are *inharmonic* (p. 117).

## 4.5 Generation of Complex Standing Vibrations in Wind Instruments

The results of the preceding section are idealizations that would become true only for hypothetical air columns (cylinders, cones) of diameters that are very small as compared to their length *L.* This, however, is not the case with real musical instruments and organ pipes. Besides, the air cavities in these instruments are cylindrical or conical only along a certain portion of their length, with more complex shapes near the mouthpiece and the open end (open finger holes, bells, etc.).

In order to analyze the physical behavior of real wind instruments, we must inspect in more detail all phenomena involved. Let us first turn to the *excitation mechanism.* There is no equivalent to the "plucking" or "hitting" of a string in the case of an air column. The reason is that the vibrations of a freely oscillating air column decay almost instantaneously. One may verify this easily by tapping with the hand on one end of an open pipe (the longer the better), or by knocking sharply on its wall, while holding the ear near the other end. A sound burst of pitch equal to the fundamental frequency of the pipe can indeed be heard, but the decay takes place within a fraction of a second. It is thus necessary to have a primary excitation mechanism equivalent to the bowing of a string, which continuously supplies energy to the vibrating air column at a given rate.

There are two distinct types of such mechanisms. The first one consists of a high speed air stream blown with velocity $v$ against a rigid, sharp edge $E$ (Fig. 4.22), located at a certain distance $d$ exactly above the slit $S$. This system is aerodynamically unstable: the air stream alternates back and forth between both sides of the edge, breaking into "rotating puffs" of air called "vortices" or "eddies," which travel upward along both sides of the edge. As the velocity of the

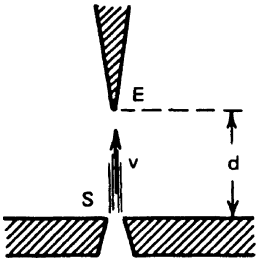



FIGURE 4.22 Generation of an edge tone.

stream increases, vortices are created at an increasing rate. Since they represent a periodic perturbation of the ambient air, sound waves are generated when the vortex generation rate falls into the audio domain. The resulting sound is called an *edge tone.*[11] The edge tone mechanism is the primary excitation process for all wind instruments of the flute family and for flue organ pipes. The air stream oscillations are in general complex; for very small flow intensities, they become nearly sinusoidal. The fundamental frequency of a free edge tone depends on the air stream velocity $v$ and the distance to the edge $d$ (Fig. 4.22). In the low frequency range, it is proportional to the ratio $v/d$, that is, it increases with increasing $v$ and with decreasing $d$.

The other excitation mechanism of importance in music is the *reed,* a thin plate made of cane, plastic, or metal, placed in front of a slit of nearly the same shape and of slightly smaller size than the reed (Fig. 4.23). As air is blown into the cavity from below (i.e., the pressure therein is increased), the excess air flows through the small space between the slightly lifted reed and the slit into the shallot. During this flow, the reed is drawn toward the slit.[12] This eventually interrupts the flow; the flow pressure overcomes the reed's own elasticity, opening the slit again, and the whole game starts anew. In other words, the reed starts oscillating back and

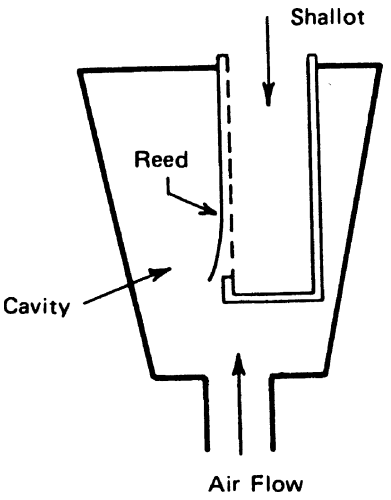



FIGURE 4.23 Reed mechanism for a reed-stop organ pipe.

[11] Vortices are even formed in absence of the edge, provided the slit S is small enough and the velocity $v$ high enough (see Fletcher and Rossing, 1998). This represents the basic physics of the *human whistle* where slit size (lip opening) and velocity of the air stream (blowing pressure) determine the fundamental frequency.

[12] By the difference in *dynamic* pressure (not static pressure) on both sides of the reed—the same effect that keeps a flying aircraft aloft!

forth, alternately closing (partially or totally) and opening the slit. The air moves in periodic puffs into the shallot, giving rise to a sound called *reed tone.* The fundamental frequency of a *free* reed tone depends on both the elastic properties of the reed and the excess pressure in the cavity (blowing pressure). In general, the vibratory motion of a free reed is complex, except at very small amplitudes for which it is nearly sinusoidal. Some instruments (oboe, bassoon) have *double reeds,* beating against each other. Also, the lips of a brass instrument player can be considered as a (very massive) double reed system.

The edge and reed tones discussed above are seldom used alone (free edge and reed tones). In the woodwind instruments, they merely serve as the primary excitation mechanism, the energy supplier to the air column in a pipe. In those cases, not only the spectrum, but also the frequency of the vibrations of the air stream or the reed are controlled by the air column via a (nonlinear) feedback mechanism. This is accomplished by the sound waves in the air column: the first compressional wave pulse to be produced travels along the pipe, is reflected at the other end (open or closed), and comes back to the mouthpiece (as a rarefaction pulse in open pipes, as a compression pulse in closed ones). There it causes a pressure variation that in the case of woodwinds "overrides" all other forces (aerodynamic or elastic) and thus controls the motion of the air stream or the reed. The resulting pitch is quite different (usually much lower) from that evoked when the corresponding edge or reed tone mechanism is activated freely, in absence of the pipe. This is quite different from the case of a string mounted on a soundboard, whose pitch remains practically unaffected by the resonator. In the case of brass instruments, the mass of the player's lips is so large that the feedback from the pipe can only influence, but not override, their vibration; the latter must be controlled by the player himself by adjusting the tension of his lips. There are a few musical instruments with open reeds (accordion, harmonica, reed organ).

The tone buildup process in a wind instrument is very complicated, but it is of capital importance for music. In many instruments, upper harmonics build up faster than the fundamental; sometimes this can be artificially enhanced, giving a characteristic "chiff" to the resulting tone.

To understand the *steady state* sound generation in woodwinds, brasses, and organ pipes, it is necessary to analyze the resonance properties of their air columns and the coupling of the latter with the primary excitation mechanism (air stream, reed, or lips). To that effect, let us state the following experimentally verified facts: (1) The primary excitation mechanism sustains a periodic oscillation that is complex, of a certain fundamental frequency, and with a series of harmonics of given spectrum. (2) Fundamental frequency and spectrum of the primary oscillations are controlled (influenced, in the case of brass instruments) by the resonance properties of the air column; the total amplitude of the oscillations is determined by the primary energy supply (total air stream flow, blowing pressure). (3) The spectrum of the pressure oscillations outside the instrument (generated sound wave) is related to the internal spectrum by a transformation that is governed by the detailed form and distribution of finger holes and/or by the shape of the bell (for

a detailed discussion, see for instance Benade (1990), and Fletcher and Rossing (1998)).

To explore the resonance properties, that is, the resonance curve, of the air column in a given wind instrument, one must devise an experimental setup in analogy to the alternating-current-driven metal string (Sect. 4.1) or the vibrator-excited violin plate (Sect. 4.3). This is accomplished by replacing the natural excitation mechanism with a mechanical oscillation driver (e.g., an appropriate speaker membrane) and by measuring with a tiny microphone the pressure oscillation amplitudes in the mouthpiece (where a pressure antinode is formed in this setup). The resonance curve is then obtained by plotting the pressure oscillation amplitudes as a function of frequency, for *constant* driver oscillation amplitude. The measured amplitudes are usually expressed in decibels *(R* in expression (4.4)), referred to some standard level. Curves of this type are also called *input impedance* plots.

Figure 4.24 sketches typical resonance curves obtained for clarinet-type and oboe-type air columns (Benade, 1971) (*without* mouthpieces, bells, and open

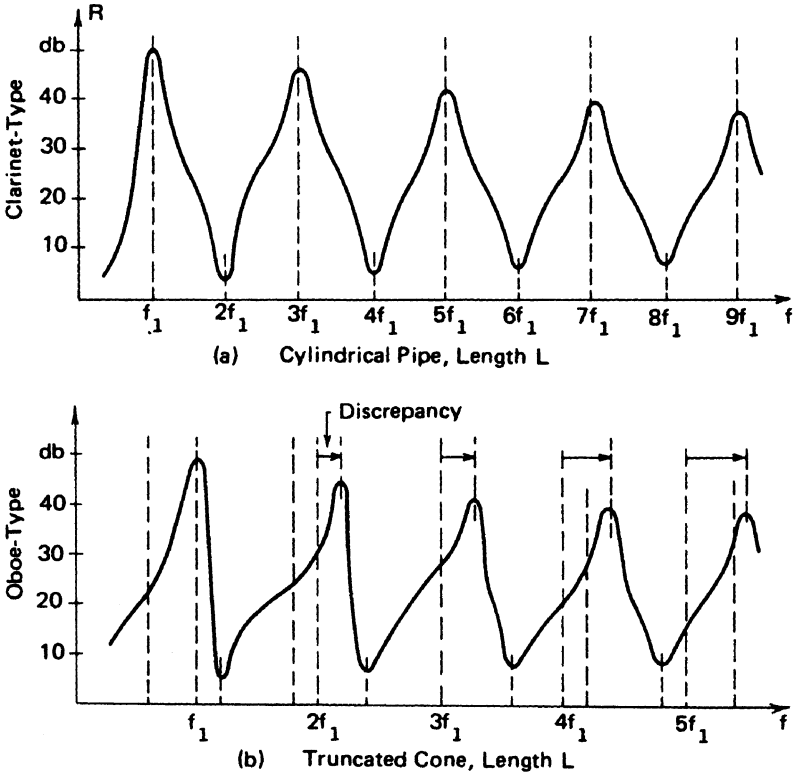



FIGURE 4.24 Typical resonance curves (after Benade, 1971) for clarinet-type (cylindrical) and oboe-type (conical) air columns (without mouthpiece, bell; closed finger holes).

finger holes). It is important to note that the resonance peaks obtained in this manner correspond to the vibration modes of an air column *closed* at the end of the primary driver, that is, to a real case in which a *reed* is used as excitation mechanism at one end of the column (pressure antinode, vibration node, at the place of the reed). To find the resonance curve of the same air column corresponding to the case in which it is excited by an air stream (flutes, recorders, organ flue pipes), it is sufficient to plot the negative $-R$ of the values obtained in the previous measurement:[13] resonance peaks become dips and dips become peaks (just turn Fig. 4.24 upside down). The main justification for this procedure is that, in the mouthpiece, the pressure antinode (that appears when a reed is used) is replaced by a vibration antinode, that is, a pressure node, in the case of a flute with an open mouth hole.

Note in Fig. 4.24 that resonance peaks are not at all "sharp"; possible oscillation modes thus do not correspond to unique, discrete frequencies, as it appeared to be in the case of infinitely thin air columns (Sect. 4.4). Furthermore, in the case of the truncated cone (b), the resonance peaks are asymmetric and inharmonic (see discrepancy from the harmonic multiples $2f_1$, $3f_1$, etc.). Toward high frequencies, the resonance peaks of the truncated cone resemble those of the cylinder. On the other hand, if the cone were complete (till the tip), the resonance peaks would all lie close to the dips of curve (a) (even harmonics of the cylinder) with very little inharmonicity.[14]

Let us discuss qualitatively how the resonance curve controls the primary excitation mechanism, say, of a reed. For very low intensities (small amplitudes of the reed), its motion is nearly sinusoidal, and, in principle, any resonance peak frequency (Fig. 4.24) could be evoked. In practice, however, it is found that only the frequency corresponding to the *tallest* resonance peak is excited at very low intensity level (pianissimo). Usually this is the peak with the lowest resonance frequency; the evoked sound pertains to the *"low register"* of the instrument.

As the amplitude of the reed oscillation increases (by increasing the blowing pressure), the nonlinear character of the feedback from the air column destroys the harmonic, sinusoidal, vibration of the reed, upper harmonics appear with increasing strength (in general, the intensity of the $n$th harmonic grows proportionally to the $2n$th power of the intensity of the fundamental), and the resulting sound becomes "brighter."

At the same time, the fundamental frequency readjusts itself if the upper resonance peaks are somewhat inharmonic. The rule governing this pitch readjustment is the following: the fundamental frequency locks into position in such a way as to *maximize the weighted average height of all resonance values* $R_1$, $R_2$, $R_3$, ...[15] corresponding to the harmonics $f_1$, $2f_1$, $3f_1$, ... (Benade, 1971). If, for instance,

[13] Only if $R$ is expressed in *decibels.*

[14] In real instruments, clarinet-type resonance curves (a) also display a discrepancy from harmonicity (Backus, 1974).

[15] Weighted with the corresponding spectral intensity values $I_1$, $I_2$, $I_3$, ...

the upper resonance peaks deviate from harmonicity as shown in Fig. 4.24(b), the pitch of the tone must become sharper as its intensity increases, in order to accommodate the set of increasingly important upper harmonics and place each of them as close as possible to a resonance peak. Due to this effect, a truncated cone will not work as a woodwind unless something is done to minimize the inharmonicity.[16]

An interesting situation arises with the clarinet-type resonance curve (Fig. 4.24(a)). There, the resonance peaks are situated at only odd integer multiples of the fundamental (see also Sect. 4.4). Hence, all even harmonic components of the reed oscillation will be strongly attenuated. Starting from a piano-pianissimo in the low register (single-frequency excitation at the fundamental peak) and gradually increasing the blowing pressure will at first tend to evoke the second harmonic. Its energy, however, will be efficiently drained because of the dip at the corresponding frequency (Fig. 4.24(a)). The resulting increase in loudness (and "brightness") will thus be considerably less for a given increase in blowing pressure than in an oboe-like air column (curve (b)), where the second harmonic can build up unhindered. This is why transitions from *ppp* to *pp* are more easily managed in a clarinet than in an oboe or saxophone.

Finally, another noteworthy fact in Fig. 4.24 is the almost identical location of the *dips* in both curves. When these dips are converted into peaks by plotting $-R$ instead of $R$ to obtain the resonance curve of these air columns when they are excited with an air stream (open mouth hole), one obtains a practically identical harmonic series in both examples. Consequently, cylinders and truncated cones can be used almost indistinctly to make flute-type instruments.

So far, we have considered standing waves in which the fundamental frequency is determined by the first (lowest frequency) peak of the resonance curve, yielding the low register tones of a woodwind instrument. In the "middle register", the fundamental frequency lies close to the second resonance peak. In reed instruments, this is realized by decreasing the size of the first resonance peak below that of the second and by shifting its position away from the harmonic series. The register hole or speaker hole accomplishes this function. In the flute, this transition, or "overblow," is accomplished by means of a change (increase) of the air speed blown against the wedge. Notice that in the first overblow of a clarinet-type reed, the pitch jumps to the third harmonic or *twelfth* (second peak, Fig. 4.24(a)), whereas conical-bore reeds (and all flutes) have their first overblow on the *octave* (second harmonic, Fig. 4.24(b)). Another overblow leads to the "top register" of the woodwinds, with fundamental frequencies based on the third and/or fourth resonance peaks. To accomplish this with a reed instrument, the first two resonance peaks must be depressed and shifted in frequency to destroy the harmonic relationship.

---

[16]It is important to understand clearly this "automatic adjustment process" for another reason: it is remarkably similar to one that has been proposed for the pattern recognition mechanism in the central pitch processor, where the set of shifting harmonics $f_1$, $2f_1$, $3f_1$, . . . . is called the "template" (see Secs. 2.9 and 4.8, and Appendix II).

*Organ pipes* work on essentially the same principles as a flute (flue pipes, open and stopped) or as a reed woodwind (reed pipes). The main difference is that since there is one pipe for each note for a given stop, tone holes and overblowing are unnecessary. Organ pipes are always operated in the low register (with a few exceptions in "romantic" organs). Resonance curves of open-flue organ pipes have peaks located near the integer multiples of the fundamental frequency, with a slight inharmonicity which depends on the ratio $r$ = diameter/length. Stopped pipe resonance curves resemble that of the upper graph in Fig. 4.24 with maxima at odd multiples of the fundamental. The larger the value of $r$, the greater will be the inharmonicity of the upper resonances. As a result, there will be a (usually downward) shift in the fundamental frequency of the resulting tone, plus an increasing attenuation of the higher harmonics (which will be increasingly displaced away from the inharmonic resonance peaks). Consequently, *the sound of wide organ pipes is less rich in upper harmonics* ("flutey" sound). Narrow pipes (small $r$) have resonance peaks that lie closer to the integer multiples of the fundamental frequency, and there will be a stronger excitation of the higher harmonics (the sound is bright or stringy). The fundamental frequency is slightly misplaced with respect to the value given by relation (4.6) (open pipe) or (4.7) (stopped pipe). These relations, however, may still be used if a correction of value $0.3 \times$ diameter is added, for each open end, to the length $L$ ("end correction").[17] Reed organ pipes range from the type in which the reed vibration is strongly controlled by the feedback from the air column (e.g., trumpet family stops) to a type in which the reed vibration is practically autonomous (regal family stops).

## 4.6 Sound Spectra of Wind Instrument Tones

Resonance characteristic of the air column and excitation mechanism collaborate to determine the power spectrum and the intensity of the standing wave in the bore of the instrument. Figure 4.24 showed two hypothetical resonance curves; real instruments, however, display a more complicated behavior due to the peculiar shape of the mouthpiece (and mouth pipe), the shape and distribution of open finger holes, the effect of the bell, and, in the case of flutes, the effect of air speed on the width and position of the resonance peaks (Benade, 1971). Here, we can only summarize briefly the most important effects. The *finger holes*, besides of course determining the effective length of the air column and, hence, the absolute position of the resonance peaks, are partly responsible for a *cutoff* of the resonance peaks above 1500–2000 Hz. This cutoff has an important effect on timbre (attenuation of high harmonics) and on the dynamic control of loud woodwind tones, particularly in the middle and high registers. In the oboe, the reed cavity and the constriction in the staple contribute to decrease the inharmonicity of the resonances of a truncated cone.

[17] For a mathematical treatment of air pressure oscillations in cylindrical pipes see Fletcher and Rossing (1998).

The brasses deserve special attention in this section. As already pointed out, the feedback mechanism is less efficient in the determination of the fundamental frequency, and the player must set the buzzing frequency of his lips near the wanted frequency in order to elicit the right pitch. In a brass instrument, the upper harmonics are created by the oscillating resonance properties of the mouthpiece caused by the alternate opening and closing of the lips (Backus and Hundley, 1971) rather than by a feedback-controlled nonsinusoidal motion of the latter. Mouthpiece, tapered mouth pipe, main cylindrical bore, and bell combine in such a way as to yield a characteristic resonance curve rather different from that of a typical woodwind. Figure 4.25 gives an example (Benade, 1971). Notice the marked cutoff frequency (mainly determined by the bell) and the large hump of peaks and dips in the mid-frequency range (mainly governed by the shape of the mouthpiece). This hump plays a crucial role in shaping the tone quality of brass instruments. Finally, the first resonance peak lies *below* the fundamental frequency (marked by the arrow) that corresponds to the rest of the peaks. Notice

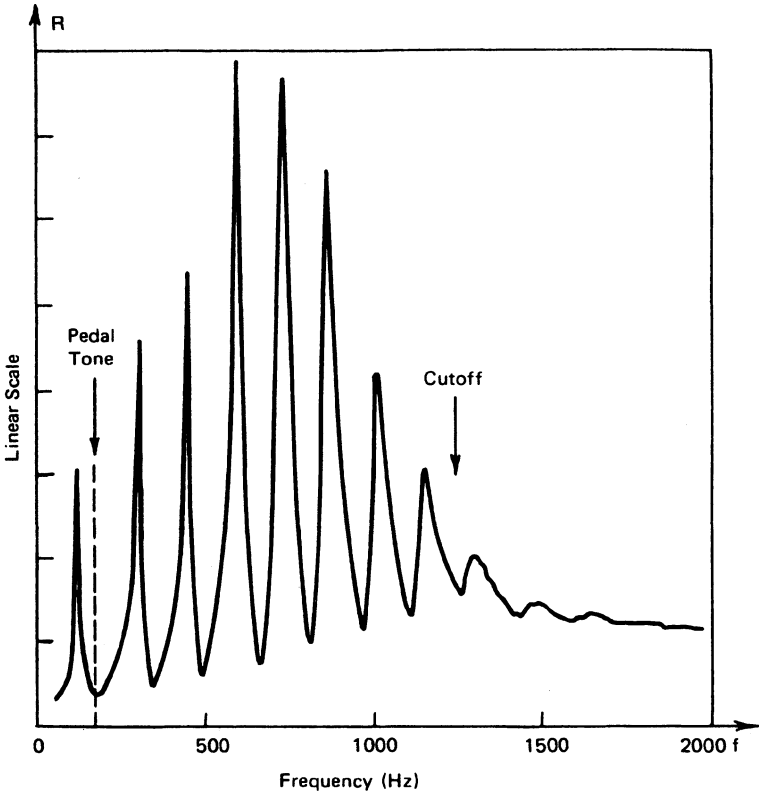



FIGURE 4.25 Resonance curve of a trumpet (Benade, 1971) (given in linear scale). By permission of Professor A. Benade.

also the characteristic asymmetry of the peaks in the low frequency range (similar to those of a truncated cone in Fig. 4.24(b)), as compared to the high frequency region (where their shape is inverted).

Brasses do not have tone holes to alter the effective length of their air columns—changes in pitch are mainly effected by overblowing, that is, making the fundamental frequency jump from one resonance peak to another. This is accomplished by appropriately adjusting the lip tension. Up to the eighth mode can be reached in the trumpet, and the sixteenth mode in the French horn. To obtain notes between resonance peaks, a system of valves offers a limited choice of slightly different pipe lengths. In the trombone, a continuous change of tube length (hence of pitch) is possible through the slide. Since the lowest resonance peak is out of tune with the rest of the almost harmonic series of peaks, it cannot be used. Rather, the fundamental frequency of the lip vibration is set at the value of the *missing* fundamental (arrow in Fig. 4.25) that corresponds to the second, third, etc., peaks. This leads to the so-called "pedal note" of a brass instrument (used only in the trombone). It can be played only at considerable loudness levels.

The spectral composition of the sound waves emitted by a wind instrument is different from that of the standing vibrations sustained in its air column. The bell and/or the open finger holes are mainly responsible for this spectral transformation. This transformation leaves the spectral composition above the cutoff frequency relatively unchanged, while it tends to attenuate the lower harmonics. In other words, woodwind and brass tone spectra are richer in higher harmonics than the vibrations actually produced inside the instrument (Benade, 1973, 1990).

Some wind instrument tone spectra have formants, that is, characteristics that are independent of the fundamental frequency of the tone (Sect. 4.3). The bassoon and English horn are examples, with (not too well-defined) spectral enhancement around 450 and 1100 Hz, respectively. These formants, however, are caused by the excitation spectrum characteristics of the *double reeds;* they are not determined by the resonance properties of the instrument's bore. Although not an explicit topic in this book, we must mention here the *human voice* as the most notable example of a "wind instrument" in which formants play a crucial role: they are the determining characteristic of all *vowel sounds.* Formants in the human voice are mainly determined by the resonance properties of the nasopharyngeal cavity (Flanagan, 1972). The shape of this cavity determines which of two main frequency ranges of the vocal chord vibrations are to be enhanced. These, in turn, determine whether the outcoming sound is "ah," "eh", "ee," "oh," "uh," etc.

## 4.7 Trapping and Absorption of Sound Waves in a Closed Environment

Musical instruments are usually played in rooms, concert halls, auditoriums, and churches. The sound a listener perceives under these conditions is not at all identical to the one emitted by the instrument. For this reason, the enclosure in which an

instrument is played may be considered a natural extension of the latter, with the difference that whereas a given musical instrument has certain immutable acoustic properties, those of the enclosure vary widely from case to case and from place to place. The subject of room acoustics is as important for music as is the physics of musical instruments.

To analyze the effect of an enclosure on a musical sound source placed somewhere inside, let us consider a musical instrument at position *S* and a listener at position *L* (which may coincide with *S* if the listener is the player) in a room of perfectly reflecting walls (Fig. 4.26). The instrument starts playing a given note at time $t = 0$, holding its intensity steady afterwards. We assume that the sound is emitted equally into all directions (which in reality never happens in practice). As sound waves propagate away from *S,* the listener will receive a first signal after the short interval of time *SL/V* which it takes the *direct* sound to travel from *S* to *L* (for instance, if $SL = 10$ m, $V = 334$ m/s (relation (3.6)), the time of direct arrival is 0.03 s; for the player it is practically zero). As we shall see in Section 5.1, the direct sound plays a key role in the perceptional process (*precedence effect*). Immediately thereafter, reflected waves (trajectories 2, 3, 4, 5, etc.) will pass through point *L* in rapid succession (in the figure, the reflections on the floor and the ceiling have been ignored). The first few reflections, if very pronounced and well separated from each other, are called echoes. This game continues with secondary, tertiary, and multiple reflections (not shown in the figure). As time goes on, and the instrument keeps sounding, acoustical energy passing through point *L* will keep building up. If there were no absorption at all, the sound waves would "fill" the room, traveling in all directions and the acoustic energy emitted by instrument would accumulate and remain trapped in the enclosure;

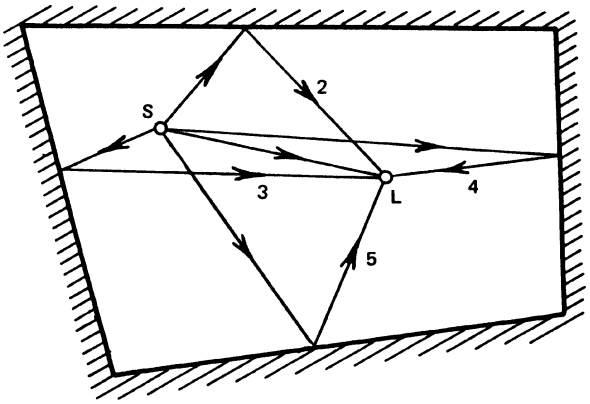



FIGURE 4.26 Example of different propagation paths of a sound wave from source point *S* to a listener *L*.

loudness would thus build up gradually at any point inside.[18] In the real case, of course, there is absorption every time a sound wave is reflected—and lost through any openings in the enclosure. Hence, the sound wave intensity will not increase indefinitely but level off when the power dissipated in the absorption and escape processes has become equal to the rate at which energy is fed in by the source (a similar situation to the vibration buildup in a bowed string, Fig. 4.10). This equilibrium intensity level $I_m$ of the diffuse sound is much higher than that of the direct sound (except in the neighborhood of the sound source).

When the sound source is shut off, an inverse process develops: first, the direct sound disappears, then the first, second, etc., reflections. Figure 4.27 schematically depicts the behavior of the sound intensity at a given point in a typical enclosure. The sound decay, after the source has been shut off, is called *reverberation* and represents an effect of greatest importance in room acoustics. This decay is nearly exponential (e.g., see Fig. 4.8); quite arbitrarily, one defines *reverberation time* as the interval it takes the sound level to decrease by 60 db. According to Table 3.2, this represents an intensity decrease by a factor of one million. Desirable reverberation times in good medium-sized concert halls are of the order of 1.5–2 s. Longer times would blur too much typical tone successions; shorter times would make the music sound "dry" and dull (see Sect. 4.8).

We may discuss a few simple mathematical relationships that appear in room acoustics. Let us imagine an enclosure of perfectly reflecting walls with no absorption whatsoever, but with a built-in *hole* of area $A$. Whenever the maxi-

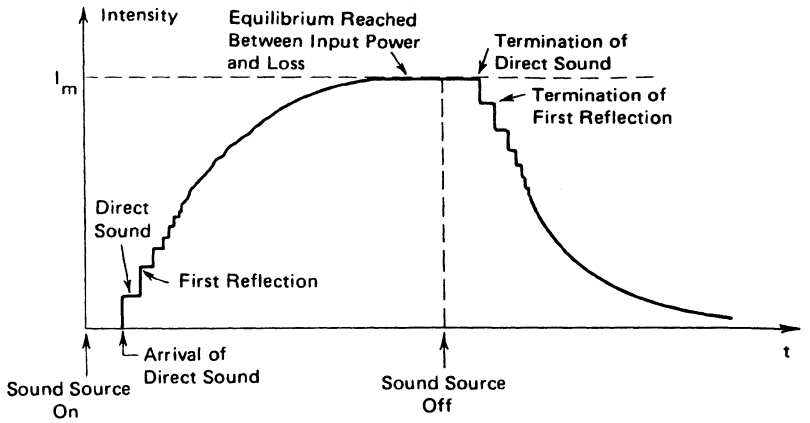



FIGURE 4.27 Typical tone intensity buildup and decay in a hall (linear scale).

[18]For almost any kind of shape of rooms and positions of the source therein, there may be regions practically inaccessible to sound waves emitted from $S$ (blind spots), or regions into which sound waves are focused (e.g., the focal points in elliptic enclosures).

mum intensity $I_m$ is reached (Fig. 4.27), acoustic energy will be escaping through the hole at a rate given by the product $I_m A$.[19] Since this corresponds to the steady state in which the power $P$ supplied by the instrument equals the energy loss rate, we can set $P = I_m A$, or

$$I_m = \frac{P}{A} \tag{4.9}$$

In a real case, of course, we do not have perfectly reflecting walls with holes in them. However, we still may *imagine* a real absorbing wall as if it were made of a perfectly reflecting material with holes in it, the latter representing a fraction $a$ of its total surface; $a$ is called the *absorption coefficient* of the wall's material. A surface of $S$ square meters, of absorption coefficient $a$, has the same absorption properties as a perfectly reflecting wall of the same size but with a hole of area $A = Sa$. Absorption coefficients depend on the frequency of the sound (usually increasing for higher frequencies), and have values that range from 0.01 (marble, an almost perfect reflector) to as much as 0.9 (acoustic tiles). Taking all this into account, we can rewrite relation (4.9) in terms of the actual wall surfaces $S_1, S_2, \ldots$ with corresponding absorption coefficients $a_1, a_2, \ldots$:

$$I_m = \frac{P}{S_1 a_1 + S_2 a_2 + \cdots} \tag{4.10}$$

This relation can be used to estimate auditorium sizes needed to achieve wanted values of $I_m$, for a given instrument power $P$, and a given distribution of absorbing wall materials.

The reverberation time $\tau_r$ is found to be proportional to the volume $V$ of the hall and inversely proportional to the absorbing area of the walls $A = S_1 a_1 + S_2 a_2 + \ldots$. Experiments show that, approximately,

$$\tau_r = 0.16 \frac{V}{S_1 a_1 + S_2 a_2 + \cdots} \tag{4.11}$$

with $V$ in cubic meters, $S$ in square meters, and $\tau_r$ in seconds. Since the absorption coefficients usually increase with sound frequency, $\tau_r$ will decrease with increasing pitch: bass notes reverberate longer than treble notes.

One of the problems in room acoustics is that the *audience* greatly influences (increases) the absorption properties of a hall. This must be taken into account in the design of auditoriums. In order to minimize the effects related to the unpredictable size of an audience and its spatial distribution, it would be necessary to build a seat whose absorption coefficient is nearly independent of whether it is occupied or not. The absorbing effect of the audience is maximally felt in enclosures with very long reverberation times, such as in churches and cathedrals. No

[19] It is assumed here that $I_m$ represents the diffuse, *omnidirectional* sound energy flow.

single performer is so exposed to (and harassed by) a changing acoustic environment as an organist.

Tone distribution, buildup, and decay, as well as the frequency dependence of the absorption coefficients, have a profound effect on the physical characteristics of musical tones emitted by an instrument in a real environment, hence on the perception of music by the listener. The time dependence of tones is deeply affected: transient characteristics are altered, and, for instance, a staccato note becomes stretched in time depending on the reverberation properties of the hall. The tone spectrum is also affected, because the absorption coefficients are all frequency-dependent. Finally, taking into account that the phases of the waves passing through a given point in a field of reverberant sound are randomly mixed, it can be shown that the resulting SPL of each harmonic component will also fluctuate at random, introducing a limit to the listener's ability to recognize timbre in a closed musical environment (Plomp and Steeneken, 1973).

There are other second-order effects, usually neglected, related to a wave phenomenon called *diffraction.* When a sound wave hits an obstacle (e.g., a pillar in a church or a person sitting in front of the listener), three situations may arise: (1) If the wavelength of the sound wave is much smaller than the size (diameter) of the obstacle (e.g., a high-pitch tone) (Fig. 4.28(a)), a sound "shadow" will be formed behind the obstacle, with normal reflection occurring on the front side. (2) If obstacle and wavelength are of roughly the same magnitude, a more complicated situation arises, in which the obstacle itself acts as a sound re-emitter, radiating into all directions (not shown in Fig. 4.28). (3) If the wavelength is much larger than the obstacle (Fig. 4.28(b)) (e.g., bass tones), the latter will not affect the sound wave at all, which will propagate almost undisturbed. Regular arrangements of obstacles (as the distribution of seats or people in the audience) may lead to interference patterns for given wavelengths and directions of propagation. Finally, standing waves may build up for certain configurations of the enclosure, certain frequencies, and certain positions of the source. This leads to the formation of annoying nodes and antinodes (Sect. 3.3) in the room. For scientific and technical details of room acoustics, see Ando (1985).

Diffraction and the precedence effect play an important role in electroacoustic sound reproduction. Electronic compensation for sound wave diffraction effects

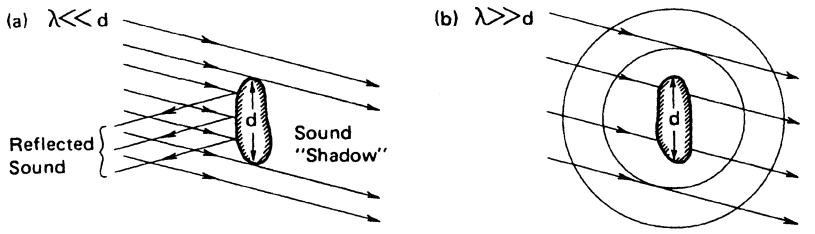



FIGURE 4.28 Sound wave interaction with an obstacle. (a) Short wavelengths (reflection and/or absorption); (b) Long wavelengths (diffraction).

at the head of the listener is desirable when stereophonic signals are received from two loudspeakers (Damaske, 1971). A correct stereo reproduction of the "direct sound" (i.e., the precedence effect) is necessary to prevent headphones from giving the sensation of a sound image localized "inside the head" (see also p. 67).

## 4.8 Perception of Pitch and Timbre of Musical Tones

Whereas considerable research has been done on the perception of pitch and loudness of *pure* tones (Secs. 2.3, 2.9, 3.4, and 3.5), much remains to be done in the study of the perception of complex tones (e.g., see Yost and Watson (1987) and Plack et al. (2005)). That the timbre of a tone can be modified by reinforcing certain overtones has been known for many centuries. Indeed, genuine tone synthesis was first performed by pipe organ builders in the 13th or 14th century. Organs of those times did not have multiple stops; rather, each key sounded a fixed number of pipes called "Blockwerk," composed of one or several pipes tuned to the fundamental pitch of the written note, plus a series of pipes tuned to the octave, the twelfth, the fifteenth, etc., respectively, following the series of upper harmonics (excluding the much feared seventh). The particular combination of loudness chosen for each component pipe determined the particular quality of sound of the instrument. Later on, the first multiple hand-activated stops appeared: they allowed the organist to selectively turn on or off the various ranks of pipes corresponding to the upper harmonics in the Blockwerk, and thus choose from among several options the particular timbre of sound of the organ (and alter the loudness—Sect. 3.4). It was only one or two centuries later that new independent stops were added, in the form of ranks of pipes of individually different timbre.[20]

Synthesis of sound is thus rather old hat. However, *analysis* of sound, that is, the individualization of upper harmonics that appear simultaneously in a naturally produced tone, was not explicitly mentioned in the literature until 1636, when the remarkable French scientist-philosopher-musician, Père M. Mersenne (Mersenne, 1636), published the first study of the (qualitative) analysis of the upper harmonics present in a complex tone.

Two main questions arise regarding perception of complex tones: (1) Why does a complex tone, made up of a superposition of different frequencies, give rise to only one pitch sensation? (2) What is it that enables us to distinguish one tone spectrum from another, even if pitch and loudness are the same? Although we have already

[20]Italian organs in the Baroque have preserved the basic timbre control through the inclusion of many mutation (upper harmonic) stops; much later, the sounds of the first electronic Hammond organs were based entirely on the possibility of a separate intensity control of individual, electronically generated, harmonics.

answered in part the first question (Sect. 2.9),[21] it is useful to re-examine once more the perception process of a *complex* sound wave impinging on the eardrum. The eardrum will move in and out with a vibration pattern dictated by the complex, nonsinusoidal but periodic vibration pattern of the wave. This motion is transmitted mechanically by the chain of ossicles to the oval window membrane, which reproduces nearly the same complex vibration pattern. Neither eardrum nor bone chain "know" that the vibration they are transmitting is made up of a superposition of different harmonics. This analysis is only made in the next step.

The complex vibration of the oval window membrane triggers traveling waves in the cochlear fluid. This is the stage at which the separation into different frequency components takes place. As shown in Sections 2.3 and 3.2, the resonance region for a given frequency component (region of the basilar membrane where the traveling wave causes maximum excitation) is located at a position that depends on frequency. A complex tone will therefore give rise to a whole multiplicity of resonance regions (Fig. 2.25), one for each harmonic, the actual positions of which can be obtained using Fig. 2.8 as a guide. In view of the near-logarithmic relationship between $x$ and $f$, the resonance regions will crowd closer and closer together as one moves up the harmonic series (Fig. 2.25(a)). Because each resonance region is extended over a certain length (Sect. 2.4), overlap between neighboring resonance regions will occur, particularly for higher harmonics (Fig. 2.25(b)). Actually, as explained in Section 2.9, beyond about the seventh harmonic, all resonance regions overlap and it is increasingly difficult to hear them out (Plomp, 1964).[22]

A single complex tone thus elicits an extremely complicated situation in the cochlea. Why, then, do we perceive this tone as one entity, of well-defined pitch, loudness, and timbre? As explained in Section 2.9, this may be mainly the result of a *spatial pattern recognition process*. The characteristic feature that is recognized in this process, common to all periodic tones regardless of their fundamental frequency and Fourier spectrum, is the *distance relationship* between oscillation maxima on the basilar membrane. The pitch sensation is to be regarded as the "final output signal" of this spatial recognition process. This central pitch processor mechanism ("template fitting") can work even if part of the input is missing (e.g., a suppressed fundamental). In such a case, it may commit matching errors or yield ambiguous or multiple pitch sensations (Sect. 2.7). A more detailed discussion of how this recognition process may actually work is given in Appendix II.

All psychoacoustic experiments reveal that our subjective reaction to complex tones depends appreciably upon the *context* of which they are a part. Performance of tasks that are musically "meaningful," such as recognition of melodies or harmonies and the identification of the tone source, that is, the instrument, greatly

[21] We strongly recommend that the reader review Sect. 2.9.

[22] It should be pointed out that what is shown in Fig. 2.25(b) are the mathematical predictions for a cochlear model which does not include a sharpening mechanism mediated by the outer hair cell motility (Sect. 3.6).

influence how complex tones are processed in the brain. This applies even to pitch perception. Experiments with electronically generated spectral components convincingly show that the attachment of a single pitch to complex sounds is greatly facilitated by, or sometimes even requires, the presentation of the test tones in the form of a meaningful melody. Individual electronically synthesized complex tones, taken out of a musical context, may often lead to ambiguous or multiple pitch sensations (see the example in Sect. 2.7, p. 54).

A startling experiment with "real" music to test this context-dependent fundamental pitch tracking effect can be performed on the organ. Play a piece (e.g., Bach's Orgelbüchlein chorale "Wenn wir in höchsten Nöten sein") with a single soprano melody on a cornet-type combination 8′ + 4′ + 2 2/3′ + 2′ + 1 3/5′ + 1 1/3′ + 1′, accompanied at all times with a soft 8′ + 4′ and 16′ + 8′, respectively. Ask a musically trained audience to carefully monitor the pitch of the melody, but warn them that there will be changes of timbre. After the first five to six bars, repeat the piece, but eliminate the 8′ from the melody. Repeat again, eliminating the 4′; then the 2′, finally the 1′. At the end, make the audience aware of what was left in the upper voice and point out that the pitch of the written note was absent altogether (in any of its octaves)—they will find it hard to believe! A repetition of the experiment, however, is likely to fail, because the audience will redirect their pitch-processing strategies!

When listening to a complex tone, our auditory system pays more attention to the output of the central pitch mechanism (which yields one unique pitch sensation) than to the primary pitch of the individual harmonic components. If we want to "hear out" the first six or seven upper harmonics of a steadily sounding complex tone, we must command a "turn-off" (inhibition) of the dominating subjective pitch mechanism and focus our attention on the initially choked output from the more primitive primary or spectral pitch mechanism, determined by the spatial position of the activated regions of the basilar membrane (see horizontal fibers in Appendix Fig. AII.2). This process of inhibiting and refocusing takes time—considerably longer than the buildup of the general tone-processing mechanism (Sect. 3.5). This is why upper harmonics cannot be "heard out" in short tones or rapidly decaying tones.[23]

It is important to point out that, in view of the asymmetry of activity distribution along the basilar membrane (e.g., Fig. 3.5) and the effect discussed briefly on p. 53, primary pitch matching of the overtones of a complex tone *always yields slightly stretched intervals*, for instance, between the first and the second har-

[23]The fact that the seventh harmonic is a dissonance has been worrying musicians for a long time. This worry is unfounded though: the seventh harmonic is extremely difficult to be singled out, even in constantly sounding tones from musical instruments. This is now recognized in the fact that some large modern organs do have a 1 1/7′ mutation stop sounding the seventh harmonic of the written note, which gives a very particular timbre when used judiciously with other stops, but does not disturb in any way the "smoothness" of the sound.

monic (stretched octave), and so on (Terhardt, 1971). This shift is caused by the perturbing influence of the ensemble of all other harmonics on the one harmonic whose primary pitch is being matched. The effect is small (up to a few percent), but may be musically relevant (Secs. 5.4 and 5.5 and Appendix II). An interesting effect occurs when just one of the harmonics of an electronically generated tone is mistuned: even if of higher order, it will suddenly be heard out as a separate entity. Not only heard out, but the matched pitch does not agree with the actual frequency difference—it is always exaggerated. It can be demonstrated that this effect provides a powerful argument in favor of the "place theory" of complex tone pitch perception (Lin and Hartmann, 1998). But it also has a practical application: it serves to demonstrate why all mutation stops of the organ must be painstakingly tuned to the *exact* intended frequency (just fifth, just major third, just seventh), and not, as it often is done, to the corresponding tempered intervals (Sect. 5.3)!

Of course, there are other key features of the primary auditory stimulus (ignored by the pitch processor) that yield perceptional output from other stages of the sound pattern recognition process. For a complex tone, we perceive *loudness* (linked to the total rate of neural impulses, Sect. 3.5) and tone quality or *timbre*. Here, we must make a clear distinction between the static situation that arises when we listen to a steady sounding complex tone of constant fundamental frequency, intensity, and spectrum, and the more realistic dynamic situation when a complex tone with transient characteristics is perceived in a musically relevant context. Let us analyze the static case. Psychoacoustic experiments with electronically generated steady complex tones, of equal pitch and loudness but different spectra and phase relationships among the harmonics, show that the timbre sensation is controlled primarily by the *power spectrum* (Sect. 4.3) (Plomp, 1970, 1976). Phase changes, although clearly perceptible, particularly when effected among the low frequency components, play only a secondary role.

The static sensation of quality or timbre thus emerges as the perceptual correlate of the activity distribution evoked along the basilar membrane—provided that the correct distance relationship among resonance peaks is present to bind everything into a "single-tone" sensation. By dividing the audible frequency range into bands of about one-third octave each (roughly corresponding to a critical band, Sect. 2.4), and by measuring the intensity or sound energy flow that for a given complex tone is contained in each band, it was possible to define quantitative "dissimilarity indices" for the (steady) sounds of various musical instruments, which correlate well with psychophysically determined timbre similarity and dissimilarity judgments (Plomp and Steeneken, 1971). It is important to point out that the timbre sensation is controlled by the *absolute* distribution of sound energy in fixed critical bands, not by the intensity values relative to that of the fundamental. This is easily verified by listening to a record or magnetic tape played at the wrong speed. This procedure leaves relative power spectra unchanged, merely shifting all frequencies up or down; yet a clear change in timbre of all instruments is perceived.

The static timbre sensation is a "multidimensional" psychological magnitude related not to one but to a whole set of physical parameters of the original acous-

tical stimulus—the set of intensities in all critical bands.[24] This is the main reason *semantic* descriptions of tone quality are more difficult to make than those of the "unidimensional" pitch (high-low) and loudness (loud-soft). Except for broad denominations ranging from dull or stuffy (few upper harmonics), to "nasal" (mainly odd harmonics), to bright or sharp (many enhanced upper harmonics), most of the qualifications given by musicians invoke a comparison with actual instrumental tones (flutey, stringy, reedy, brassy, organ-tone-like, etc.). This is similar to the description of psychophysical sensations of smell—consider the (sometimes rather contrived) descriptions of the "nose" of a good wine!

Timbre perception is just a first stage of the operation of *tone source recognition*—in music, the identification of the instrument. From this point of view, tone quality perception is the mechanism by means of which information is extracted from the auditory signal in such a way as to make it suitable for: (1) Storage in the memory with an adequate label of identification, and (2) Comparison with previously stored and identified information. The first operation involves learning or conditioning. A child who learns to recognize a given musical instrument is presented repeatedly with a melody played on that instrument and told: "This is a clarinet." His brain extracts suitable information from the succession of auditory stimuli, labels this information with the qualification "clarinet," and stores it in the memory. The second operation represents the conditioned response to a learned pattern: When the child hears a clarinet play after the learning experience, his brain compares the information extracted from the incoming signal (i.e., the timbre) with stored cues, and, if a successful match is found, conveys the response: "a clarinet." On the other hand, if we listen to a "new" sound, for example, a series of tones concocted with an electronic synthesizer, our information-extracting system will feed the cues into the matching mechanism, which will then try desperately to compare the input with previously stored information. If this matching process is unsuccessful, a new storage "file" will eventually be opened up for this new, now identified, sound quality. If the process is only partly successful, we react with such judgments as "almost like a clarinet" or "like a barking trombone." The neural processes responsible for all this will be discussed in the next section.

Musicians will protest and say that there is far more to the sensation of timbre than merely providing cues to find out "what is playing." For instance, what is it that makes one instrument sound more beautiful than another of the same kind? First, we should point out that this is obviously related to a further degree of sophistication of the identification mechanism mentioned above—we can learn

[24] Although there are about 15 critical bands in the musically relevant frequency range the intensities of which ought to be specified in order to determine the spectrum, a study of vowel identification (Klein et al., 1970) indicates that only *four* independent intensity parameters (each one a specific linear combination of the intensities in all critical bands) are sufficient to specify a complex tone within the "timbre resolution capability" of the auditory system.

to extract an increasingly refined amount of information from the sound vibration patterns of an instrument, so as to be able to distinguish among different samples of instruments of the same kind. Why some vibration patterns appear to be more beautiful than others is really not known. A great deal of research has been attempted, for instance, to find out what physical characteristics make a Stradivarius violin indeed a great instrument (e.g., Saunders, 1946). Many of these characteristics are dynamic in character, and most of them seem to be more related to the major or minor facility with which the *player* can control the wanted tone "color" (spectrum and transients), than to a "passive" effect upon a listener (a beginner will sound as bad on a Stradivarius as on any other violin!). For instance, a significant aspect of violin tones seems to be related to the effect of the narrowly spaced resonance peaks (e.g., Fig. 4.15) on loudness and timbre when the fundamental frequency of the tone is modulated by the player in a *vibrato* (Matthews and Kohut, 1973). Under such circumstances, the frequencies of the harmonic components sweep back and forth past the narrow and unequally spaced resonance peaks. As a result, the amplification of each component varies periodically, and so will loudness and timbre of the tone. Depending on the particular microstructure of the resonance curve of his instrument, the string player has the possibility of inducing extremely fine changes in loudness and timbre coupled to his vibrato. In the case of a famous baroque organ, the quality of its sound is determined as much by the spectrum and initial transients of its pipes as by the room acoustics—and the organist's dexterity in exploiting those characteristics in the phrasing of the tone passages.

## 4.9 Neural Processes Relevant to the Perception of Musical Tones

In Chap. 1, we stated that music is information. From a neurobiological point of view, the question of what information is involved in music becomes a question of identifying the specific neural patterns elicited by musical stimuli in the human brain, and finding out which affective behavioral consequences are innate and which are acquired through cultural conditioning.

Before we can address this matter from an objective, scientific point of view, we must clarify how information is actually represented in the brain (e.g., Roederer, 2005). There are two fundamental modes. One is dynamic, expressed in the form of a rapidly changing pattern of neural activity, specifically, the *spatial and temporal distribution of electrical impulses* which individual neurons send to other neurons (Sect. 2.8), representing the operating state of the neural network. The other mode is quasi-static, given by the *spatial distribution and efficacies* of inter-neuron connections (the synapses), representing the internal state or "hardware" of the neural network (also designated "synaptic architecture"). In addition, there is also a *chemical information transmission* system: certain substances (neurotransmitters, hormones) injected into the blood stream under neural

control play the role of a temporary modulator (stimulator or inhibitor) of the neural activity in specific brain regions; they determine the affective responses of brain activity and control many internal organ functions and the immune system.

The dynamic mode varies on a time scale of a few milliseconds to seconds and usually involves millions of neurons even for the simplest information-processing tasks, requiring a substantial supply of energy to be maintained. It is the increased vascular blood flow and oxygen consumption that appear mapped as images in functional magnetic resonance (fMRI) and positron emission tomography (PET), respectively (Moonen and Bandettieri, 1999; Herholz, 2004). Unfortunately, such techniques do not reveal the exact, neuron-by-neuron distribution of neural activity patterns; such a task seems hopeless, at least today: in the human brain, there are over 100 billion neurons in the cortex, each one connected to thousands of others. Yet, as anticipated in Section 2.8, it is the detailed microscopic *spatio-temporal distribution of electrical activity at the neuronal level* and the spatial distribution of synapses which taken together represent the integral state of the functioning brain at any instant of time.

Let us first discuss neural information processing in its broad term, expressed in the form of the brain region's enhanced neural activity detected in fMRI, PET, MEG, and EEG. Each region usually will encompass hundreds of thousands of neurons. First of all, let us identify what kind of external information needs to be recognized in the various sensory systems. There are well-defined stages of information processing in the neural circuitry of the brain, extending from the primary sensory receiving areas of the cortex to the prefrontal lobes, and from there to the motor areas that command the muscles. There is also a feedback system from the higher processing stages back to the primary areas. In the visual sense, it is the discrete entities with usually well-defined boundaries and textured surfaces which we call "the objects in space," their mutual spatial relationships and changes in time. The auditory sense recognizes "objects in acoustical space" (which in reality are "objects in time," like musical messages, Sect. 1.3), that is, discrete trains of sound waves with well-defined signatures, their relationships to each other and to their sources in the environment.

As mentioned in Section 2.9, the neural circuitry in the periphery and afferent pathways up to the primary cortical receiving area is mostly "prewired." Neurons in the primary cortical areas to which the afferent transmission system from a sensory organ is wired, are "feature detectors." If I look at a tree, there is no activated region on the cortex that has the form of a tree, and if I hear a trumpet sound, nothing blips in my cortex with a pattern that emulates the acoustic oscillation patterns of a trumpet. Rather, the afferent processing stages and the primary cortex have taken the original images apart, and remapped their component features in quite different ways and locations. For instance, in the primary auditory cortex (Heschl's gyrus), there are neurons that are "tuned" to a specific frequency interval, but they tend to respond only to certain complex sound stimuli in that frequency domain and specific transients (e.g., the famous "meow detectors" in cats). In the visual cortex, while individual neurons do have a specific receptive field

(i.e., a small region on the retina) for incident photons to which they respond, they only do so only for certain well-defined patterns appearing in that receptive field, such as a dark or light bar inclined with a specific angle, an edge moving in a certain direction, and so on (e.g., Marr, 1982).

At the next stage, these disjoint patterns have to be bound together in such a way that features belonging to one and the same object (spatial or temporal) elicit a pattern that is specific and univocal to *that* object—regardless of where in visual space or frequency space it is located. In other words, the incoming information has to be assigned into categories that have to do with *meaning*. In vision, responses to edges and lines belonging to the same object have to be transformed into *one* neural pattern that is in one-to-one correspondence with that object; in hearing, the spatially dispersed responses corresponding to the resonance regions of harmonics of a musical tone have to be transformed into *one* pattern specific to the pitch and timbre of a single musical tone (Secs. 2.9 and 4.8). In other words, the recognition of a given complex tone as "one entity" (regardless of its pitch, timbre, and intensity) is informatically equivalent to the recognition of a physical object in 3-D space as "one entity" (regardless of its color, size, or brightness).

Let us examine in more detail how the various levels or stages of visual information processing and representation operate in the brain (see sketch of Fig. 4.29 [25] (Roederer, 2005)). The neural circuitry in the periphery and afferent pathways up to and including the so-called primary sensory receiving area of the cortex (stage 1 in the occipital lobe) carry out some basic preprocessing operations mostly related to the above-mentioned *feature detection*. The next cortical stage 2a in the parietal lobe executes geometric transformations that assign "identity" to information coming from the same three-dimensional object seen at different distances, positions, and orientations. A second stream (stage 2b in the temporal lobe) carries out the above-mentioned *feature integration* or *binding* process, needed to sort out from an incredibly complex input those features that belong to one and the same spatial or temporal object. In other words, both operations transform radically different patterns (originating in the different projections on the retina from one and the same object) into single patterns that are in correspondence with the topological properties of the form of that object (for object recognition 2a), and with the spatial position of the object in the environment (for eventual motor actions, 2b). Some of these transformations may be learned, that is, the result of experience during the first months of an infant, with the sense of touch providing a signal of "uniqueness" to an object that is being handled and simultaneously looked at (it may be not by chance that region 2b is near the somatosensory area). They certainly can be learned at a later stage in life (for instance, one can easily learn to read upside-down or mirror-image text).

[25] Although this is a book on *sound*, it easier to visualize in a picture the information routes of the visual system. The principal lower-stage auditory areas of the cortex are hidden behind cortical folds.

Ablation studies with animals have shown that at stage 2, the brain "knows" that it is dealing with an object, but it does not yet know *what* the object is. This requires a complex process of comparison with existing, previously acquired information (stage 3) and must rely on the process of associative recall (see below). For instance, in the human medial temporal lobe, neurons were found that respond whenever faces of certain persons, environmental scenes, and specific objects and animals are seen (Marr, 1982). The final stage 4 in the frontal lobes corresponds to full recognition of objects and integration into the complete perceived scene (the landscape being seen). As one moves up along the stages of Fig. 4.29, the information-processing becomes less automatic and more and more centrally controlled; in particular, more *motivation-controlled* actions and decisions are necessary, and increasingly the *previously stored* (learned) information will influence the outcome (see Sect. 5.8).

The acoustic system has some equivalent processing stages from the informational point of view, except for the existence of a striking hemispheric lateralization, in which a division of tasks into sequential and synthetic operations

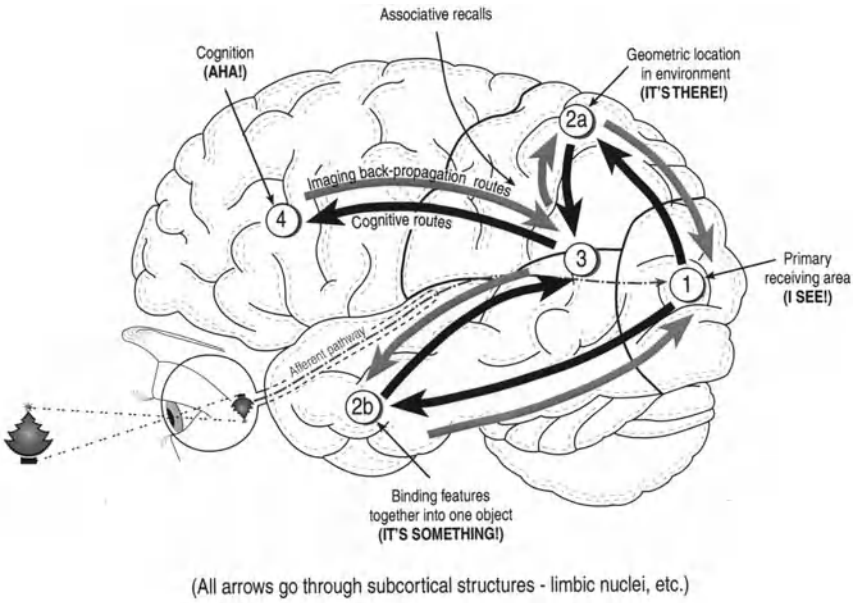



FIGURE 4.29 Ascending information routes and processing levels for visual information (from Roederer (2005)). The routing through lower, subcortical levels (not shown) checks on current subjective relevance of the information on its way to the prefrontal cortex. Through feedback pathways (gray arrows), the imagination of a given object triggers neural activity distributions at lower levels that would occur if the object was actually perceived by the eye.

takes place, described in more detail in Section 5.7. The evolutionary reason for this division is probably the brain's own version of "time is money": spoken language processing, generally accepted as the most distinguishing ability of human information processing and perhaps the most significant step in human evolution (see Sect. 5.6), puts enormous demands on the rate and speed of cerebral information processing. The brain simply cannot afford the ∼50 milliseconds it takes to exchange information between both hemispheres (mainly through the corpus callosum, see Sect. 2.9) when it comes to speech perception.[26] So, the fast sequential tasks are kept together in spatial proximity in one temporal lobe, which happens to be the left one in 97% of the individuals. In this speech hemisphere (also called the "dominant" hemisphere), there is a well-delineated three-step processing path (Binder, 1999) from the superior temporal gyrus (next to Heschl's gyrus, the primary auditory receiving area), to the phonemic pattern recognition systems around the superior temporal sulcus, and the first stage of lexical–semantic processing in the ventro-lateral (bottom) part of the temporal lobe. From there, the information proceeds to several "higher" areas, including the posterior cingulate (the cingulate gyrus is an important part of the cortex, buried deep in the middle groove, that interacts two-way with many other cortical and subcortical areas), the prefrontal cortex and angular gyrus, in a very complex and not yet fully explored series of steps for the full linguistic analysis, in which, like in the visual pathways, associative recall mechanisms play a fundamental role. The acoustic information processing in the minor hemisphere is more diffuse and less explored.[27] Basically, the equivalents to phoneme-lexical processing would be complex tone, chord and melodic analysis (for a more detailed discussion of sound lateralization, see Secs. 5.6 and 5.7).

Let us now return to the actual neuron-by-neuron representation of information in the brain, which, as mentioned above, defines the integral instantaneous state of the brain. How a specific spatio-temporal neural activity distribution elicited by listening to a sound or by the sight of an object becomes a specific "felt" sensation and *mental image* is an old question that has puzzled biologists and philosophers alike. Today, neurobiology provides a radical answer: the pattern doesn't "become" anything—it *is* the image! In other words, there is no need to postulate the existence of any scientifically indefinable, immaterial higher level instance such as a "mind" (although nobody should feel prevented from imagining such—see Sect. 5.8).

Let us restate this with an (oversimplified) example. When you see a "shiny red apple"; when you close your eyes and imagine a "shiny red apple"; when somebody says the words "shiny red apple"; or when you are reading these very lines, there appears a spatio-temporal distribution of neural activity in certain specific regions of your brain, part of which is nearly the same in all cases. That

[26] We have a similar situation with present-day electronic computers. The main limitation to their speed is given quite simply by the spatial distance between computing units!

[27] Research in music processing has a lower priority and receives less funding than speech!

common part represents the cognition of "shiny red apple," and is your mental image—your *neural correlate*—of the concept "shiny red apple." It is yours only; physically/physiologically, it would be very different from the one that forms in my brain or in anybody else's under the same circumstances (only the participating regions would be the same)—but still these patterns are all expressions of the *same pragmatic information* (the concept of "shiny red apple"). Also, the ulterior behavioral response to a pattern representing cognitive information may be the same for you and me (we both may feel pleasure, desire to eat it, etc.). What counts is the univocal character of the correspondence "object→neural activity distribution," not the actual form of the activity (which because of its enormous complexity could not be represented mathematically anyway (see however Rabinovich et al., 2008)).

There are no direct experimental proofs of this yet, but indirect evidence is overwhelming. For instance neurons have been found that respond consistently to one very special type of complex input like the face of a person, or to an expected input feature even if it is absent from the current stimulus (e.g., Koch, 2004; Tsao et al., 2006). More recently, so-called "mirror neurons" have been found that respond to specific feelings whether experienced by the laboratory animal itself or recognized by that animal to occur in another (empathy).

One fundamental brain function is *memory*. From what was discussed thus far, it should be clear that, as an information-processing system, the brain can be viewed as consisting of many discrete interacting modules, arranged in ascending levels, several of them in parallel) from periphery (sensory organs, brain stem, and primary areas) to the executive (frontal lobe) level, and from there to the motor output areas. But from the executive level, there are also many feedback connections back to the primary sensory areas (e.g., Fig. 4.29) and even out to the sensor organs (efferent system, Sect. 2.9). If an input stimulus-specific neural activity distribution in some level $L_n$ persists for, say, a few seconds after it has been formed, we say that this activity is a *short-term* or *working memory* image of the original stimulus—it represents temporarily stored information on the given stimulus. If, on the other hand, that same activity distribution at level $L_n$ is triggered at some *later* time by feedback from some higher level $L_{n+k}$, we say that the triggered image represents information on the original stimulus stored in *long-term* or *structural memory* at the higher level. That information would have to be of the quasi-static mode of synaptic architecture, mentioned at the beginning of this section. A very important region of the brain, the hippocampus (one in each hemisphere) is responsible for the long-term information storage operations (e.g., Whitlock et al., 2006). The feedback or "top-down" process, in which the original stimulus-specific neural activity distribution is being reconstructed *without* the corresponding full external sensory stimulus is called a *memory recall.* In other words, the memory recall of a sensory (or any other type) event consists of the *reenactment* of neural activity patterns that were present when that event was actually perceived. There also exists a process stretched in time. In the so-called *procedural memory* recall, a given motor skill is executed during an extended

period of time (e.g., driving a car while talking on a cell phone, riding a bike, etc.). In music practice and performance, it plays a fundamental role! The cerebellum plays a key role in the storage and recall operations of this type of "time release" memory.

The brain's information storage and retrieval processes are of a very different type than those with which we are familiar in daily life. To retrieve the photograph of my grandmother, I must know its "address" in the family album or on the wall, go there and access the image physically. To mentally recall or remember the image of the face of my grandmother, my brain must recreate at least part of the neural activity that was in one-to-one correspondence with that elicited by the actual perception of her face. This is called *distributed* or *content-addressable memory* and the mode of *holologic representation*. In other words, to retrieve a specific piece of information from everyday memory systems, we must know its address, or else scan through the entire storage register until we find what we want. But, if we had a *content-addressable* memory system of, say, recorded music (which does not yet exist in practice), we would be able to play or sing in the famous four notes ta-ta-ta-taah and retrieve the whole Fifth Symphony of Beethoven. This is precisely how our brain operates! What is stored in the brain is not the perception-triggered neural pattern itself, but the *capacity to regenerate* that pattern. Of course, there is no way yet to demonstrate that exactly the same complex spatio-temporal activity distribution is recreated every time an image is recalled, yet single cell recordings (and to a certain extent, functional MRI and PET tomography) convincingly show that the same cell *clusters* that had been involved in the corresponding perception become active again once that specific image is being recalled by association, voluntarily imagined or just only expected.

Let us show on the basis of an oversimplified scheme (Fig. 4.30, Roederer, 2005) how associative memory may work. Suppose that there are two nearly simultaneous input patterns $P_S$ and $Q_S$ (e.g., the visual image of a musical instrument and the acoustic pattern of its timbre, respectively), triggering the different patterns $P_A$ and $Q_A$, respectively, at some primary neural level $A$; $P_B$ and $Q_B$ at a higher, secondary level $B$, and so forth. We now assume that when these input patterns are repeated several times in near-simultaneity, changes occur in the synaptic architecture at level $B$ such that: (1) A *new* pattern $(PQ)_B$ emerges that is a representation of the simultaneously occurring stimulus *pair* $P$ and $Q$; (2) This new pattern appears even if either *only* $P$ or *only* $Q$ is presented as input; (3) Whenever pattern $(PQ)_B$ is independently evoked, *both* $P_A$ and $P_B$ will appear at the lower level $A$ due to feedback. As the result of this process, a sensory input $P_S$ will trigger an additional response $Q_A$ at level $A$ even *in the absence* of input $Q_S$, and an input $Q_S$ will trigger $P_A$ (as feedback from level $B$). This is the essence of what we called an *associative recall*. The formation of a new image of the pair $(PQ)_B$ at the secondary level $B$ is the result of a neural *learning process*; the change in hardware that made this possible represents the *long-term storage* of information on the correlated inputs $P$ and $Q$.

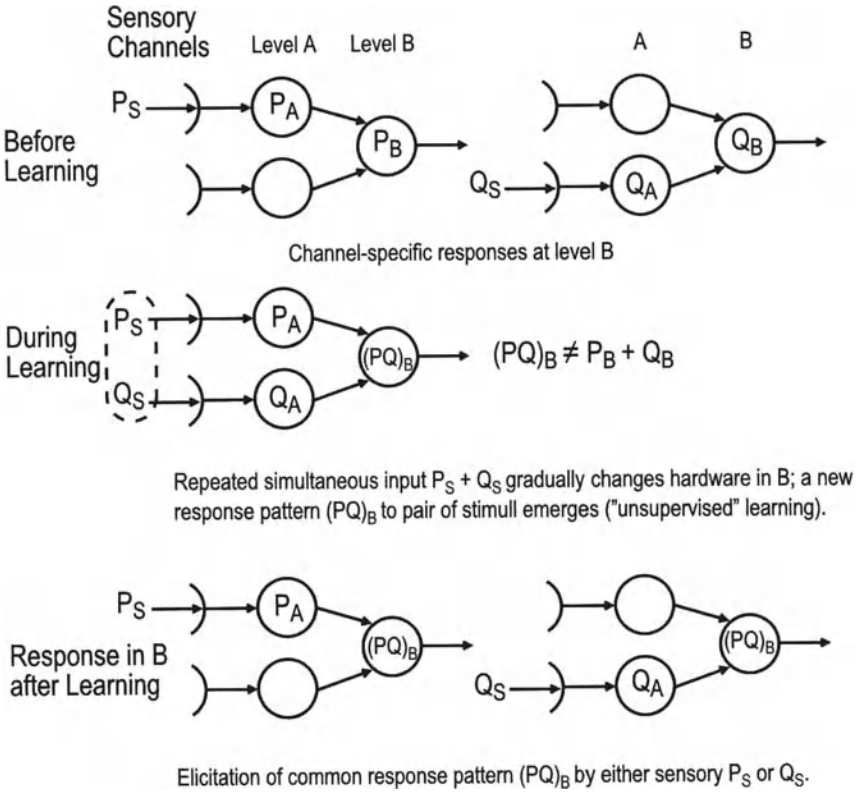



FIGURE 4.30 Sketch of the basic mechanism of learning and associative memory in the neural system (from Roederer (2005)). The change in "hardware" necessary to elicit the new, combined pattern $(PQ)_B$ consists of a change in the spatial configuration and efficacy of synapses (neural "plasticity"). This new spatial configuration embodies the representation of information stored in long-term memory.

$P_S$ and $P_Q$ need not be simultaneous—the new combined pattern can establish itself on the basis of the respective temporarily stored patterns (conditioning through short-term memory). Notice that $P_S$ and $Q_S$ could be inputs from two different sensory modalities as mentioned above; or they could be two different parts of the same object (e.g., the face of your dog and the body of your dog). One of the two inputs could be "very simple" compared to the other; in that case we call it a *key*; a very simple key as input can thus trigger the recall of a very complex image consisting of the superposition of different component patterns (e.g., a four-letter word, and the complex action it represents). This means that the replay of a specific neural pattern $P_A$ can be triggered by cues other than the full reenactment of the original sensory input $P_S$—a *partial* re-enactment may suffice to release the full image $P_A$ (this is an *auto-associative recall*). This also goes

for a partial input from the higher stages, and represents a fundamental property of the mechanism of associative memory recall (see also Appendix II). Finally, an incomplete or noisy signal as input (e.g., the missing fundamental effect; optical illusions; the "cocktail party effect") can trigger the recall of the "clean" original. These operations represent most basic algorithms for "intelligent information processing." They have been discussed extensively in the literature using numerical simulations with mathematical models of neural networks (e.g., Arbib, 1987; Kohonen, 1988).

Computer buffs will recognize that in brain operation, there is no software: memory, instructions, and operations are all based on appropriate changes in the hardware (the architecture and efficiency of synaptic connections between neurons). This self-organization is also the principle on which today's "neural computers" operate (although the "hardware changes" are still simulated with appropriate software programs—e.g., see Hinton (1992)). More on this and a simple model of a hololog ically operating central pitch processor is presented in Appendix II.

Let us turn to some specific musical examples. Consider a particular distribution of neural signals which is specific to an original sensory event such as the tone of a clarinet. When this pattern is triggered externally while we listen to the instrument, we recognize or remember that this tone comes, say, from a clarinet. When this activity is released internally (by some association or by a voluntary command—for the sense of vision, see the white arrows in Fig. 4.29), we are remembering the sound of a clarinet in absence of a true external sound. This then represents the simplest form of activation of the *acoustical imaging mechanism*. Experiments with vision have shown that, for instance, the mere imagination of a geometrical form evokes activity in the visual cortex very similar to that externally evoked when the subject actually sees that form (for details of this "top down control" see Miyashita (2004)). In the auditory system, "internal hearing" operates on the following basis: The imagination of a melody or the recall of spoken words is the result of the activation, or "replay," of neural activity originally triggered somewhere in the prefrontal cortical areas (depending on the specific recall process), which then feed information back down the line of the auditory processing stages creating sensory images without any sound whatsoever entering our ears. In animals, such backward transfer of information happens "automatically," triggered by associations or other environmental input; only in humans, it can be willed without an external input (see Sect. 5.7). As mentioned above, the experimental evidence does not yet prove that, in such a backward-propagating situation, the elicited spatio-temporal distribution of activity is actually identical to that which would occur for an equivalent external (sensory) input, but at least it shows that the participating areas are the same. When the experiments involve thinking about words (e.g., Premack, 2004), that is, semantic processing during absence of acoustic input or during silent reading or lip-reading (in which case the incoming information is visual but the tested areas are acoustic), fMRI imaging shows the active serial involvement of four regions - the angular gyrus, dorsal prefrontal cortex, posterior cingulate, and ventral temporal lobe, which is exactly the

reverse order followed when the subject actually *hears* the words (more on this in Sec. 5.6).

Another important example is the fact that the expectation or *anticipation* of a sensory input will trigger neural activity in relevant sensory and association areas of the cortex even *before* an actually occurring external feature can elicit the corresponding response; if the expected feature is missing, the corresponding neural pattern appears anyway. This has been verified in many experiments in which the response of feature-detecting neurons in the visual primary cortex is measured while the laboratory animal is exposed to stimuli in which the expected feature is sometimes present, sometimes missing. This effect plays a fundamental role in speech and music perception (e.g., Sect. 5.5).

Finally, although the rhythmic structure of music is not being addressed in this book, let us state that it may be strongly related to the natural clocks in the brain that control body functions and motor response, as well as the characteristic times of the working (short-term) memory. Neuroimaging studies have shown that the cerebellum and the basal ganglia (a neural network controlling stereotyped motion) may function as a central timing mechanism (see Peretz and Zatorre (2005), and references therein). An astounding feature in music is the ability of a performer or conductor to keep time in the long term (for instance, regarding the total duration of an hour-long musical piece) within an accuracy of seconds. There is indeed a mechanism in the brain responsible for the operation of a "stopwatch": it involves the so-called striatocortical loops connecting the basal ganglia (a region that coordinates automatic, stereotyped muscle movements) with the frontal cortex, and a neurochemical action by yet another nucleus, the substantia nigra, in which squirts of dopamine secreted by the neurons of the latter act upon an "accumulator" in the basal ganglia (the "hourglass" effect—see Morell (1996)), thus providing a longer-time (minutes, hours) time signal to the information-processing machinery of the brain.

# 5
# Superposition and Successions of Complex Tones and the Integral Perception of Music

*". . . it has become taboo for music theorists to ask why we like what we like. . . if different people have different preferences, we must not simply ignore the problem. Instead we must try to account for how and why that happens!"*

M. Minsky, in *Music, Mind and Brain,*
M. Clynes, ed., Plenum Press, 1982.

In the course of Chaps. 2, 3, and 4, we have been moving gradually up the ladder of neural processing of acoustic signals, from the mechanisms leading to the perception of spectral pitch, loudness, subjective pitch, and timbre, and to the recognition of a musical instrument. On the physical side, we have analyzed how the sound characteristics leading to these sensations are actually generated in musical instruments. These psychological attributes are necessary, but by no means sufficient, ingredients of music. Music is made of *successions* and *superpositions* of tones that convey integral information that is more than the sum of its parts, which can be analyzed, stored, and intercompared in the brain.

## 5.1 Superposition of Complex Tones

Polyphonic music consists of superpositions of complex tones. Even if only a single melody is played in monophonic music, a superposition of reverberant sounds usually reaches our ear, leading to the superposition of complex tones. The psychophysical study of complex tone superposition effects is still very much incomplete. This applies particularly to the understanding of how the brain is able to disentangle the "mess" of fundamental and upper harmonic frequencies that belong to different simultaneously sounding complex tones, so as to keep the sensations of these tones apart.

When two complex tones of different pitch are superposed, either of two situations may arise: The fundamental frequency of the higher tone is equal to one of the upper harmonics of the lower tone, or it is not. In the first case, the upper tone will reinforce certain upper harmonics of the lower tone. Why don't we, then, simply detect a change in *timbre* of the lower tone, instead of clearly singling out the upper tone and even keeping the timbre of both tones apart? A similar problem arises with the second case where each tone produces its own multiplicity of resonance regions on the basilar membrane. How does our brain single out from

J.G. Roederer, *The Physics and Psychophysics of Music*,
DOI: 10.1007/978-0-387-09474-8_5, 

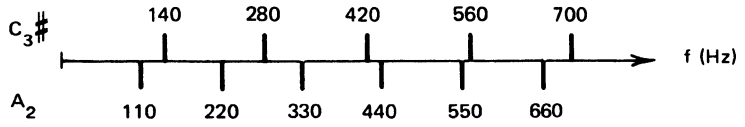



FIGURE 5.1 Harmonic frequencies of two complex tones forming a major third interval.

the resulting mixture which sequence belongs to what tone? For instance, consider the superposition of two complex tones, say, $A_2$ (110 Hz) and $C_3^{\#}$ (140 Hz). Harmonics of both tones are shown in Fig. 5.1 on a linear frequency scale. For each one of these frequencies there is a corresponding resonance region on the basilar membrane (Fig. 2.8). Unless there are slight changes in intensity or pitch (which happens in real music—see below), the sensor cells do not get the slightest cue as to which tone each resonance region belongs to. This discrimination must therefore be performed at a higher center in the auditory neural system.

The pitch of the two tones is discriminated by the central pitch processor (Sects. 2.9 and 4.8, and Appendix II). A most astounding capability of the auditory neural system is that of discriminating the *timbre* of two simultaneously sounding complex tones. No real music would be possible without such capability. For instance, assume that you listen monaurally with earphones to the sound of instrument 1 playing *exactly* the note $A_4$, and instrument 2 playing exactly $A_5$, at nearly the same intensity level. Fig. 5.2 shows the hypothetical superposition. The total length of the vertical bars represents the total intensity of each harmonic actually reaching the ear. How does our brain manage to keep both tones apart, in terms of timbre? This discrimination mechanism is not yet well understood; a *time element* seems to play the key role. First, the initiation (attack) of two "simultaneous" tones is never exactly synchronized, nor does the tone buildup develop in the same manner, particularly if both tones come from different directions (stereo effect). The onset of a tone is a most important attribute for timbre and tone identification (Iverson and Krumhansl, 1993). During this transient period, the processing mechanism in our brain seems able

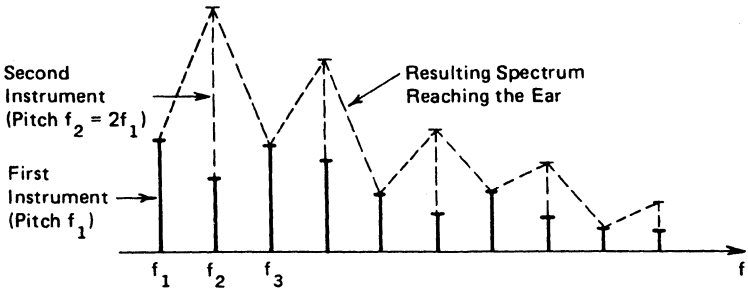



FIGURE 5.2 Resulting spectrum of two complex tones of different timbre (spectrum) an octave apart.

to lock in on certain characteristic features of each instrument's vibration pattern and to keep track of these features, even if they are garbled and blurred by the signal from the other instrument. During performance, too, slight excursions in frequency and intensity (the so-called *chorus* effect) that are *coherent* for the whole harmonic series of each tone are used by the auditory processing mechanism (Sect. 4.9). Other, more pronounced, variations that seem to provide important cues to the mechanism for discriminating tone quality are the periodic variations in pitch (vibrato) and intensity (tremolo, vibrato) that can be evoked (voluntarily or involuntarily) in otherwise steady tones of many musical instruments. Superposition of multiple complex sounds totally deprived of these small coherent time-dependent perturbations—as happens when multiple stops are combined in organ music—are indeed much more difficult to discriminate in timbre.

What probably aids the discrimination mechanism of both pitch and timbre of complex tones the most is the information received from a *progression* of tone superpositions. In such a case not only the above-mentioned primary cues given by the coherent fluctuations in timing and frequency of each tone can be used, but the contextual information extracted from the melodic lines (the musical "messages") played by each instrument (see also p. 154) will also be available.

The complex tone discrimination mechanism has its equivalent (and probably, its primordial root) in the mechanism that steers our auditory perception when we follow the speech of one given person among many different conversations conducted simultaneously at similar sound levels. This ability has been pointedly called the "cocktail party effect," and very likely uses the same cues, primary and secondary, as the complex tone discrimination mechanism. Finally, it probably is this same mechanism which enables us to disentangle individual sounds from among the messy tone superposition in a strongly reverberating hall (Sect. 4.7). In this latter case, again it is a time effect that seems to play the main role: The "first" arrival of the direct sound (Figs. 4.26 and 4.27) provides the key cues on which our perception system locks in to define the actual tone sensation and tone discrimination (precedence effect).

As happens with two pure tones, there is a minimum difference in fundamental frequency that two complex tones must have in order to be heard out separately (Fig. 2.13). When two complex tones differ in pitch less than the limit for tone discrimination, first-order beats (Sect. 2.4) may arise between all harmonics. If, for instance, both complex tones are out-of-tune unisons with fundamental frequencies $f_1$ and $f_1 + \varepsilon$, respectively, *all* resonance regions will overlap on the basilar membrane and produce beats of different frequencies. The fundamentals will beat with frequency $\varepsilon$, the second harmonics with frequency $2\varepsilon$, and so on. Only the first few harmonics are important; usually the beats of the fundamental ($\varepsilon$) are the most pronounced. First-order beats between corresponding harmonics will also appear if we superpose two complex tones forming other mistuned musical intervals. These are quite different (though equal in frequency) from the second-order beats that arise in out-of-tune intervals of *pure tones* (Sect. 2.6).

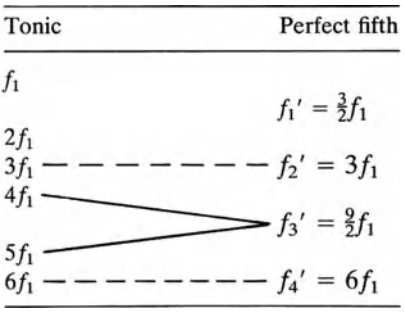
## 5.2 The Sensation of Musical Consonance and Dissonance

Consonance and dissonance are subjective feelings associated with two (or more) simultaneously sounding tones, of a nature much less well defined than the psychophysical variables pitch and loudness, and even timbre. Yet tonal music of *all* cultures seems to indicate that the human auditory system possesses a sense for certain special frequency intervals—the octave, fifth, fourth, etc. Even months-old infants pay special attention to these musical intervals as compared to any other ones (Trehub, 2001). It is most significant that these intervals are "valued" in nearly the same order as they appear in the harmonic series (see Fig. 2.19).

When two complex tones are sounded in unison or in an octave exactly in tune, *all* harmonics of the second tone will pair up exactly with harmonics of the first tone; no intermediate frequencies will be introduced by the second tone. This property puts the octave in a very special situation as a musical interval (in addition to the arguments concerning vibration pattern simplicity put forth in the analysis of pure tone superpositions in Sect. 2.6). The situation changes when we sound a perfect fifth of complex tones (Table 5.1). All odd harmonics of the fifth have frequencies that lie *between* harmonics of the tonic; only its even harmonics coincide. In particular, the third harmonic of the fifth, of frequency 9/2 $f_1$ lies "dangerously close" to the frequencies of the fourth and the fifth harmonic of the tonic: their resonance regions on the basilar membrane may overlap and either beats or "roughness" may thus appear (Sect. 2.4), even if the interval of fundamental frequencies is perfectly in tune. By constructing tables similar to Table 5.1, the reader may verify that for such other musical intervals as the fourth, thirds, and sixths, the proportion of "colliding" harmonics increases rapidly and moves down in harmonic order. Historically, this effect was thought to be the main cause for the sensations of consonance and dissonance.

Indeed, since the times of von Helmholtz, dissonance was associated with the number, intensity, and frequency of beating harmonics—and consonance with the

TABLE 5.1. Comparison of the first few harmonics of two complex tones a fifth apart. Solid lines: possible beating pairs.



absence thereof. In other words, it was assumed that for some unspecified reason our auditory system "does not like beats." As a result, it prefers, above all, the perfect unison and the perfect octave because, in these intervals, all harmonics of the upper tone form matching pairs with harmonics of the tonic. In the fifth, according to Table 5.1, the third harmonic of the upper tone may beat with the fourth and fifth harmonics of the tonic. The increasing proportion of beating pairs of harmonics that appear as one proceeds to the fourth, the sixths and thirds, the seventh, the second, etc., would thus explain the decreasing consonance—or increasing dissonance—of these intervals. This assumption was particularly attractive because—as it is easy to show mathematically—in order to maximize the number of matching harmonics of two complex tones (and hence minimize that of nonmatching ones), it is necessary that their fundamental frequencies $f_1$ and $f_1$' *be in the ratio of integer numbers,* and that these numbers *be as small as possible.* Indeed, if

$$\frac{f_1'}{f_1} = \frac{m}{n} \qquad \text{n, m: integers} \tag{5.1}$$

then the $m$th harmonic of $f_1$' will have the same frequency as the $n$th harmonic of $f_1$: $mf_1' = nf_1$ (and so will the $2m$th with the $2n$th, etc.). All other harmonics will not match and thus may give beats if their frequencies are near enough to each other. Table 5.2 shows the intervals within one octave that can be formed with small numbers *m, n* accepted in the Western musical culture (and most others) as consonances (in decreasing order of "perfection").

On the basis of sophisticated monaural and dichotic experiments (Plomp and Levelt, 1965) on consonance judgment involving pairs of pure tones and inharmonic complex tones, it became apparent that beats between harmonics may not be the major determining factor in the perception of consonance. Two *pure* tones an octave or less apart were presented to a number of musically naive (untrained) subjects who were supposed to give a qualification as to the "consonance" or "pleasantness" of the superposition. A *continuous* pattern was obtained that did

TABLE 5.2. Frequency ratios for musical intervals.

| Frequency ratio ($n/m$) | | Interval |
|---|---|---|
| "Perfect" consonances | 1/1 | Unison |
| | 2/1 | Octave |
| | 3/2 | Fifth |
| | 4/3 | Fourth |
| "Imperfect" consonances | 5/3 | Major sixth |
| | 5/4 | Major third |
| | 6/5 | Minor third |
| | 8/5 | Minor sixth |

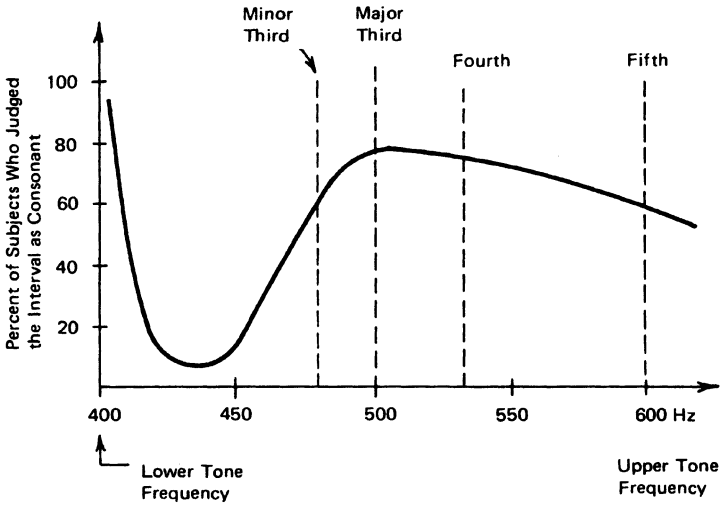



FIGURE 5.3 Consonance "index" for the superposition of two *pure* tones (after Plomp and Levelt, 1965).

not reveal preference for any particular musical interval. An example is shown in Fig. 5.3. Whenever the pure tones were less than about a minor third apart, they were judged "dissonant" (except for the unison); intervals equal or larger than a minor third were judged as more or less consonant, regardless of the actual frequency ratio.[1] The shape of the curve really depends on the absolute frequency of the fixed tone. All this is related to the roughness sensation of out-of-tune unisons and to the critical band discussed in Sect. 2.4. The results of these experiments may be summarized as follows: (1) When the frequencies of two pure tones fall *outside* the critical band, the corresponding pure tone interval is judged to be consonant. (2) When they coincide, they are judged as "perfectly" consonant. (3) When their frequencies differ by an amount ranging from about 5% to 50% of the corresponding critical bandwidth, they are judged as "non-consonant." We shall call an interval of two pure tones under these latter conditions a "basic dissonance."

We now turn back to the musically more significant case of two simultaneously sounding *complex* tones and apply the above results individually to each pair of neighboring upper harmonics. If the total number of pairs that are more or less consonant (see (1) above) and perfectly consonant (2) is weighted against that of basic dissonances (3), a "consonance index" may be obtained for each interval of complex tones (Plomp and Levelt, 1965; Kameoka and Kuriyagawa, 1969). It can be shown that this index indeed happens to attain peak values for tones whose

[1]Trained musicians were excluded from this experiment, because they would have been strongly compelled to identify consonances on the basis of training.

fundamental frequencies satisfy condition (5.1): the height of the peaks (degree of consonance) follows approximately the decreasing order given in Table 5.2. Moreover, in view of the dependence of the critical bandwidth on frequency (Fig. 2.13) *a given musical interval has a degree of consonance that varies along the frequency range.* In particular, moving toward lower frequencies, a given musical interval becomes less and less smooth sounding—a fact well known in polyphonic music, where in the bass register mainly octaves and, eventually, fifths are used.

The degree of consonance also depends upon the timbre or spectrum of the component tones, that is, the *relative intensity* of disturbing pairs of upper harmonics. This, too, is well known in music; there are instrument combinations that "blend" better than others in polyphonic music. Even the *order* in which two instruments define a musical interval is relevant. For instance, if a clarinet and a violin sound a major third with the clarinet playing the lower note, the first dissonant pair of harmonics will be the seventh harmonic of the clarinet with the sixth harmonic of the violin (because the lower even modes of the clarinet are greatly attenuated, Sects. 4.4 and 4.5). This interval sounds smooth. If, however, the clarinet is playing the upper tone, the third harmonic of the latter will collide with the fourth harmonic of the violin tone, and the interval will sound "harsh."

Terhard's (1974) theory of consonance perception postulates that tonal music is based essentially upon the pattern recognition mechanisms that operate in the auditory system (Sects. 2.9 and 4.8). One of these—the central pitch processor—responsible for the extraction of a single pitch sensation from the complex activity distribution elicited by a musical tone, acquires knowledge of the specific relations that exist between the resonance maxima and ensuing foci of neural activity evoked by the lower six to eight harmonics of such a tone (Sect. 4.8). The corresponding primary pitch intervals (octave, fifth, fourth, major third, minor third) thus become "familiar" to the central processor of the auditory system, and convey *tonal meanings* to all external stimuli whose (fundamental) frequencies bear such relationships (Appendix II).[2] According to this theory, both minimum roughness *and* tonal meaning play a determining role in the sensation of consonance. However, in view of the phenomenon of primary pitch shift of individually perceived harmonic components (p. 155), both of these principles may impart

[2]Quite generally, the hypothesis that the central pitch processor is a neural unit that must *learn* to extract meaningful information from complex input signals through repetitive exposure to natural sounds (Terhardt, 1972; Sect. 2.9), if demonstrated to be correct, could have far-reaching impact in many ways. In music, for instance, one may set out and try to *relearn* a whole new set of "invariant" characteristics pertaining to, say, a given class of inharmonic tones with the chance of *building entirely new tonal scales and schemes* thereupon (Terhardt, 1974). On the more practical side, this intrinsic learning capability would inject additional hope to the present-day efforts of developing electronic prostheses for the deaf, based on *microelectrode implants in the cochlea.* Whereas the spatial activation pattern of these implants is extremely difficult to predetermine, the interpretation of the elicited excitation patterns may well be *learned* by the patient's central processor.

conflicting instructions, and, in actual musical situations, force the central processor to a compromise (Terhardt, 1974). The preferred "stretched" tuning of pianos (as compared to the equally tempered scale, Sect. 5.3), and the observed fact that the upper note of a melodic interval of successive tones is preferentially intonated sharp (Sect. 5.4), may both be a result of this compromise.

There are more complicated factors that influence the sensation of consonance, most notably experience and training and the ensuing prejudice (i.e., musical tradition). It is interesting to note that, historically, musical intervals as explicit harmonic ingredients have been gradually "accepted" in Western civilization in an order close to the one given in Table 5.2. This seems to point to a gradual tolerance of our auditory processing ability. Of course, this was not the result of biological evolution but, rather, that of a sophistication of the *learning* experiences to which humans were being exposed as time went on. This development, as that of civilization as a whole, went on stepwise, in "quantum" jumps—it always took the mind of a revolutionary genius to introduce daring innovations the comprehension of which required new and more complex information-processing operations of the brain, and it was the charisma of the genius that was needed to persuade people to learn, and thus to accept *and preserve* these daring innovations. However, the human predilection for consonant intervals, linked to the fact that they are the intervals between the first harmonics of a complex tone, is universal and hence should be considered innate—determined by the mode of function of the central pitch processor, described in Sects. 2.9 and 4.8.

So far, we have been considering musical intervals smaller than, or equal to, the octave. For large intervals (e.g., $C_3$-$G_5$), it is customary to project the upper tone down by octaves ($G_5$-$G_4$-$G_3$) until an interval smaller than one octave is obtained ($C_3$-$G_3$). The degree of consonance of that latter interval is then considered "equivalent" to that of the original one. This cyclical property of intervals that repeats within successive octaves has been called the *chroma* of musical tones. It is a basic property that assigns an equivalent ranking to all tones whose pitch differs by one or more octaves, and which makes us call their notes by the same name—this is a property that is truly universal. What is responsible for this curious cyclic character of musical tones, which repeats every octave (every time the frequency doubles)? There is no equivalent in any of the other sensorial modalities.[3] It obviously is related to the key property of the almighty octave—that of having all its harmonics coincident with upper harmonics of the lower tone. There is no other

[3]In vision, "octaves" could never arise: the electromagnetic spectrum of visible light covers slightly less than one octave (frequencies from about 390–770 trillion Hz). Careful, however, with the temptation to consider color (frequency of light) equivalent to pitch (frequency of sound): whenever we superpose two or more pure (single-frequency) colors our eye perceives only one, but different, hue (think of how your color television works—with only three basic colors!). In other words, color superposition works very differently from tone superposition! It may sound strange, but the visual equivalent of auditory frequency (detected as position of the resonance region on the basilar membrane) is the *direction* of the incident light ray (detected as position of the stimulated region on the retina).

musical interval with this property (except the unison, of course). Quite generally, the existence of the chroma, that is, the fact that pitches differing by an octave have a degree of similarity that is considered identical to that of the unison, indicates that the pattern recognition process in our auditory system must respond in some "special," perhaps simplified, way when octaves are presented. Note again that the octave is the first interval in a harmonic series, and that the repetition rate of two notes an octave apart is *identical* to that of the lower tone. Any other consonant musical interval (fifth, fourth, etc.) has an associated fundamental repetition rate (relations (2.7)) that is *not* present in the original two-tone stimulus. If we remember how the pitch processor might work (Sect. 2.9 and Appendix II, Fig. AII.3), we realize that, whenever presented with two complex tones whose fundamental frequencies $f_1$ and $f_2$ are a musical interval apart, the output from the pitch processor should contain two prominent signals representing the pitch of each tone (corresponding to $f_1$ and $f_2$), *plus* other less prominent signals representing the repetition rate (2.7) corresponding to the pair of first harmonics $f_1$ and $f_2$ and its multiples (Appendix II). Under normal conditions, these additional signals are *discarded* as pitch sensations, a process that requires an additional "filtering" operation. Note, however, that this additional operation is not needed whenever an *octave* is presented, because no such third output signal is present! As a matter of fact, the "tonal meaning" mentioned above may be strongly related to the number, intensity, and position of "parasitical" signals in the output from the pitch processor. The more complex the multiplicity of these signals (i.e., the more complex the sound vibration pattern), the "lower" will be the tonal meaning of the original tone superposition. This is reflected already in the time-distribution of neural impulses in the acoustic nerve (Sect. 2.9 and Fig. 2.23), which for dissonant intervals exhibit coarse fluctuations ultimately leading to the perception of roughness and dissonance (Tramo et al., 2001). In other words, von Helmholtz' assumption that our auditory system "does not like beats or roughness between overlapping harmonics" turned out basically correct!

Finally, when *three or more* tones are sounded together it is customary to analyze the resulting chord into pairs of tones and to consider their individual consonance values. It is obvious that, as more and more complex tones are combined, a more complicated configuration of resonance regions arises on the basilar membrane. Overlap will increase to the extent that sensor cells covering a large extension of the basilar membrane will respond all at the same time. In light of the various pitch theories (Sects. 2.9 and 4.8, and Appendix II), we may state that in this case, too, the degree of consonance (or dissonance) may be related to both, the proportion of beating harmonics *and* the number, intensity, and position of parasitical signals (see Fig. AII.3) in the output from the pitch processor. Note that the *major triad* is a three-tone combination whose components, taken two at a time, always yield repetition rates that only differ from the tonic by octaves (i.e., have the same chroma).

There is a limit to multitone intelligibility, though. When the vibration patterns are randomized (i.e., their periodicity is destroyed), or when their complexity exceeds a certain threshold, the neural processing mechanism simply gives up:

no definite pitch and timbre sensations can be established. The ensuing sensation is called *noise.* Any nonperiodic pressure oscillation leads to a noise sensation. However, noise can be highly organized. Just as a periodic oscillation can be analyzed into a discrete superposition of pure harmonic oscillations of frequencies that are integer multiples of the fundamental frequency (Sect. 4.3), aperiodic vibrations can be analyzed as a *continuous* superposition of pure vibrations of *all* possible frequencies. Depending on how the intensity is distributed among all possible frequencies, we obtain different *noise spectra.* Noise plays a key role in the formation of consonants in speech. But it also plays a role in music; the importance of noise components in percussion instruments is obvious. The noise burst detected during the first tenths of a second in a piano and harpsichord tone has been shown to be a key element for the recognition process (Sect. 4.2, p. 124). The effect of noises of electronically controlled spectra on our auditory perception is being studied extensively with tone and noise synthesizers. A vast new territory in auditory sensations (music??) is being unveiled (see also Sect. 5.7).

## 5.3 Building Musical Scales

For a purely practical purpose, let us define a scale as a *discrete set of pitches arranged in such a way as to yield a maximum possible number of consonant combinations* (or minimum possible number of dissonances) *when two or more notes of the set are sounded together.* With this definition, and keeping in mind Table 5.2, it is possible to generate at once two scales in an unequivocal way, depending on whether all consonant intervals are to be taken into account, or whether only the perfect consonances are to be considered. In the first case, we obtain the *just scale;* in the second, the *Pythagorean scale.*[4]

### *1 The Just Scale*

We start with a tone of frequency $f_1$, which we call *do*. [5] The first most obvious thing to do is to introduce the octave above, which we denote as *do′*. This yields the most consonant interval of all. The next obvious thing is to add the fifth of frequency 3/2 $f_1$, which we call *sol.* That yields two new consonant intervals, besides the octave, of frequency ratios 3/2 $f_1$ (*do-sol*) and 4/3 $f_1$ (the fourth *sol-do′*), respectively. For the next step, there are two choices if we want to keep a maximum number of consonant intervals. They are the notes 5/4 $f_1$ or 6/5 $f_1$ which we call *mi* and $mi^b$, respectively. We choose the first one, *mi,* because this will guarantee a number of consonances of higher degree. Fig. 5.4 shows the resulting

[4]We envisage here a scale (also called temperament) as a set of tones with *mathematically defined frequency relationships.* This is to be distinguished from the various scale *modes,* defined by the particular *order* in which whole tones and semitones succeed each other.

[5]The solfeggio notation do-re-mi-fa-sol-la-ti-do' is used here to indicate *relative* position in a scale (i.e., the chroma), not actual pitch.

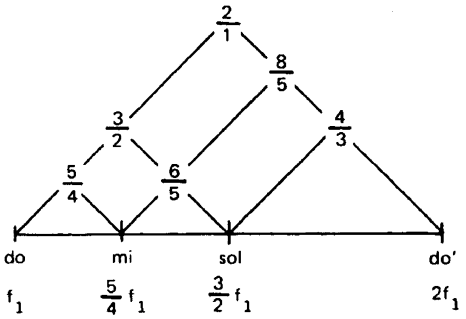



FIGURE 5.4 First set of consonant intervals obtained in the process of generating a just scale (see text).

intervals, all of them consonant. The notes do-mi-sol constitute the *major triad,* the building stone of Western music harmony (our second choice $6/5f_1$ or $mi^b$ would have yielded a *minor triad*).

We may continue to "fill in" tones, in each step trying to keep the number of dissonances to a minimum and the number of consonances (Table 5.2) to a maximum. We end up with the *just diatonic scale* of seven notes within the octave (Fig. 5.5). These seven notes can be projected in octaves up and down to form a full diatonic scale over the whole compass of audible pitch. Note in Fig. 5.5 the two intervals of quite similar frequency ratios 9/8 and 10/9 representing *whole tones.* The interval 16/15 defines a *semitone.* With the notes of this scale taken in pairs, we can form 16 consonant intervals, 10 dissonant intervals (minor and major sevenths, diminished fifth, whole tones, semitones), and—rather unfortunately—two *out-of-tune* consonances: the 1.5% too sharp minor third *re-fa* (32/27) and the 1.9% too flat fifth *re-la* (40/27). Finally, and perhaps most importantly, with the just diatonic scale, we can form three just major triads: *Do-mi-sol, do-fa-la*, and *re-sol-ti*; two just minor triads: *mi-sol-ti* and *do-mi-la*, and an out-of-tune minor triad re-fa-la.

Considering the existence of uneven spacings between neighboring notes; it is possible to still implement this scale by parting the larger gaps (whole tones) into two semitones each. Unfortunately, the resulting new intervals get more and more complex (e.g., several kinds of semitones, more out-of-tune consonances), the choices are not unique, and different frequency values result for the so-called

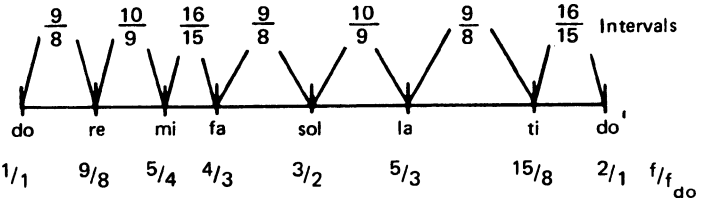



FIGURE 5.5 The just diatonic scale.

*enharmonic* equivalents $do^{\#}$–$re^{b}$, $re^{\#}$–$mi$, etc. Seeking to keep the proportion of possible consonances to a maximum, the following notes are introduced: $mi^{b}$ (6/5 $f_1$), $ti^{b}$ (9/5 $f_1$), $sol^{\#}$ (25/16 $f_1$) (or $la^{b}$, 8/5 $f_1$), $do^{\#}$ (25/24 $f_1$), and $fa^{\#}$ (45/32 $f_1$). The result is a *chromatic just scale* of 12 notes within the octave.

## 2 The Pythagorean Scale

We now restrict ourselves to the so-called perfect consonances, the just fifth and the just fourth (and the octave, of course), and build our scale on the basis of these intervals alone. We may proceed in the following way: After introducing *sol,* we move a just fifth down from $do'$ to introduce *fa* ($2/3 \times 2f_1 = 4/3f_1$). Then we move a just fourth down from *sol* to obtain *re* ($3/4 \times 3/2\, f_1 = 9/8\, f_1$), and a fifth up from the latter to get *la* ($3/2 \times 9/8f_1 = 27/16\, f_1$). Finally, we fill the remaining gaps by moving a fourth down from la to obtain *mi* ($3/4 \times 27/16\, f_1 = 81/64\, f_1$) and up a fifth from there, to *ti* ($3/2 \times 81/64\, f_1 = 243/128f_1$). The result is the so-called Pythagorean scale (Fig. 5.6). Notice that there is only *one* whole tone interval, the *Pythagorean whole tone* of frequency ratio 9/8 (equal to the "short" whole tone of the just scale). The interval 265/243 is the *Pythagorean diatonic semitone.*

We can convert this scale into a chromatic one, by continuing to jump up or down in just fourths and fifths. We thus obtain $fa^{\#}$ (a fourth below *ti*), $do^{\#}$ (a fourth below $fa^{\#}$), $sol^{\#}$ (a fifth above $do^{\#}$), $ti^{b}$ (a fourth above *fa*), and $mi^{b}$ (a fifth below $ti^{b}$). In this way, a new semitone appears (e.g., *fa*–$fa^{\#}$) defined by the odd-looking ratio 2187/2048, called the *Pythagorean chromatic semitone.* This whole procedure again leads to enharmonic equivalents of different frequency. In particular, if we continue to move up and down in steps of just fourths and fifths, we eventually will come back to our initial note *do—but* not exactly! In other words, we shall arrive at the enharmonic equivalent $ti^{\#}$ whose frequency is *not* equal to that of $do'$ ($= 2f_1$).

So, based on some "logical" principles, we have generated two scales. Each one has its own set of problems. By far the most serious one is the fact, common to both, that only a very limited group of tonalities can be played with these scales without running into trouble with out-of-tune consonances. In other words, *both*

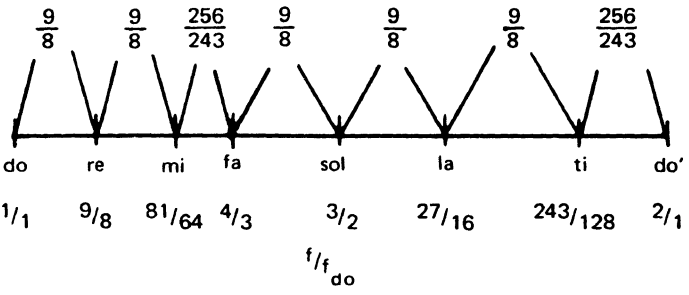



FIGURE 5.6 The Pythagorean diatonic scale.

*scales impose very serious transposition and modulation restrictions.* This was recognized as early as in the 17th century. There is no doubt, though, that both scales do reveal a quite specific character when music is performed on instruments tuned to either of them. But the type of music that can be played is extremely limited.

## *3 The Equally Tempered Scale*

It thus became apparent that a new scale was needed, which on the basis of a reasonable compromise, giving up some of the "justness" of musical intervals, would lead to *equally spaced intervals,* regardless of the particular tonality. In other words, a semitone would have the same frequency ratio, whether it was a *do–do*$^{\#}$, a *mi–fa*, or a *la–ti*$^{b}$, and a fifth would be the same whether it was *fa–do*$'$ or *do*$^{\#}$*–sol*$^{\#}$. This was accomplished in the *tempered scale,* enthusiastically sponsored by none other than J. S. Bach, who composed a collection of Preludes and Fugues ("Das Wohltemperierte Clavier") with the specific purpose of taking full advantage of the new frontiers opened up by unrestricted possibilities of tonality change on a keyboard instrument.

In the tempered scale, the frequency ratio is the same for all 12 semitones lying between do and do$'$. Let us call $s$ this ratio. This means that

$$f_{do\#} = s f_{do} \quad ; \quad f_{re} = s f_{do\#} = s^2 f_{do} \ \ldots \ f_{do'} = s^{12} f_{do}$$

Since we know that $f_{do'} = 2f_{do}$ (only the octave is kept as a "just" interval!), the twelfth power of $s$ must be equal to 2. Or

$$s = \sqrt[12]{2} \tag{5.2}$$

This is the frequency ratio for a *tempered semitone.* The frequencies that ensue for the notes of the chromatic tempered scale are *integer powers of s* times $f_{do}$. Table 5.3 shows the frequency ratios for consonant intervals in all three scales.

It is convenient to introduce a standard subdivision of the basic interval of the tempered scale, in order to be able to express numerically the small differences between intervals pertaining to different scales. This subdivision is used to describe small changes in frequency (vibrato), changes in intonation (pitch), and out-of-tuneness of notes or intervals. The most accepted procedure today is to divide the tempered semitone into 100 equal intervals, or, what is equivalent, to divide the octave into 1200 equal parts. Since what defines a musical interval is the *ratio* of the fundamental frequencies of the component tones (not their difference), we must divide the semitone frequency ratio $s$ (5.2) into 100 equal *factors* c:

$$\underbrace{c \times c \times c \times c \times \ldots\ldots \times c}_{\text{Hundred times}} = c^{100} = s$$

In view of relation (5.2) the value of $c$ is

TABLE 5.3. Frequency ratios and values in cents of musical intervals, for the three scales discussed in the text.

| | Just scale | | Pythagorean scale | | Tempered scale | |
|---|---|---|---|---|---|---|
| Interval | Ratio | Cents | Ratio | Cents | Ratio | Cents |
| Octave | 2.000 | 1200 | 2.000 | 1200 | 2.000 | 1200 |
| Fifth | 1.500 | 702 | 1.500 | 702 | 1.498 | 700 |
| Fourth | 1.333 | 498 | 1.333 | 498 | 1.335 | 500 |
| Major third | 1.250 | 386 | 1.265 | 408 | 1.260 | 400 |
| Minor third | 1.200 | 316 | 1.184 | 294 | 1.189 | 300 |
| Major sixth | 1.667 | 884 | 1.687 | 906 | 1.682 | 900 |
| Minor sixth | 1.600 | 814 | 1.580 | 792 | 1.587 | 800 |

$$c = \sqrt[100]{1.0595} = 1.000578 \tag{5.3}$$

The unit of this subdivision is called a *cent.* To find out how many cents are "contained" in a given interval of arbitrary frequency ratio *r,* we must determine how many times we must multiply $c$ with itself to obtain *r:*

$$c^n = r \tag{5.4}$$

$n$ is then the value of $r$ expressed in cents. By definition, one tempered semitone is 100 cents, a tempered whole tone ($s^2$) is 200 cents, a tempered fifth ($s^7$) is 700 cents, etc. To find the value in cents of any other interval, we must use logarithms. Taking into account the properties described in Sect. 3.4, we take logarithms of relation (5.4): $n \log c = \log r$. Hence

$$n = \frac{\log\ r}{\log\ c} = 3,986\ \log r \tag{5.5}$$

Using this relation, we find the values in cents for the various consonant intervals that have been given in Table 5.3.

## 5.4 The Standard Scale and the Standard of Pitch

The tempered scale has been in use for more than 200 years and has de facto become the standard scale to which all instruments with fixed-pitch notes are tuned. Since its inception, though, it has come under attack on several occasions—until this day. The main target of these attacks is the "unjustness" of the consonant intervals of the tempered scale, particularly the thirds and sixths (Table 5.3), which do, indeed, sound a little bit out-of-tune when listened to carefully and persistently, especially in the bass register.

Let us critically compare the scales discussed in the previous section with each other. There is no doubt that for *one given tonality* the just scale is the

"theoretically" perfect scale, yielding a maximum possibility of combinations of just or pure (i.e., beatless) intervals. For this reason, the just scale should indeed be taken as a sort of reference scale; this is precisely why we have introduced it in the first place. But the big question is: Does our auditory system really care for absolutely beatless intervals? And then: Would we give up tonality transposition and modulation possibilities in favor of obtaining these pure intervals? A 300-year history of music has answered these questions unmistakably with a loud and clear *no*! So the just scale is ruled out.

The Pythagorean scale may be one step forward in the right direction (while fifths and fourths are kept as just intervals, thirds and sixths are slightly out-of-tune, Table 5.3), but it still does not allow unlimited transposition and modulation possibilities. There have been other scales, introduced as minor modifications of the Pythagorean scale, that we will not even mention here, however. None has succeeded in preventing the equally tempered scale from being universally accepted.

There have been attempts to settle experimentally the question of which scale is really preferred (setting tonality modulation capability arguments aside). There are two possible approaches. (1) Use fixed-frequency instruments (piano, organ) and carefully compare the subjective impressions of a given piece of music played successively on two instruments of the same kind, respectively tuned to different scales. The piece of music, of course, should be very simple, without modulations into distant tonalities. And the instrument really should be one with nondecaying tones (such as the organ) in order to bring out beats or roughness more clearly. (2) The other possibility is to measure experimentally the average frequencies of pitch intonation chosen by singers or by performers of variable-pitch instruments (strings), and determine whether they prefer one temperament to another.

The second approach is more appropriate for yielding quantitative results. Electronic instrumentation has made possible very precise instantaneous frequency measurements on performers. The musical intervals to be watched closely are the major third and the major sixth, for which the differences between scales are most pronounced (Table 5.3). Notice, in particular, that the upper note in both of these intervals is flat in the just scale and sharp in the Pythagorean scale (with respect to the tempered scale). The experimental results very convincingly show that, on the average, singers and string players perform the upper notes of melodic intervals with *sharp* intonation (Ward, 1970). This seems to point to a preference for the Pythagorean scale. However, one should not jump to conclusions. The same experiments revealed that also fifths and fourths and even the almighty octave were played or sung sharp, on the average! [6] Rather than revealing a preference for a given scale (the Pythagorean), these experiments point to the existence of a previously unexpected *universal tendency to play or sing sharp the upper notes of all melodic intervals.* This stretched intonation could be caused by the primary pitch shift of the harmonic components of a musical tone (Sects. 4.3

[6] A reciprocal effect exists: just melodic intervals are consistently judged to sound flat (Terhardt and Zick, 1975).

and 5.2), which leaves a "slightly wrong" record in the central pitch processor (for a detailed recent discussion, see Hartmann (1993)). Furthermore, a perhaps even more significant result of these experiments is that individual fluctuations of the pitch of a given note during the course of a performance are very large. This includes vibrato as well as variations of the average pitch of a given note when it reappears throughout the same piece of music. In these pitch fluctuations of a given written note, a frequency range is scanned that goes far beyond the frequency differences between different scales—actually, it makes the latter completely irrelevant! Quite generally, all these results point to the fact that *musical intervals are perceived in a contextual, categorical mode,* with actual fluctuations being easily ignored. Ethnomusicologists often argue that all this (and many topics discussed in other chapters) only pertains to tonal *Western* music and that, in general, what we have called "music universals" are invariant only within the Western culture. They are mistaken. Leaving some tones out of a scale, deliberately starting a scale at a point different from the dominant tone, enlarging or modulating consonant intervals, are indeed culturally dependent "embellishments"—but fundamental *common* features of tonality and harmony such as pitch scales, consonance, and rhythmic organization can be recognized in *all* musics from the five continents! The present-day worldwide propagation of Western pop music and the fact that military music and national anthems everywhere follow the Western music style, are clear indications that the latter "resonates" with common natural inborn functions of the human auditory system (e.g., Tramo et al., 2001).

So far, we have been dealing with intervals, that is, frequency ratios. What about the absolute frequencies per se? Once a scale is adopted, it is sufficient to prescribe the frequency of only one note; it makes no difference which one. However, if musical instruments of fixed frequencies are to be easily interchangeable all over the world, this has to be done on the basis of an international agreement. One has prescribed for the "middle *A*" of the piano ($A_4$) a fundamental frequency of 440 Hz. Different "regional" standard frequencies had been in use since the tuning fork became available in the 17th century. Over the last three centuries, there has been a gradual rise of the "standard" frequency from about 415 Hz to as high as 461 Hz.[7] We can only hope that the present standard will, indeed, remain constant.

In the tempered scale, all intervals of the same kind (e.g., fifths, major thirds, etc.) are exactly "the same thing," except for the actual pitch of their components. A melody played in C major is in no way different from the same tune played in D major (except for the pitch range covered). Absolute "key colors" or different

[7] This has a serious consequence for famous historical instruments that are still in use today. For instance, a Stradivarius violin, originally built for a standard pitch of, say, $A_4 = 415$ Hz, today has to be tuned higher, which means *higher tension* for the strings (relation (4.3)). This alters the quality (spectrum) of the tone. A baroque organ, also built for $A_4 = 415$ Hz, when retuned to the higher pitch of $A_4 = 440$ Hz has to have its flue pipes partly cut, in order to shorten their effective length (relation (4.6)).

"moods" of certain tonalities have no psychoacoustic foundation, as experiments have shown long ago (Corso, 1957). There can be slight differences in the sounds of various keys, though, due to *physical* circumstances: The greater occurrence of black keys in the piano (which are struck in a slightly different way) for certain tonalities,[8] the greater occurrence of open strings for certain tonalities in string instruments, or the effect of the fixed-frequency resonances or range of formants (Sect. 4.3) in soundboards and other resonance bodies.

One final word is in order on *absolute pitch perception.* The few persons endowed with the ability to recognize or to vocally reproduce a given note in absolute manner (this is also called "perfect" pitch) are usually greatly admired. We have stated several times that the information most relevant to music is *relative* pitch changes, and that this is what our perceptional system is geared to pay attention to. In other words, our brain is set to interpret and store a melody as a sequence of pitch *transitions* rather than pitch values; the information on absolute pitch, although reaching our brain, is discarded as nonessential in the cognitive process. It can, however, be retained by all normal persons during short intervals of time, ranging from 10 s up to a few minutes (Rakowski, 1972). It is quite possible that "perfect pitch" could be learned at an early stage of mental education and retained thereafter. A recent study (Schellenberg and Trehub, 2008) finds no relation between absolute pitch memory and race or language (tone-language like Chinese).

## 5.5 Why Are There Musical Scales?

Our ear is sensitive to sound waves over a wide range of frequencies. We can detect very minor changes in frequency; the DL of frequency is typically only 0.5% or less (Fig. 2.9). Yet Western music (and that of most other cultures) is based on scales, that is, tone transitions and tone superpositions that differ from each other by more than 20 times the limen of our frequency resolution capability. Why don't we make music with continuously changing pitches that sound like, for instance, the "songs" of whales and dolphins which have a very sophisticated acoustic communications system based on continuous frequency "sweeps"? Why does pitch always have to "jump" in discrete steps?

There are no simple answers to these questions. First, let us remember that a given musical tone has to last a certain minimum period of time in order to be processed fully by the brain (Sect. 3.4). This probably has prevented sweeping tones from becoming basic and lasting elements of music. Second, let us note that different musical cultures are using or have used different scales—thus scales are somehow related to, or were influenced, by training and tradition. Third, most early musical instruments were fixed-pitched. The existence of scales has also

[8] For example, Chopin's piano music!

been justified on the basis of consonance; this would imply that scales appeared in connection with polyphonic music. However, scales had already been in use when melodies were only sung (or accompanied) monophonically in unison (or, at the most, in octaves or fifths), and they may have existed already in Paleolithic times (Fig. 1.1) (Gray et al., 2001). Perhaps, the underlying neuropsychological reason for the existence of scales is that it is easier for the brain to process, identify, and store in its memory a melody that is made up of a time sequence of discrete pitch values that bear a certain relationship to each other, somehow given by the "familiar" harmonic series, rather than pitch patterns that sweep continuously up and down over all possible frequencies, and whose processing, identification, and storage in the memory would require far more information than a discrete sequence.

The explanation of the existence of scales, that is, discrete tone sequences, has also been attempted on a *dynamic* basis of tone-tone relationships in time, that is, based upon *melodic* rather than harmonic intervals. This line of thought is based on the musically so important, but psychophysically still little-explored field of the sensations of "direction" or expectation of a two (or more) tone sequence, of dominance of a given pitch therein and of "return" to that leading pitch (also called "finality"). For instance, we tend to assign a natural direction to a two-tone sequence that is upward (in pitch) if the tones are a semitone apart, and downward, if they are a whole tone apart. In both cases, we assign a dominance to the second tone; the natural sense is then equivalent to the direction toward, the contranatural to the direction away from, the leading tone. Similarly, a sequence like *C-G-C-G-C-G*. . . "begs" to be ended on *C*, whereas the sequence *C-F-C-F-C-F*. . ."cries" for an *F* as the terminating note. And if we listen to *E-G-E-G*. . . neither of the components is satisfactory as an ending—we want to hear *C*! The whole diatonic orientation of music listening is based on these effects.

As a historical aside, at the beginning of last century, Meyer (1900) and Lipps (1905) attempted to "explain" the preference for certain melodic endings and tonic dominance in terms of numerical properties of the frequency ratios of a melodic interval. In the above examples, the dominating tone is that one whose frequency corresponds to a power of *two* in the integer number ratio. For instance, $f_G/f_C$ = 3/2 (lower tone leads); $f_F/f_C$ = 4/3 and $f_C/f_B$ = 16/15 (upper tone leads). Later investigations, however, have tended to attribute these effects mainly to cultural conditioning. Still, the question remains: Why did these and not any other preferences emerge? It is worthwhile to note, in this connection, that when the musical intervals *C-G, C-F,* and *E-G* of the above examples are thought of as neighboring tones of a harmonic series, the fundamental note of that series happens to give the leading pitch as determined by the sense of return (*C, F, C*, respectively). Again, this expectation may be dictated by the "familiarity" of the harmonic interrelations acquired by our central pitch processor (Sects. 2.9, 4.8, 5.2, and Appendix II) or, at a higher cognitive level, by the familiarity acquired through worldwide exposure to the Western musical culture (Bharucha, 1994). However, we must emphasize again that these perceptional effects have been observed in infants and may well be considered universal characteristics of music (Trehub, 2001).

Another perceptual phenomenon related to time sequences of tones, of importance to music, is that of *stream segregation* (Bregman and Campbell, 1971). If a melody is played in which the tones succeed each other fast with melodic intervals of several semitones alternating up and down, coherency is lost and *two* (or more) independent melodic lines are perceived. In this case, our brain tends to group the tones according to their *proximity in pitch*, rather than their contiguity in time. This effect has been profusely used, especially during the baroque period, to make it possible to play multipart music on a single-tone instrument. For comprehensive reviews of this and other related time-sequence phenomena, see van Noorden (1975), Deutsch (1982b, 1995), Sundberg (1992), and Bregman (1990).

## 5.6 Cognitive and Affective Brain Processes in Music Perception: Why Do We Respond Emotionally to Music?

Multidisciplinary research during the last three decades has led to notable progress in the understanding of the relationships between aspects of music common to all cultures and characteristic features of acoustic information processing in the human brain. As we anticipated in Sect. 1.7, increasing evidence of a parallelism between many structural aspects of music and human language points to a common, perhaps even simultaneous, origin of music and language during the early phase of human brain evolution. And robust arguments are emerging about the neural mechanism of musical emotions and the possible origin of the human drive to listen to music, make music, and compose music. In short, answers to the questions of *why did music develop in the early days of human evolution* and *why is there music still now* may be around the corner.

After a vigorous development of psychoacoustics which shed light on the perception of individual complex musical tones, tone superpositions and sequences, and fundamental psychological attributes such as pitch, loudness, timbre, consonance, roughness, chroma, tonal dominance, and scales (see preceding chapters and sections), the scientific interest shifted more and more to the underlying physiological and neural mechanisms, particularly to the higher-level processing involved in musical imaging (internal hearing, composition) and the affective response to music. This required a truly interdisciplinary approach, with a cooperative participation of musicians, physicists, psychologists, physiologists, and neuroscientists. The goals became rather demanding, addressing ultimate questions like: Why are humans from all cultures virtually “immersed” in something like music? Why do specific musical forms lead to different moods like happiness, sadness, courage, or fear? We enjoy gay music—so why do we also enjoy sad music? What was the survival value of music during the early history of human evolution, notwithstanding the fact that, apparently, music does not convey any “concrete” information like language? Why can music be used to treat mental illnesses and why does it affect the immune system?

If extraterrestrial intelligent civilizations exist, would they have music? These are all "transcultural" questions, addressing universal properties of music, and as such, can only be answered by finding out in detail *how* the human brain works at its highest level. This will be the topic of the remainder of this chapter.

We begin with a discussion of some basic concepts of life, information, and the evolution of human brain function (Roederer, 1978, 2005). Biological systems, from microorganism to primate, are "islands" of organized matter which persistently evolve toward increasing order. To be able to generate order, a living system must preserve itself invariant over some finite interval of time in spite of interactions with a changing environment. It must operate in a future-oriented, self-organizing way, that is, maintain itself functional by following courses of action that are favorably adapted to environmental and somatic change. In short, it must be able to effect changes consistently that are to its own advantage. The single most significant difference between living and nonliving systems is marked by the fact that the interactions of living systems with the environment and with each other, however simple or complex, are based on processes involving information (information-driven interactions, Sect. 1.6). Interactions between natural (not man-made) *inanimate* systems, like celestial bodies, rocks, atoms and elementary particles, etc., do not involve any information processing.

As species evolved, information about the environment was gradually incorporated and stored in the genetic memory structures of the organism. Very slow changes in the environment, of time-scales, orders of magnitude longer than that of one generation of the species, could be incorporated in the genome through mutation and survival of the fittest by Darwinian evolution. But, as the species became more complex and as the reaction to more and more unpredictable characteristics of the terrestrial environment became determinant of survival, the capability of ontogenetic adaptation during the lifetime of an organism became a fundamental requirement.

When locomotion appeared in multicellular organisms some half-billion years ago, the number of relevant environmental variables to be monitored increased drastically, with the time-scale of change down to a fraction of a second. It became necessary to absorb an enormous amount of information through increasingly sophisticated sensory system. Most of this influx of information is irrelevant but it carries embedded in an a priori unpredictable way those signals or patterns that are decisive for the organism's survival. The nervous system evolved to endow higher organisms with the capacity to detect, sort out, and identify relevant information contained in the complex sensory input, and anticipate and react appropriately to fast changes in the environment. To "react appropriately" also requires receiving and processing information on the organism itself, its overall metabolism and posture, position in and relation to the environment. In the course of this development, what started out as a simple environmental signal conversion, transmission, and muscle reaction apparatus in cnidarians like jelly fish and sea anemones, evolved into the central nervous system of higher vertebrates, with sophisticated input-analysis and response-planning capabilities. And, within the central

nervous system, the animal brain emerged as the "central processor" carrying out the fundamental operations of monitoring and control of somatic functions, environmental representation, prediction of environmental events, and the execution of a life-preserving and species-preserving behavioral response.

The main output function of an animal brain is the control of the organism's striate musculature for posture, voluntary and stereotyped movement, and control of some of the smooth muscles of internal organs and the chemical endocrine system. The more advanced an animal species, the more options it will have for the response to a momentary constellation of the environment. This requires decision making based on some priorities. Such instructions are coordinated by brain structures historically called the *limbic system*.[9] These structures include a group of subcortical nuclei located near the midline of the brain, which in conjunction with the hypothalamus, the amygdala, the hippocampus, and the basal forebrain, check on the state of the environment and the organism, ultimately directing the animal's attention and motivation (see sketch in Fig. 5.7), and making sure that the output—the integral *behavioral response*—is beneficial to the survival of the organism and the propagation of the species in conformity with evolutionary and ontogenetic experience (for a review see Dolan (2002)).

The limbic system works in a curious "binary" way by dispensing sensations of "reward or punishment": Hope or anxiety, boldness or fear, love or rage, satisfaction or disappointment, happiness or sadness, and so on. These are the *emotional states* of the brain (generated by the deeper subcortical nuclei). They evoke the anticipation of pleasure or pain whenever certain environmental events are expected to lead to something favorable or detrimental to the organism, respectively. Since this anticipation comes before any actual benefit or harm could arise, the emotional state helps guide the animal's motivation (controlled by the anterior cingulate cortex) to respond in a direction of maximum chance for survival and procreation, dictated by information acquired during evolution and experience—the so-called *instincts and drives*. Of course, only human beings can report in detail to each other on emotional states or feelings, but all higher vertebrates experience such a "digital" repertoire. The pleasantness or unpleasantness of a feeling is controlled by the chemical information system of the brain (opioids, monamines—Sect. 4.9).

Most behavioral responses of vertebrates are governed by this cortical-limbic interaction. As a matter of fact, in the case of Fig. 4.29 for example, there are routes (not shown) that communicate with the limbic nuclei to check on subjective relevance of the information being handled. Without the guiding mechanism of such a "control mechanism," animal intelligence could not have evolved; without instincts and drives the survival of a complex mobile organism in a rapidly changing, unpredictable environment would be highly improbable. Without the motivation to

[9]Brain scientists are quite reluctant to use such "umbrella" designations for processing centers that interact in different ways depending on context—but for our limited purposes it is reasonable to use it.

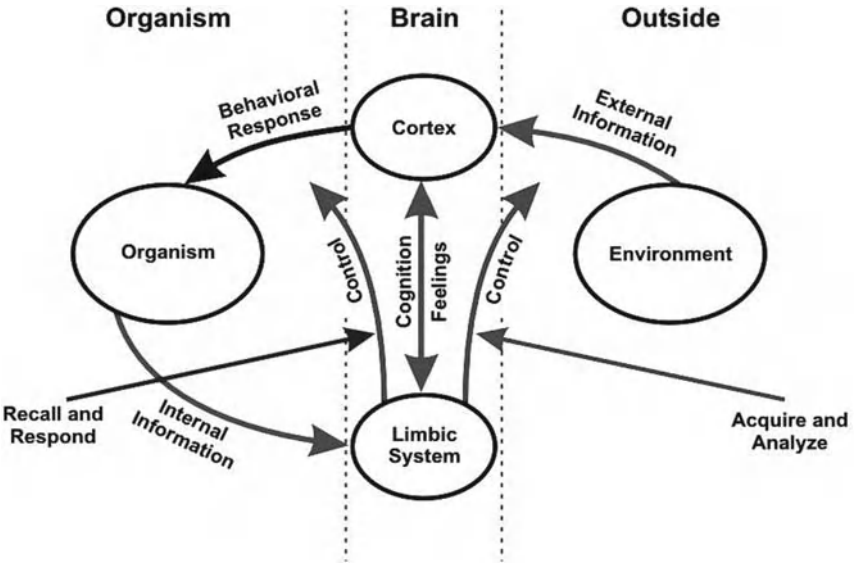



FIGURE 5.7 Basic functions of the limbic system which assure that the cognitive functions and behavioral response are beneficial for the organism and the propagation of the species. The limbic system controls emotion and feelings and communicates interactively with higher processing levels of the cortex. In higher species, the coherent operational mode of the cortico-limbic interplay gives rise to consciousness. The human brain is able to control this interplay and to overrule limbic dictates; this leads to self-consciousness (Roederer, 2005).

acquire information even if not needed at the moment, a repertoire of environmental events and appropriate responses thereto could not be built up in the memory. Without the coherent, cooperative mode of two distinct information-processing systems, a cognitive one (mainly handling ontogenetic, recent information) and an instinctive one (mainly working with phylogenetic past information), giving rise to the single "main program" we call *consciousness* (see for instance Damasio (1999) and Koch (2004)), animal intelligence would not be possible.

Even simple perceptional acts elicit responses from the limbic system. A musical example is the elementary sensation of pleasure when a consonant superposition of two complex periodic tones is heard—these are two simultaneous tones whose acoustic signal is easier to process because it has many coincident overtone frequencies (Sect. 5.2). Other examples are the reward for a confirmed short-term prediction, as in the resolution of a chord sequence or the goose flesh elicited by the challenge of a sudden, unexpected turn of a chord progression.

Here, we come to a fundamental question: Wouldn't all the preceding discussion mean that higher animals should experience these musical sensations, too? Experiments with cats, chinchillas, and chimps indeed show that they process primary attributes of complex tones pretty much like humans do—but do they *enjoy*

consonant intervals, do they anticipate endings of a tone sequence, do certain chord progressions ruffle their fur, do they instinctively move in synchrony with music? Before we turn to an answer, we must examine the fundamental differences between human and animal brains.

Aristotle already recognized that "animals have memory and are able of instruction, but no other animal except man can recall the past at will." More specifically, the most fundamentally distinct operation that the human, and only the human, brain can perform is to recall stored information as images or representations, manipulate them, and re-store modified or amended versions thereof *without any concurrent external sensory input* (Roederer, 1978, 2005). In other words, the human brain has internal control over its own feedback information flow (e.g., the gray arrows in Fig. 4.29); an animal can anticipate some event on a short-term basis (seconds), but only in the context of some real-time somatic and/or sensory input, that is, triggered by "automatic" associative recall processes. The act of information recall, alteration and re-storage *without* any external input represents the *human thinking process* or *reasoning*.

The evolution of a capability of recalling information without any concurrent input had vast consequences. In particular, the capability of re-examining, rearranging, and altering stored images led to the discovery of previously overlooked cause-and-effect relationships, to a quantitative concept of elapsed time and to the awareness of future time. In animals, the time interval within which causal correlations can be established (trace conditioning) is of the order of tens of seconds and decreases rapidly if other stimuli are present (Han et al., 2003); in humans, it extends over the long-term past and the long-term future. Along with the ability of ordering events in time came the possibility of *long-term prediction* and *planning*, i.e., the mental representation of events that have not yet occurred. Combined with the capacity of making decisions unrelated to real-time environmental and somatic input, these capabilities led to the emergence of *self-consciousness* (see discussion in Roederer (2005), and references therein).

In parallel developed the ability to encode complex mental images into simple acoustic signals and the emergence of *human language.* This was of such decisive importance for the development of human intelligence that certain parts of the auditory and motor cortices began to specialize in verbal image coding and decoding (next section), and the human thinking process began to be influenced and sometimes controlled by the language networks (e.g., Premack, 2004) (but it does not mean that we always think in words).

Concomitantly with this development came the postponement of behavioral goals and, more specifically, the capacity *to overrule the dictates of the limbic system* (e.g., sticking to a diet even when you are hungry) and also *to willfully stimulate the limbic system*, without external input (e.g., evoking pleasure by remembering a musical piece). In short, the body started serving the brain instead of the other way around. Mental images and emotional feelings can thus be created that have no relationship with momentary sensory input—the human brain indeed can go "off-line" (Bickerton, 1995). It is important to point out that the capabilities of recalling and rearranging stored information without external input, making

long-term predictions, planning and having the concept of future time, stimulating or overruling limbic drives, and developing language, most likely evolved hand-in-hand as neural expressions of human intelligence and self-consciousness.

Quite generally, the human thinking process involves the *creation of new images*, that is, spatio-temporal distributions of neural activity that do not correspond to any previously sensed or experienced information input. Patterns or objects can be crafted and changes in the environment can be effected that did not exist before and which would never be a deterministic consequence of physical laws and natural initial conditions. In particular, let us consider the case of *music imagery*, the "tune inside of your head." This neural process is now being actively explored scientifically with the new tomographic techniques (Sect. 4.9). Gradually one is reaching the conclusion, anticipated years ago, that in many regions of the brain, the neural activity involved in the processing of imagined sounds is nearly the same as that evoked by actually perceived sound (Halpern, 2001). In addition, there is evidence that in the process of imagining music, other nonacoustic brain centers are also activated systematically: The motor areas controlling hands (in instrument performers), larynx (in singers), arms (in conductors), and legs (rhythm), as well as the vision areas (imagery of the score, the instrument, the audience).

We may now dare to ask: How did Mozart compose? And do it so fast, so prolifically? There is evidence that he possessed the equivalent of an "eidetic" acoustic memory: apparently he could remember in detail any musical piece he once heard, but also those he only has imagined. As he willfully retrieved sound images and pieced them together in different, novel ways or created entirely new combinations of tones, he obviously was able to store immediately in memory everything that was being pieced together, while experiencing new emotional sensations triggered by these recently stored images—a true "reverberation in the composer's head," not of sound but of the correlated neural activity distributions representing the mental images of the sound. Of course, any composer must have such abilities—only that Mozart was quite unique in it.

## 5.7 Specialization of Speech and Music Processing in the Cerebral Hemispheres

In the introductory chapter (Sect. 1.6) and in Sect. 4.9, we alluded briefly to the remarkable division of tasks found among the left and right cerebral hemispheres of the human brain. We now expand the discussion of this phenomenon, mainly in the light of its relevance to music (e.g., see Bradshaw and Nettleton (1981); Peretz (2001a), and references therein).

The body of vertebrates exhibit a bilateral symmetry, especially with respect to the organs concerned with sensory and motor interaction with the environment. This symmetry extends to the brain hemispheres, with the left cortex connected to the right side of the body, and vice versa. This crossing mainly pertains to the systems capable of sensing directional dimension such as vision and audition

(e.g., see flow chart in Fig. 2.26), and to the efferent motor control of legs and arms. It probably developed because of the need to keep together within one cortical hemisphere the interaction mechanisms connecting incoming information and outgoing motor instructions regarding events from the same spatial half-field of the environment. The optical image is physically inverted in the eye lenses projecting the right panoramic field onto the left half of the retina and vice versa in each eye. The left halves of both retinas are connected to the left visual cortex, in order to reunite in one cerebral hemisphere the full information pertaining to the same spatial half-field.

As mentioned in Sect. 2.9, both hemispheres are connected with each other by the 200 million fibers of the corpus callosum (and the about one million fibers of the anterior commissure), which thus restores the global unity of environmental representation in the brain. In the afferent auditory pathways, there are lower-level connections between the channels from both sides (e.g., Fig. 2.26), through which the left and right side signals can interact to provide information on sound direction.

In the evolution of the human brain, the immense requirements of information processing that came with the development of verbal communication crystallized in the emergence of hemispheric specialization. In this division of tasks, the analytic and sequential functions of language became the target of the "dominant" hemisphere (on the left side in about 97% of the subjects, Penfield and Roberts, 1959). The minor hemisphere emerged as being more adapted for the perception of holistic, global, synthetic relations.[10] That the speech centers are located in one hemisphere has been known for over 150 years, mainly as the result of autopsy studies on deceased patients with speech and language defects (aphasias, alexias, anomias, agraphias) acquired after a vascular hemorrhage (stroke) in the left hemisphere (e.g., Geschwind, 1972). Right hemisphere lesions, on the other hand, were found to cause impaired visual pattern recognition (Kimura, 1963) and timbre and tonal memory loss (Milner, 1967). Quite generally, all nonverbal auditory tasks are impaired in these patients. Convincing examples have been documented in studies of "split-brain" patients whose corpus callosum had been transected for therapeutic reasons (e.g., Gazzaniga, 1970). For instance, these patients cannot verbally describe any object, written word, or event localized in their left visual field, because the pertinent sensory information, originally displayed on the right side visual cortex, cannot be transferred to the speech centers due to a severed corpus callosum. A technique used on patients without physical traumas in the brain is the injection of a barbiturate into a carotid artery, which briefly anesthetizes one hemisphere (a procedure sometimes used to confirm the left/right side location of the speech centers); a series of tests with such patients (see summary in Borchgrevink (1982)) confirmed that pitch and tonality in music (but not in speech)

[10]It might be suspected that speech processing is located in the hemisphere that also controls the predominantly used hand (e.g., left hemisphere for the right hand). But left-handedness is much more frequent than the 3% of right side speech processing.

are handled by the right hemisphere, whereas normal speech comprehension and production as well as musical rhythm were tasks of the left hemisphere. Right ear advantage was found in speech recognition tasks, and left ear advantage for melody tests[11] (e.g., Kimura, 1963). By far, the most convincing proof of this hemispheric division of tasks is provided by fMRI and PET tomography (see Peretz and Zatorre (2005)). In particular, neuronal networks in and close to the superior temporal gyrus in the right temporal lobe participate in music processing in a decisive and exclusive manner (Peretz, 2001b). Table 5.4 (based on the review by Bradshaw and Nettleton (1981), and references therein) summarizes some basic features of hemispheric specialization in auditory tasks.

Why did this curious dichotomy of hemispheric function appear in the course of human evolution?[12] The most plausible reason for this development, already mentioned in Sect. 4.9, was the need to keep the areas responsible for processing speech input and directing the vocal, gestural, and mimical output as close as possible to each other, in order to minimize transmission delays between the participating networks. The complex sequential operations of speech processing simply could not afford the time it takes (approximately 50 milliseconds) to transmit neural signals from one cerebral hemisphere to the other. As a result of this development, substantial "processing space" in the left hemisphere became unavailable for the other, slower tasks of holistic, integrative nature, which then "by default" were taken over by the right hemisphere. It is thus important to realize that the specialization of the cerebral hemispheres is of a much more basic nature, involving two quite different operational modes. One mode involves sequential analysis of subparts (subparts time-wise) of information such as required in language processing. The other involves spatial integration or synthesis of instantaneous

TABLE 5.4. Comparative listing of hemispheric specialization in auditory tasks (based on Bradshaw and Nettleton (1981)).

| Left hemisphere | Right hemisphere |
|---|---|
| Stop consonants | Steady vowels |
| Phonological attributes, syntax | Stereotyped attributes, rhyme in poetry |
| Comprehension of speech | Intonation of speech, environmental and animal sounds |
| Propositional speech | Emotional content of speech |
| Analysis of nonsensical speech sounds | Pitch, timbre, tonality, harmony |
| Spoken text (verbal content) | Sung text (musical and phonetic content) |
| Rhythm, short-term melodic sound sequences | Holistic melody |
| Verbal memory | Tonal memory |

[11] Although one auditory cortex also receives information from the contralateral ear (Fig. 2.26), tests show that the left auditory cortex will pay more attention to the right ear input.

[12] Our primate ancestors do not exhibit such a distinct hemispheric specialization (although this is still controversia1: some animals do exhibit an operational hemispheric asymmetry when it comes to sequential vs holistic processing tasks. See Denenberg (1981)).

patterns of neural activity, to accomplish the determination of holistic qualities of input stimuli (e.g., Papçun et al., 1974). However, both modes must coexist and cooperate in order to process information on, and program the organism's response to, the complex human environment.[13] In particular, sequential tasks (like visual scanning) may be necessary for pattern recognition and image construction and, conversely, holistic imaging may be required as an operation collateral to sequential programming.

Because music is preferentially handled by the minor hemisphere, does this mean that music mainly involves synthetic operations of holistic quality recognition? Regarding complex tone recognition, this indeed seems to be in agreement with the "template-fitting" theories of pitch perception (Sect. 4.8 and Appendix II). The holistic quantity in a musical stimulus is the current, instantaneous spatial distribution of neural activity (corresponding to the resonance maxima on the basilar membrane), leading to complex tone pitch (Sect. 4.8), to multiple-tone discrimination (Sect. 5.1), to consonance (Sect. 5.2), and tonal return and expectation (Sect. 5.5). Another quantity is the relative distribution of the amount of activity, given by the power spectrum, leading to timbre and tone source identification (Sect. 4.9). We may identify here a formal analogy with vision: the incoming sound pattern (in time) is "projected" as a pattern in space on the basilar membrane—the result is a spatial image, much like the spatial image projected on the retina. From there on, both systems operate on their respective inputs in formal analogy, eventually leading to musical and to pictorial sensations.

An apparent paradox emerges when we consider melodies and the time dependence of musical messages. Wouldn't they require sequencing, that is, dominant hemisphere operations? This is not necessarily so. Our brain recognizes the typical musical messages as being of holistic nature, long-term patterns in time, rather than short-term sequences. The phenomenon of melodic stream segregation (p. 185) is a most convincing example of this. Expressed in other words, music seems to be recognized by our brain as the representation of integral, holistic auditory images (the harmonic structure), whose (long-term) succession in time bears in itself a holistic "Gestalt" value (the melodic contours).

All this is quite germane to the understanding of the evolution of Western music since the Middle Ages. In a broad sense, we may depict this evolution as a gradual transition between two extreme configurations. At one extreme, we find highly structured, clearly defined, emphatically repeated, spatial (harmonic), and temporal (melodic) sound patterns, each one of which bears a value as an unanalyzed whole (e.g., a given chord and a given voice or chord progression, respectively). At the other extreme (contemporary music), we identify tonal forms whose fundamental value is recognized in the current state of the short-term temporal sound signatures. In the light of what we have said above about hemispheric specialization, we may speculate that these two extreme configurations are

[13] By "human environment" we mean an environment containing other humans with whom to communicate.

intimately related to the two distinct processing strategies of the human brain. Only the future will tell whether the current trends in music merely represent a more or less random effort to "just break away" from traditional forms (which in part had emerged quite naturally as the result of physical properties of the human auditory system), or whether these trends can be channeled into a premeditated exploration and exploitation of vast, still untested, processing capabilities of the central nervous system.

## 5.8 Why Is There Music?

We are now in a better condition to address the "ultimate" question: *Why* is there music? Let me state at the outset in quite general terms: Without a cortical-limbic interplay and without the capacity of internal information recall and image manipulation detached from current sensory input, there could be no music (nor any art and science). We can envision robots programmed (by a human being!) to compose and perform music according to preset rules, but it would be hard to imagine a robot *enjoying* to listen to and make music, *wanting* to compose, *enjoying* a painting, and *wanting* to find out how the Universe works!

It is not difficult to try to trace the origin of the motivation to perform certain actions that have no immediate biological purpose, such as climbing a mountain (instinct to explore), playing soccer (training in skilled movement), or enjoying the view of a sunset (expectation of the shelter of darkness). But why have "abstract" musical tones and forms been of advantage to early hominids? Of course, this question must be considered part of a more encompassing question related to the emergence of aesthetic motivation, response, and creativity.

As mentioned in Sect. 1.7, there is an increasing amount of evidence that music is a co-product of the *evolution of human language* (e.g., see Roederer (1984), Zatorre and Peretz (2001), and Koelsch (2005), and references therein). In this evolution, which undoubtedly *was* an essential factor in the development of hominids, a neural network emerged, capable of executing the ultra-complex operations of sound processing, analysis, storage, and retrieval necessary for phonetic recognition, voice identification, and analysis of syntax and grammar of speech. It is therefore conceivable that, with the evolution of human language, a drive emerged to train the acoustic sense in sophisticated sound pattern recognition as part of a *human instinct to acquire language* from the moment of birth. Animals do not possess the ability of propositional language, and they do not experience the specific motivations and drives that humans experience in relation to musical sounds[14]—this is why they hear consonances, but do not necessarily enjoy them, as we hinted at the end of the previous section.

[14]Bird song seems beautiful music to *humans*, but for the birds it is just their way of communicating very concrete messages concerning reproduction, feeding and danger. From the informational point of view, beautiful birdsong is no different from the awful-sounding shrieks of primates. . ..

During the later stages of intrauterine development, the acoustic sense of the fetus begins to register passively the intrauterine sound environment. At birth, there is a sudden transition to active behavioral response in which the acoustical communication with the mother or her surrogate plays a most fundamental role. An acoustical communication feedback cycle is thereby established, which may reinforce the emotional relationship with the mother and feed both the motivational drive to acquire language in the infant and the motivational drive of the mother to vocalize simple successions of musical tones (Roederer, 1984). Recent experiments with few-months old infants reveal a remarkable inborn predisposition of the brain for musical message processing (see review in Trehub (2001)), well before such messages could have any biological or social utility. For instance, infants recognize when a melody is shifted in pitch up- or downward provided the tone relationships are preserved; they recognize when the tempo is altered as long as relative durations are preserved; and they detect interval changes in the context of integer frequency ratios—the octave, fifth, fourth, etc.

Note that an infant reacts first to the *musical* content of speech, but so do dogs; when adults speak to infants or dogs or any pets, they use the same intonation and pitch contour in the tone of voice, quite independent of the actual language being used. But here the similarities end. The motivation to listen to, analyze, store, and vocalize musical sounds, even when there is no apparent need given by present circumstances, leads to limbic rewards, that is, triggers feelings of pleasure when this is done (Blood and Zatorre, 2001). When we sing to a puppy, nothing special will happen, except eliciting attention to the sound source. To facilitate the acoustical information processing of speech, the motivation emerged to discover symmetries and regularities, to extrapolate, predict, interpolate, to tackle with redundancy and repetition and with the surprise of sudden change. Each one of these tasks elicits affective responses, which taken together contribute to the emotional effects of music, ranging from those of instantaneous character related to the subjective sensations of timbre, consonance, tonal expectation, sense of tonal return, to the longer-term structures of melodic lines. These affective elements may be manifestations of limbic rewards in the search for the phonetic or phonemic content of sound and for the identification of grammatical organization and logical content of acoustical signals. They represent a predisposition for musical skills; the fortunate fact that these feelings are irrepressible and occur *every time* lies at the very foundation of modern music theory (e.g., Lerdahl and Jackendoff, 1983). The timing aspects involved in this kind of acoustical information processing may engage the "clockwork" circuits mentioned at the end of Sect. 4.9, and trigger limbic rewards in association with musical rhythm.

The evolution of music exhibits two stages, both historically as well as ontogenetically in each individual. First, there is a stage driven by genetic factors that has co-evolved with the appearance of language during the early days of the human species. Second, there is a "non-adaptive pleasure-seeking" stage, driven by the feelings evoked by structures and rhythms of tone superpositions and sequences (Huron, 2001). It is the latter which plays the primordial role in today's music enjoyment.

Since an early stage in life, most persons are exposed to a limited class of musical stimuli. Cultural conditioning rapidly takes hold, and emotional response

begins to be influenced by external factors, some fortuitous and subjective, like the emotional state experienced by a person during the first listening of a given musical piece or passage therein; some more controllable, such as the degree of repetition of characteristic musical forms pertaining to a given musical style. In addition, the innate drive to diversify the possibilities of human endeavor plays an important role. Technological developments such as the appearance of keyboard instruments or, more recently, electronic synthesizers have had substantial impacts on development and on why one particular style or type of music is preferred over some other kind.

A question arises about *unmusical* individuals, who are unable to experience most musical sensations. Although they *hear* everything that musical subjects hear, their central auditory system does not have the skills to extract musically relevant information from nonspeech-related sound superpositions and sequences. Musical events such as a tonality or a rhythmic change are heard but not interpreted and thus do not evoke any affective response. This impairment can range from a severe *amusia* (congenital or caused by trauma to the musical information-processing areas of the minor hemisphere (Sect. 5.7)) to the more frequent inability to carry a tune. The fact that there are unmusical but otherwise perfectly normal adults, whereas there are no "otherwise normal" adults that are unable to process language, is sometimes used to dismiss any causal link between music and language (e.g., Pinker, 1994). This argument, however, ignores the historical and cross-cultural importance of music (Trehub, 2001) and the evolution of cortical areas specialized in musical information processing (Sect. 5.7).

Concerning the development of the second "pleasure-seeking" stage of music, we may search for further contributing elements to a survival value. Like a good public speech, music can succeed in arousing and maintaining the attention of masses of people, overruling their normal limbic drives for extended periods of time. Since music conveys information on affective states, it can contribute to the equalization of the emotional states of a group of listeners just as an oral lecture may contribute to the equalization of the intellectual state (knowledge) of the audience. The role of music in superstitious and sexual rites, religion, ideological proselytism, military arousal, even antisocial behavior, clearly demonstrates the value of music as a means of achieving behavioral coherence in masses of people. In the distant past, this could indeed have had an important survival value, as the increasingly complex human environment demanded *coherent, collective actions* on the part of large groups of human society (Benzon, 2001). A fundamental aspect of this is the role of rhythm; indeed, recent studies point out the importance of the relationship between the biological rhythms of the human body (see end of Sect. 4.9) and relevant rhythms of music. When music is perceived, the biological system of the listener reacts to the signals as a whole, and a number of physiological effects can be observed and analyzed. According to chronobiology, a musical experience can be a pleasant one or a disturbing one depending on the synchronization between the rhythms of the organism and the acoustic input, especially when the experience of the music begins.

In the end, what does remain invariant from the original instincts, independent of the exposure to a given musical culture, are (1) The *fact* that there are some components of music that are common to all musical cultures; (2) The *fact* that motivation exists to pay attention to musical sounds and forms; and (3) The *fact* that an emotional reaction and feelings can be triggered.

## *1 Epilogue: Is Music a Universal Signature of Human-like Intelligence?*

Let us conclude this book with some thoughts about an "extraterrestrial" issue (after all, I am a space physicist!). Music is an art form that exploits the information processing capabilities of our sense of hearing and the elicited feelings of pleasure, just as the visual arts are based on vision and culinary art is related to the sense of smell. Whereas pictorial and culinary arts are mainly "static" or at most "slow-moving," the most relevant information in music comes in the form of rapid *time sequences* of external signals. As emphasized repeatedly in this book, information processing in the auditory system consists of an analysis of temporal *change* that involves a wide range of characteristic time scales. These are also the relevant time scales for speech and the neural information processing of language—the perhaps most distinct ability of the human being (Sect. 5.6). It is not surprising therefore, that, as we posited in the preceding section, music co-evolved with human language as a "training tool" for information-processing operations necessary for the development of oral communication.

The crux of the matter concerning the question formulated in the title resides in some statements we made on p. 187 of Sect. 5.6, which we paraphrase:

> "...there are routes (from the cortex) that communicate with the limbic nuclei to check on subjective relevance of the information being handled [see Figs. 4.29 and 5.7]. Without the guiding mechanism of such a "control mechanism," animal intelligence could not have evolved.... Without the motivation to acquire information even if not needed at the moment, a repertoire of environmental events and appropriate responses thereto could not be built up in the memory. Without the coherent, cooperative mode of two distinct information-processing systems, a cognitive one (handling mainly ontogenetic, recent information, by the cortex) and an instinctive one (mainly working with phylogenetic past information, controlled by the limbic system), giving rise to the single "main program" we call *consciousness*, animal intelligence would not be possible."

To this we now add: Human intelligence could not have evolved *without*: (1) A cortical-limbic interplay and the capacity of internal information recall and image manipulation detached from current sensory input (Sect. 5.6); (2) The acoustic system playing the central role for the conveyance of *intra-species information exchange* (i.e., a language, encoding in highly compressed form complex sensory and mental images and their relationships); (3) A brain-like central processor of information with *motivational control* of its cognitive functions, so that noninheritable components of language can be acquired efficiently during the early lifetime

of the organism; and (4) A control mechanism that dispenses reward/punishment-like *feelings* if the action taken is deemed favorable/detrimental according to evolutionary experience. We said human intelligence. But we can expand the realm to *human-like* intelligence—applicable elsewhere in the Universe.

Indeed, I believe that any extraterrestrial civilization with human-like intelligence *must* have evolved an information-processing system, similar to our central nervous system, which follows the above specifications. And I also believe that a temporally expressed "language-system related art" would then *necessarily* have developed in that extraterrestrial society. In other words, like on Earth, human-like language and the equivalent of a human-like music would have co-evolved—the latter as an evolutionary relic of a training tool for the acquisition of language.[15]

Will we ever find out if there is music out there in the Universe, besides our own? What would a program like NASA's Search for Extraterrestrial Intelligence (SETI) have to look for in order to find out? It would be naïve to listen with our radio telescopes for signals with a structure and organization that resemble, say, the tango La Cumparsita or Beethoven's Fifth. Rather, once signals from a civilization of intelligent beings have been identified, the key task would be to somehow find out if, among the messages with which those beings communicate *with each other*, a substantial fraction appears to have no relation to their language code and no specific purpose for, or relation to their immediate needs for somatic well-being and survival (Roederer, 2009). Such "purposeless" messages, I bet, would encode *their* music. . .

[15]It could be argued that, in principle, a language communications system in another civilization would not necessarily have to be acoustic—for instance, it could be optical (some ultra-sophisticated firefly-like emission organ, or some very advanced gesticulation capability). The key should be the capacity of acquiring, encoding in highly compressed form, processing at high (neural) speed, storing efficiently, and sending out very complex information. However, in a gaseous or liquid environment (which we *must* assume for living beings), an acoustical intra-species communications system is far more advantageous than the line-of-sight optical mode, because of the possibilities of diffraction around obstacles, very low absorption in the medium, propagation through fog and along natural wave guides (caves), etc. Besides, *if there were* an optical language system, as explained in a footnote on page 174 an "optical system of music" would necessarily be lacking many of the fundamental attributes of our music (consonance, chroma, harmony, etc.) because of the extremely limited range of frequencies of electromagnetic radiation (barely an octave) that can be handled by any conceivable physiological system built with carbon-based biochemistry.

# Appendix I
## Some Quantitative Aspects of the Bowing Mechanism

Let us consider an idealized situation: A very, very long string, bowed at a point $A$ with an infinitesimally thin bow Fig. AI.1. The bow moves with speed $b$ in the upward direction. Let us furthermore assume that right from the beginning (time $t_0$) the string sticks to the bow. This means that the point of contact $A$ (we really should say the segment of contact) moves upward with the same speed $b$. The result will be a deformation of the string in the form of a transverse wave that will propagate away from point $A$ as shown in Fig. AI.1 for instants of time $t_1$, $t_2$, $t_3$. Since transverse waves propagate with a speed $V$ given by relation (3.3), which happens to be much higher than any reasonable bowing speed $b$, the slope $b/V$ of the kinked portions of the string $AP$, $AQ$ in reality will be extremely small. Under these conditions, the transverse force $F$ applied by the bow (not to be confused with the bowing pressure which would be directed *perpendicular* to the paper) holds the balance with the projections along $OA$ of both tension forces $T$. This means that $F = 2\,Tb/V$. In order to actually have a regime of static friction, with the string sticking to the bow, the force $F$ must be less than a certain threshold $F_s$, called the limit of static friction. Experimental results show that this limit is proportional to the bowing "pressure": $F_s = \mu_s P$. The parameter $\mu_s$ (Greek letter mu) is the *coefficient of static friction;* it depends on the "roughness" of

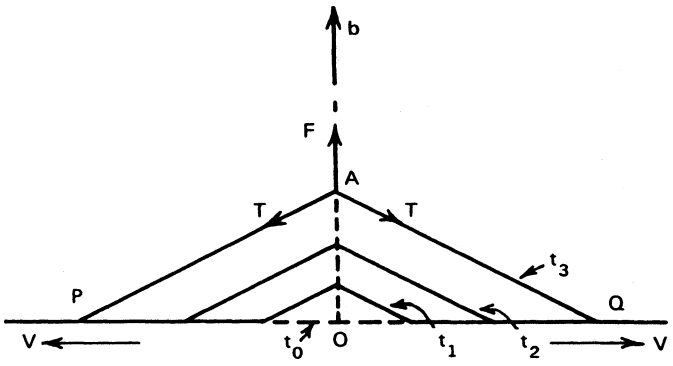



FIGURE AI.1 Idealized progression of the deformation of a long string bowed at point A with constant velocity $b$ (not in scale!).

the contact surfaces (in this case, on the amount of rosin on the bow′s hairs). The condition for "sticking" is then $F = 2\,Tb/V < \mu_s P$. [1]

Likewise, the condition for slipping will be $F = 2\,Tb/V > \mu_s P$. Since the quantities *V, T,* and $\mu_s$ are constant parameters for a given string, we may summarize both expressions in the following, more physical way:

| Quatity controllable by the player | Quantity fixed for each string | Type of string motion with respect to the bow | |
|---|---|---|---|
| $\frac{b}{P}$ | $< \frac{\mu_s V}{2T}$ | sticking | (AI.1) |
| | $> \frac{\mu_s V}{2T}$ | slipping | |

Notice in relations (AI.1) that what matters is the *ratio* of bow velocity to bow pressure, not *b* or *P* separately. The ratio *b/P* thus defines the nature of the motion of the bowed string.

What now, if the string is slipping from the very beginning (bottom relation in Eq. (AI.1))? In that case, the speed *v* of the bowing point *A* of the string will be less than (or may even be oppositely directed to) the speed of the bow *b.* We have a regime of *dynamic* friction, in which it turns out that the force $F = 2\,Tb/V$ (Fig. AI.1) is proportional to *P,* but also depends on the *relative speed b–v* between bow and string (speed of slipping). We write this in the form $F = \mu_d.P$, where $\mu_d$ is the coefficient of *dynamic friction,* which now depends on the relative velocity *b–v* (is a *function* of *b–v*). Thus, during the slipping regime:

$$\frac{b}{P} = \mu_d \frac{V}{2T} \tag{AI.2}$$

If we knew the dependence of $\mu_d$ with the speed of slipping *b–v,* we could use expression (AI.2) to determine the speed *v* of the contact point *A* of the string. Again, this relation is governed by the ratio *b/P.* But notice carefully: what is determined by this ratio is the difference *b–v,* that is, the speed of the string *relative to the bow.* The larger *b* is, the larger will be *v,* for a given value of *b/P.* While this ratio determines the *nature* of the string motion (sticking vs slipping, relations (AI.1)), the bow speed determines the actual speed of the string (for a given *b/P).* Thus, if one increases bow speed but *at the same time* increases bow pressure so as to hold their ratio constant, the nature of the string motion will not change at all—only its velocity will increase linearly with *b.* This will lead to an increase of the amplitude, that is, intensity of sound, in the real case. In other words, *the amplitude of the vibration of a bowed string (loudness of the tone) is solely controlled by the bow velocity, but in order to maintain constant the nature or type of*

[1] The symbol $<$ means "less than", $>$ means "greater than".

*the string motion (timbre of the tone), one must keep the bowing pressure proportional to the bowing speed.*

Let us now consider a little more realistic case: A string of *finite* length *L,* bowed with an infinitesimally thin bow at the midpoint *O* (Fig. AI.2). In this figure, we show schematically the shape of the string as we start bowing (again, the slopes are highly exaggerated); $v$ is the speed of the midpoint (we may either have slipping ($v < b$), or sticking ($v = b$)). Notice that at time $t_4 = L/2V$, the first "wave" (of slope $v/V$) has reached the string′s end points. There, the wave is reflected and superposed with the ongoing initial wave, unfolding the "kinked" shape shown for the times $t_5$ to $t_7$. Then, at $t_8 = L/V$, something new happens (Fig. AI.2): *the slope suddenly changes at the bowing point.* This changes the expression for the force $F$ and a new regime may arise (e.g., slipping, if previously there was sticking).

We cannot continue this discussion without running into considerable mathematical complexities (Keller, 1953). Just notice that these crucial changes in shape (whenever the wave is reflected at the fixed end points) always occur at times that are *integer multiples of L/V,* a quantity that is completely independent of the bowing mechanism. Actually, the inverse of *L/V* appears in the expression (4.3) of the fundamental frequency of the vibrating string. The reader may thus envisage how that frequency (and all upper harmonics) may indeed be excited (and maintained) by the bowing mechanism, and infer from Fig. AI.2 (with a little extra imagination) that in its real vibratory motion, a *bowed string always has an instantaneous shape that is made up of sections of straight lines;* this result has been verified experimentally long ago (see Fletcher and Rossing (1998)).

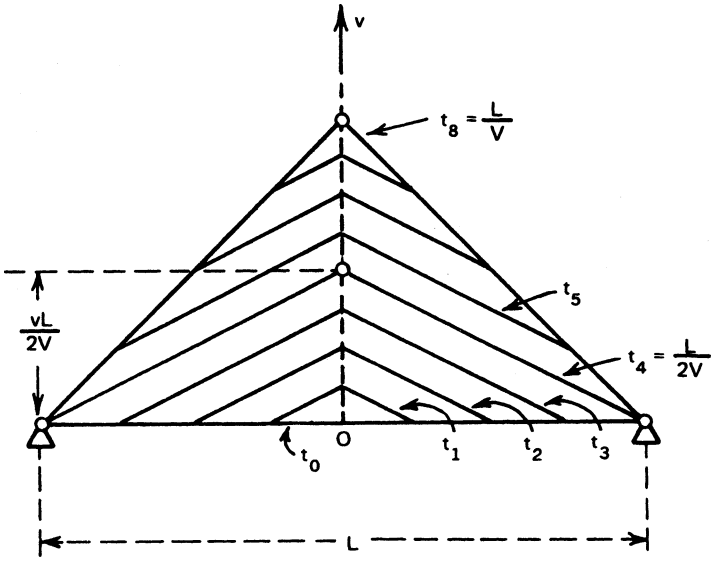



FIGURE AI.2 Same as Fig. AI.1 for a string with fixed end points.

# Appendix II
## Some Quantitative Aspects of Central Pitch Processor Models

In this appendix, we follow up on our discussion of the perception of the pitch of complex tones given in Sects. 2.9, 4.8, and 4.9. In particular, we will show how, with just a little algebra, a simplified model of "template fitting" (Sect. 2.9, p. 69) can explain some quantitative characteristics of complex tone pitch perception. In the second part, we shall speculate about neural models that may accomplish the functions of a central pitch processor and other cognitive operations.

What is a template, really? Normally we think of it as a device that gives us some standard shape or pattern which can be translated, rotated, stretched, and compressed without losing its identifying properties (which could be, for instance, a number, a word, a geometric figure, etc.) In the cases to be discussed below, the template will be the finite set of frequencies as they appear in a complex musical tone: $f_1, f_2, f_3, f_4, \ldots, f_n, \ldots, f_N$. The frequency values are all related to the fundamental frequency $f_1$ (which in the real case of a musical tone would lead to the sensation of its pitch) by the relation $f_n = nf_1$. This is the identifying property that *defines* the template; $f_1$ is the parameter that we want to adjust so as to make the entire template *fit as best as possible* the input signal, which we shall assume to consist of a set of discrete frequencies $f_a, f_b, f_c, \ldots$ (*not* necessarily harmonic) of an auditory signal. The match of the above template will provide us with a value of $f_1$ representing the predicted subjective pitch of the input tone. "Best fit" means that $f_1$ minimizes in some average way the differences between the template values $f_n = nf_1$ and the nearest components of the real signal (see Plomp (1976)).

To illustrate how such a "maximum likelihood estimation" would work, let us apply it to the Smoorenburg experiments (Sect. 2.7). Consider a two-tone stimulus of frequencies $f_a = 1000$ Hz and $f_b = 1200$ Hz. These are the exact fifth and sixth harmonic frequencies of a fundamental $f_1 = 200$ Hz. Let us take our template of frequencies $f_1, 2f_1, \ldots, Nf_1$ whose fundamental $f_1$ can be changed arbitrarily. The matching process consists of finding a frequency $f_1$ for which two *successive* harmonics $nf_1$ and $(n+1)f_1$ coincide with, or are as close as possible to, the input tone frequencies $f_a$ and $f_b$. At this stage, it is irrelevant what the order $n$ of these two harmonics is, provided that the match is the best of all possible ones (the differences $f_a - nf_1$ and $f_b - (n+1)f_1$ are as small as possible). In our example, only one match is best: for $n = 5$ and $f_1 = 200$ Hz, both frequency differences are exactly zero—the match is perfect! The reader can easily verify that there is no other fundamental frequency $f_1$, nor any other value of $n$, that can give a perfect

match. Note carefully that in our example a given match requires determination or estimation of *two* quantities: harmonic order $n$ and fundamental frequency $f_1$. The hypothesis of this template model is that $f_1$ would then correspond to the actual, single, subjective pitch sensation evoked by the two-tone stimulus.

As in Smoorenburg′s experiments, let us now shift the input frequencies to the *inharmonic* pair $f_a = 1050\,\text{Hz}$ and $f_b = 1250\,\text{Hz}$. There is *no* harmonic series of which these two are neighboring components. What pitch would a template matching model predict in this case? There are many possibilities to define "best fit" mathematically; as a first approximation, let us follow the simplest way. We want to find a pair of values $n$, $f_1$ such that the neighboring harmonics $nf_1$ and $(n+1)f_1$ of the template come as close as possible to the pair of input frequencies $f_a$ and $f_b$. Or, what is the same, we want to find $f_1$ and $n$ such that the quantities $f_a/n$ (a subharmonic "projection" of tone $f_a$ toward the fundamental) and $f_b/(n+1)$ (ditto for tone $f_b$) come as close as possible to the fundamental of the template $f_1$. For a moment, let us keep the harmonic order $n$ fixed. For that $n$, let us find the value of $f_1$ that minimizes the *mean square deviation* (= average of the squares of the errors $f_a/n - f_1$ and $f_b/(n+1) - f_1$; also called the square of the standard deviation):

$$\sigma^2 = \frac{1}{2}\left[\left(\frac{f_a}{n} - f_1\right)^2 + \left(\frac{f_b}{n+1} - f_1\right)^2\right] \tag{AII.1}$$

As the theory of experimental errors tells us, the quantity that minimizes $\sigma^2$ is precisely the arithmetic mean of the individual "data", which in our case are $f_a/n$ and $f_b/(n+1)$:

$$f_1 = \frac{1}{2}\left(\frac{f_a}{n} + \frac{f_b}{n+1}\right) \tag{AII.2}$$

Instead of the mean square deviation (AII.1), we can introduce a dimensionless quantity

$$Q(n) = f_1 \big/ \sigma \tag{AII.3}$$

which we may call "quality of fit" for the given order number $n$. A perfect match (as we obtain for $f_a = 1000\,\text{Hz}$, $f_b = 1200\,\text{Hz}$, and $n = 5$) yields $Q(5) = \infty$ ("infinitely good fit"). Given any pair $f_a$ and $f_b$, for different values of $n$ we obtain different quality of fit values $Q(n)$, $Q(n+1)$, .... If among them, one stands out as the largest, the corresponding $f_1$ and $n$ represent the best match for the template—and $f_1$ should be the pitch that is heard!

In our two-tone example of $f_a = 1050\,\text{Hz}$, $f_b = 1250\,\text{Hz}$, we obtain, for $n = 4$, $f_1 = 257\,\text{Hz}$ and $Q(4) = 28$; for $n = 5$, $f_1 = 209\,\text{Hz}$ and $Q(5) = 180$, and for $n = 6$, $f_1 = 177\,\text{Hz}$ and $Q(6) = 70$. Clearly, the choice $n = 5$ leads to the highest $Q$ value. The corresponding frequency (209 Hz) is indeed very close to the subjective pitch that is identified most easily when this inharmonic two-tone complex

is presented (Smoorenburg, 1970). Note that $n = 6$ and, to a lesser extent, $n = 4$ also give a non-negligible quality of fit; this predicts the observed fact that the corresponding fundamental frequencies 177 Hz and 257 Hz also can be heard as "secondary" pitch sensations, although with much more difficulty. This simple model thus explains quantitatively the appearance of ambiguous or multiple pitch sensations elicited by inharmonic tones. If we shift the two-tone stimulus frequencies even further away from harmonicity (but always keeping the same frequency difference of 200 Hz), we obtain the results for $f_1$ (AII.1) and $Q$ (AII.3) shown in Figure AII.1.

Note in Fig. AII.1 that for the pairs 1100/1300 and 900/1100, *two* template matches are more or less equally prominent—they are neither the repetition rate (200 Hz) of the resulting vibration pattern, nor the rate of its amplitude envelope variations (100 Hz). Note also how the harmonic order $n$ for which the perfect fit is obtained shifts from 4 to 6, as the center frequency of the pair $f_a$, $f_b$ is shifted up, while the corresponding best-fit pitch jumps down, from a value lying above 200 Hz to one below 200 Hz. Quite generally, as both $f_a$ and $f_b$ are swept continuously upward in frequency (keeping their difference constant), the main subjective pitch sensation "oscillates" about the repetition rate given by $f_b - f_a = 200$ Hz, but coinciding with the latter only in the harmonic positions. Ambiguous or multiple pitches appear most distinctly whenever the two-tone stimulus lies approximately halfway between harmonic situations. In all these cases, the quality of fit $Q$ (Fig. AII.1(b)) is related to the "clarity" or intelligibility of the corresponding pitch sensation.

It is important to point out that the theoretical values obtained with the above template-matching model do not fit exactly the experimental results, especially when the order $n$ is greater than about 7. The model described here is indeed highly simplified; an improved model must take into account the fact that the spatial resolution of resonance regions on the basilar membrane deteriorates considerably beyond the seventh harmonic (e.g., Fig. 2.25(b)), and that beyond a fundamental frequency of about 1000 Hz, the template process may not work at all (Lin and Hartmann, 1998). Another complication is that combination tones of the type (2.5) and (2.6) (Sect. 2.5) seem to play a role and should be taken into account as additional low pitch signals before any matching procedure is applied (Plomp, 1976).

Without complicating the mathematics too much, we can present a one-step improvement and apply it to the case of a *slightly* out-of-tune series of overtones. Consider a set of frequencies $f_n$, each one very close, but not equal, to an integer multiple $n$ of a fundamental frequency $F$, which we want to determine by a template adjustment. Instead of using a simple average like (AII.1), we introduce so-called weight functions $w_n$ to express the degree of contribution of each partial tone to the extraction of a pitch signal. Without entering in details, let us just state that the $w_n$ must tend to zero when $n \geq 7$. In general, those weight functions would depend on $n, f_n$ and perhaps also on the frequencies and intensities of neighboring partials (to take into account masking effects). With such weight functions, the equivalent of relation (AII.2) becomes the "weighted average":

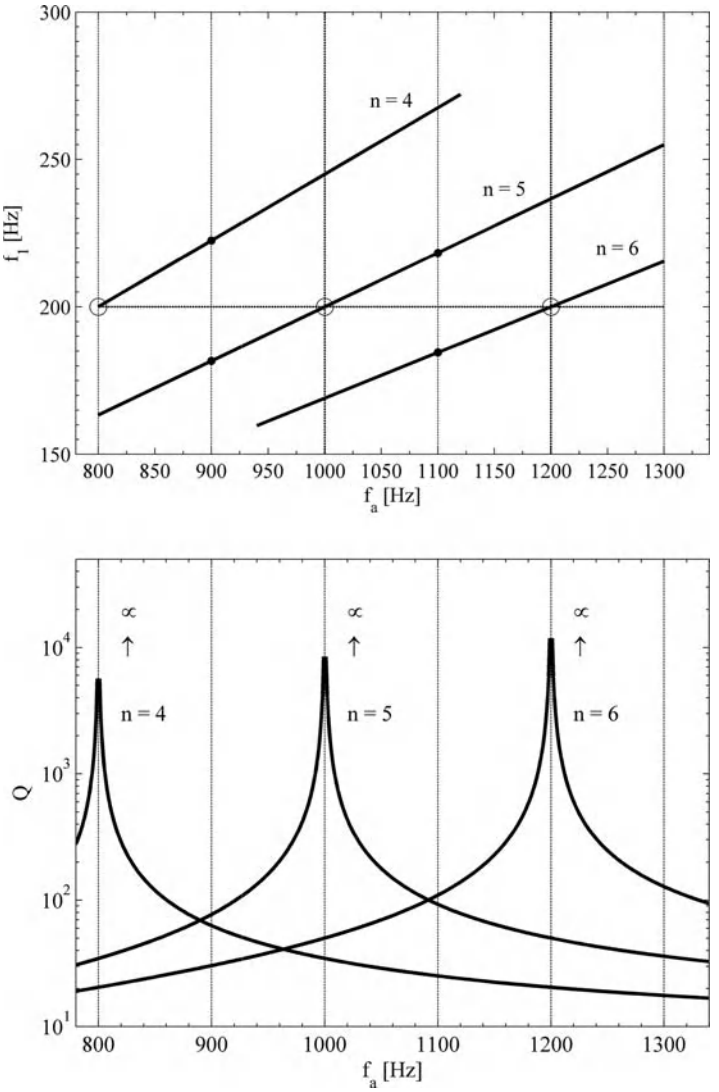



FIGURE AII.1 *Template matching* to determine the pitch of a two-tone stimulus of constant frequency difference ($f_b - f_a$ = 200 Hz). *Top graph*: Perceived fundamental frequency $f_1$ as a function of the root frequency $f_a$, given by the approximate expression (AII.2). *Bottom graph*: Quality of fit $Q$, given by expressions (AII.3) and (AII.1). Only when $f_b$ and $f_a$ are *neighboring* harmonics of a harmonic sequence will the missing fundamental be heard with maximum clarity ($Q = \infty$). Any other position of the tone pair will only elicit ambiguous, multiple pitches (e.g., full dots for $f_a$ = 900 and 1100 Hz). See accompanying text and also discussion of Fig. 2.20.

$$F = \frac{1}{2}\left(w_1\frac{f_1}{1} + w_2\frac{f_2}{2} + w_3\frac{f_3}{3} + \ldots + w_N\frac{f_N}{N}\right)\Big/(w_1 + w_2 + w_3 + \ldots + w_N) \tag{AII.4}$$

$N$ is the total number of single-frequency components of the signal. The Smoorenburg example discussed above corresponds to the case in which the weights $w$ are all zero except for $w_n = w_{n+1} = 1$.

The reader can verify that (AII.4) can be used to demonstrate the case of the "missing fundamental" discussed in Sects. 2.7and 4.8 (set $w_1 = 0$). Furthermore, expression (AII.4) shows that the change $\delta F$ of the pitch of a complex tone caused by slight mistuning of just one of the upper harmonics ($f_n = nF + \delta f_n$) is very small, of the order of $\delta F = \delta f_n/(nN)$ (for all $w$ set = 1). This means that the slight inharmonicities of upper harmonics in musical instruments mentioned in Chap. 4 could have only a very small effect on the resulting subjective pitch. Finally, it is interesting to note that the mechanism of pitch adjustment in a wind instrument mentioned in Sect. 4.5 (p. 144) works in formal analogy to the model of template fitting described above.

Thus far, we have discussed a *mathematical* model. Let us now turn to biological reality and speculate on a *neural* model for the central pitch processor, by invoking the "learning matrix" introduced by Terhardt (1974). This is done here more as an academic exercise and is not intended to represent yet another central pitch theory. We have sketched in Fig. AII.2 a neural wiring scheme capable

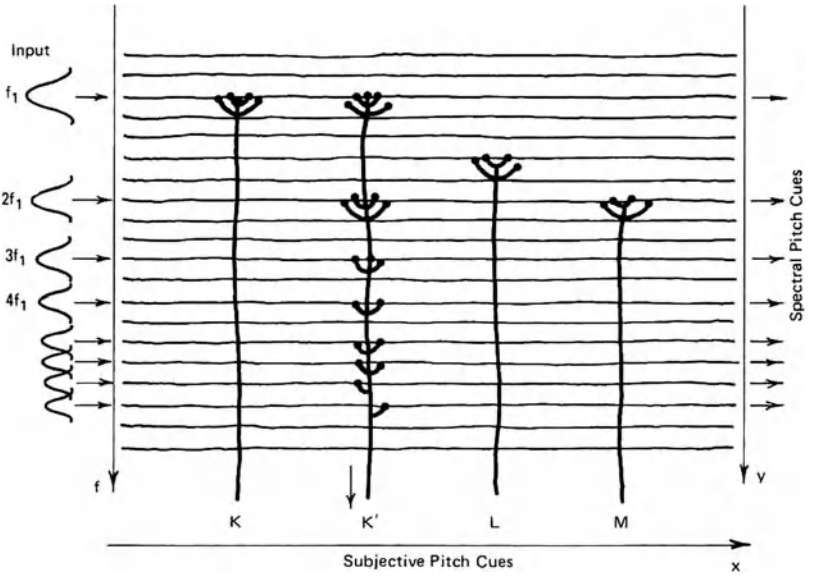



FIGURE AII.2 Model of a neural wiring scheme for pitch extraction and fundamental tracking. *K*, *L*, *M*: initial "untrained" neuron dendrites. *K'*: synaptic configuration of neuron K with new synapses after multiple exposures to a complex tone of harmonics shown along the left axis (see text).

of performing the operations needed for pitch extraction and fundamental tracking. The horizontal fibers are assumed to conduct the combined neural signals from both cochleas into the primary (spectral) pitch processor. This output carries information on the primary pitch of each harmonic component of a complex tone, but is normally ignored at the higher stages of musical tone processing. The horizontal axons are intercepted by a vertical array of neuronal dendrites (Sect. 2.8) as shown in Fig. AII.2. We assume that, initially (at birth), the active synaptic connections are distributed as shown for neurons K, L, and M. We further assume that, in order to reach threshold and fire, each vertical neuron must be activated at many synaptic contacts at nearly the same time. It is clear from this figure that in an acoustically virgin brain, the output activity distribution from the vertical neurons (along the $x$ dimension) would be nearly identical to that of the horizontal fibers (along the $y$ dimension).

Our next assumption, in line with Terhardt′s theory, is that as the ears are exposed repeatedly to harmonic tones, synaptic contacts will also be activated between a vertical neuron and all those horizontal axons that are most likely to be firing at the same time (the "essence" of the learning process in the nervous system, Sect. 2.8, p. 60). When a complex harmonic tone is presented whose fundamental is, say, $f_1$, neuron K (Fig. AII.2) will initially respond only to that fundamental. But as this complex tone stimulus is repeated, the dendrite K will develop active synaptic contacts with all those horizontal fibers whose best frequencies correspond to the higher harmonics of that tone (p. 62). As a result, the vertical neuron emerges "tuned" to the whole harmonic series of $f_1$, as shown for neuron K′ in Fig. AII.2. Trained vertical neurons thus would physically represent a first biological approximation to the "templates" discussed above. Template response will be highest wherever a local best match (local maximum of vertical neuron excitation) is achieved. We finally assume that the location of maxima of output activity from the vertical neurons (along the $x$ dimension, Fig. AII.2) leads to the sensation of subjective pitch. After the learning process, this output is indeed quite different from that of the horizontal fibers (in the $y$ dimension). For instance, if enough synapses are activated by the upper harmonics of $f_1$, neuron K′ will respond, even if the input at the fundamental $f_1$ is missing in the original tone. This represents the mechanism of fundamental tracking. The higher the order of the harmonics, the less sharply defined is the "horizontal" input, because of the proximity of the respective excitation maxima (Fig. AII.2). Vertical neurons may thus be led to respond to the "wrong" input signal (one that does not correspond to the fundamental frequency to which its apical dendritic tree has been originally wired). Multiple pitches are hence possible, as we have shown with a mathematical model in the first part of this Appendix.

Our neural model needs some improvements, though. As shown in Fig. AII.2, the tuned neuron K′ would also respond to all those complex tones whose fundamental frequencies are integer multiples of $f_1$. It may even fire when only one upper harmonic is present. To prevent this undesirable effect from happening, we may introduce an intermediate set of vertical neurons, capable of detecting

coincidences between neighboring harmonics (see above what we have said about weight functions!). It is quite straightforward to write a simple program[1] for a computer which simulates the operation of such a neural model. Some quantitative results are shown in Fig. AII.3 (for simple but realistic assumptions on the distribution of primary excitation around each harmonic and on the degradation of response at increasing harmonic order). At the top of each panel, the primary input power spectrum is given (in linear scales), corresponding to the superposition of two complex tones forming an octave, fifth, and minor third, respectively. The lower graphs represent the computed neural activity distribution along the $x$ dimension (cf. Fig. AII.2). Note the pronounced peaks corresponding to the fundamental frequencies of each original complex tone. We assume that these peaks are recognized at a higher stage of neural processing and lead to the two clear pitch sensations corresponding to the two-tone complex. The position and, to a

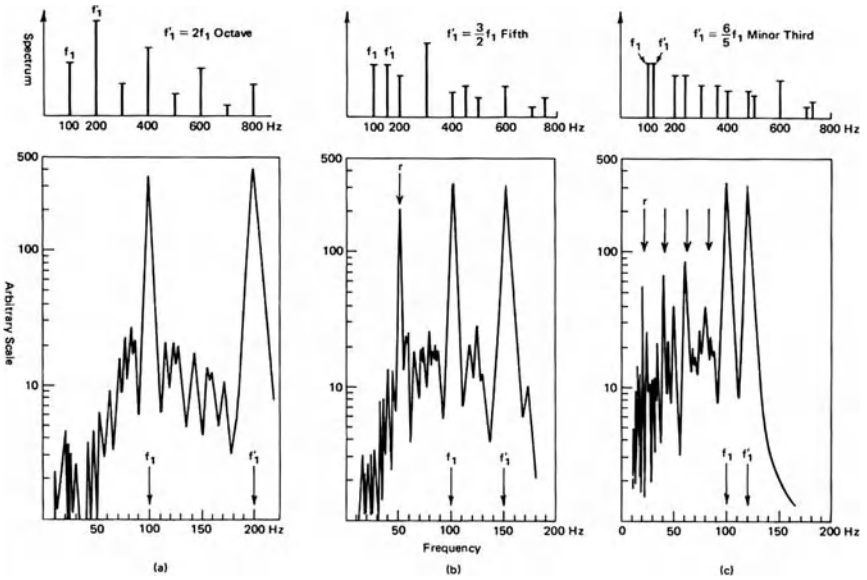



FIGURE AII.3 *Top graphs*: Hypothetical spectra of the superposition of two complex tones forming an octave, fifth, and minor third. *Bottom graph*: Computed output signals using a simple pitch processor model (see text). Note the pronounced peaks at the fundamental frequencies corresponding to each component tone.

[1]This program makes no assumptions on Fourier transforms or autocorrelation functions. The program simply runs a template of harmonic frequencies through a given fundamental frequency domain and counts, for each position, the number of *simultaneous* excitations at neighboring harmonic positions. The total number of pairs of simultaneous excitations represents the output (the intensity or probability of activation of the corresponding vertical neuron).

large extent, the shape of these primary peaks is independent of the actual power spectra of the component tones, depending only on their fundamental frequencies $f_a$ and $f_b$. Note also that, following the order of decreasing consonance (Sec. 5.2), "parasitical" peaks appear at the positions of the repetition rate $r$ and its multiples. These parasitical peaks (which must be deliberately inhibited at some higher stage) are absent in the octave. In addition, there is a background activity or "noise level" below the lower pitch tone that increases with decreasing order of consonance, and may be representative of the sensation of dissonance.

The neural model discussed above is quite primitive and, in its design, quite removed from physiological reality—except for what we called "the essence" of the learning process: the establishment of new synapses or the change of efficiency of existing synapses between neurons (Sect. 2.9). Much progress has been made in recent years in the development of more realistic models of neural networks (e.g., see Hinton (1992)), in part promoted by computer scientists and robotics engineers interested in the design of neural computers. Neural computing is now beginning to be used in the study of pitch perception (see summary in Bigand and Tillmann (2001)) and other musical cognitive tasks. However, a detailed discussion would go far beyond the scope of this book.

# Appendix III
## Some Remarks on Teaching Physics and Psychophysics of Music

It would be unrealistic to make elaborate recommendations on how to organize a truly interdisciplinary course on this subject. The main reason lies in the rather unpredictable composition of the typical student audience signing up for such a course, their widely differing background, and the broad spectrum of interests. Assuming that such a course is open to the whole student body of a university (which it should), it may include five main populations: Majors in music, psychology, the life sciences, mass communications and engineering, and physics and mathematics. The single most general difficulty is to make the course equally interesting, useful, and easily understandable for everyone. This imposes three overall requirements:

1. To minimize the use of mathematics, yet to do it in such a way as not to ridicule the presentation in the eyes of the science majors and engineers. (*Hint*: Use the course to show science majors explicitly how to teach science without math!)
2. To explain everything "from the beginning," be it a topic of physics, psychophysics, or music—yet to do it in such a way as not to appear condescending to the respective "experts." (*Hint*: Use the course to show the experts explicitly how to present concise and comprehensive reviews of topics in their own fields!)
3. To conduct class demonstrations, experiments, and to assign quizzes, essays, and problems that are meaningful, that is, conducted in such a way that the student (irrespective of his/her background) may answer the following question without hesitation: What have I learned by watching the demonstration, by doing this experiment, or by solving this problem? (*Hints*: In the experiments, do not let students watch or make dull measurements for the sake of the measurement—show them how magnitudes relate to each other in nature, how they change with respect to each other, and how they are connected through physical cause-and-effect relationships. In the problems, do not let them "solve equations"—again, show how a given relationship connects two or more quantities "dynamically" over a whole range of variability, lead them toward an intuitive feeling of quantitative relationships between magnitudes; show them how mathematical relationships can be

used to predict the behavior of a system. In quizzes, make them think intuitively yet answer with scientific precision.)

One serious difficulty is that many music (or other fine arts) majors have an inherent "fear" of scientific rigor, in the sense that they assume a priori that "they will not understand." This is exclusively a mental block that can be dispersed successfully with patience, persuasion, and dedication to the individual on part of the teacher. (*Hint*: Convince them that if they are able to balance their checking account each month, they will be able to understand the little bit of math needed in this course!)

The inclusion of psychoacoustics in an introductory course of musical acoustics (which I find absolutely *essential*!) presents a number of additional challenges to the instructor. First, there is the most obvious one: how to fit everything into the available time. No matter how short or how long that time is, hard decisions will have to be made on which topics should be left out and which should be included. Second, psychoacoustics and the related topics of neurosciences are subjects perhaps even less familiar to the students of a musical acoustics course than physics. This makes it necessary to restrict these topics to just a few relevant and interesting ones. It will help to point out right at the beginning of the course some relevant aspects of psychoacoustics. For instance, alert the student that the recent insights gleaned in music perception can be incorporated into the creation of new frontiers in music composition. Point out that many of the fallacies that exist about musical performance do have a root in the particular modes of acoustical information processing in the ear and the brain. Point out that many technical requirements of high quality electroacoustic equipment are related directly to particular aspects of signal processing in the nervous system. Point out that the understanding of music perception should not only be of interest to musicians but that it can also benefit biologists in terms of learning about certain brain functions; and that psychologists may obtain, from this field, quantitative information of relevance to music therapy.

A general difficulty is the fact that sophisticated experimental demonstrations concerning this subject involve very expensive equipment. Nevertheless, it is possible to get along with a minimum baseline that only requires equipment that most likely can be borrowed from other courses or departments. This minimum baseline, with which indeed most of the experiments briefly touched upon in this book can be demonstrated, is summarized below.

## *1 Psychoacoustic Experimentation*

(1) Two sine-wave generators, a good quality amplifier, good headphones for everybody, two hi-fi loudspeakers. (2) One oscilloscope for each group of four to six students, if possible, with double-beam display and memory trace. (3) An electronic synthesizer (a portable version is sufficient). With this equipment, it is perfectly feasible to demonstrate almost everything that is mentioned on pp. 24,

27, 33, 38, 41, 43, 45, 48, 101, 155, 170–172 and 181. An excellent description of selected experiments and classroom demonstrations can be found in Hartmann (1975). If the class has access nearby to a large pipe organ in a good acoustical environment, additional useful demonstrations can be made (e.g., see pp. 51, 101, 104, 151, 154, and 168). In all these experiments or demonstrations, it should be a rule that *whatever is being sounded should always be simultaneously displayed on the oscilloscope*.

## *2 Acoustic Experimentation*

(1) One "sonometer" (Sect. 4.1) per group of four to six students, with stroboscope and appropriate circuitry to conduct the experiments described on pp. 114–116. This setup also provides for experimental discussion of relations (4.2), (4.3), and of bowing and plucking mechanisms. (2) A piano is useful for performing the simple demonstrations of pp. 117 and 118–119. (3) Single organ pipes (usually available in physics departments) to explore resonance curves of the type of Fig. 4.24, and relations (4.5), (4.6), by using a small speaker of good quality "implanted" in the pipe. (4) Video loops and ripple tanks, also usually available, are extremely useful to demonstrate traveling waves, standing waves, and acoustical optics in general. (5) Acoustics experimentation must be accompanied by meaningful homework problems. An excellent set can be found in Savage (1977).

In addition to all this, it is advisable to assign individual students to conduct special studies and write essays on a musical instrument of their choice which, of course, will require access to appropriate literature. To sum up, this is a course that is challenging and fun to teach—the perhaps most interdisciplinary of all that a university can offer at a freshman level. It presents a chance to both teacher and student alike, to let the imagination fly high—within the strict boundaries of science!

# References


Allen, J.B., and S.T. Neely. 1992. Micromechanical models of the cochlea. *Phys. Today*, July 1992, 40.

Ando, Y. 1985. *Concert Hall Acoustics*. Springer-Verlag, Berlin.

Arbib, M.A. 1987. *Brains, Machines and Mathematics.* Springer-Verlag, 2nd ed. New York.

Ashmore, J. 2008. Cochlear outer hair cell motility. *Physiol. Rev.* **1**:173.

Askenfelt, A., and E. Jansson. 1990. In *The Acoustics of the Piano*, Publ. of the Royal Swedish Academy of Music **64**:36.

Backus, J. 1974. Input impedance for the reed woodwind instruments. *J. Acoust. Soc. Am.* **56**:1266.

Backus, J., and T.C. Hundley. 1971. Harmonic generation in the trumpet. *J. Acoust. Soc. Am.* **49**:509.

Benade, A.H. 1971. Physics of wind instrument tone and response. In *Symposium on Sound and Music, December 1971*. American Association for the Advancement of Science, Washington, D.C.

Benade, A.H. 1973. The physics of brasses. *Sci. Am.* **229**(1):24.

Benade, A.H. 1990. *Fundamentals of Musical Acoustics*. 2nd revised ed. Dover Publications, New York.

Benzon, W. 2001. *Beethoven's anvil.* Basic Books, New York.

Bharucha, J.J. 1994. Tonality and expectation. In *Musical Perceptions*. R. Aiello, ed. Oxford University Press, Oxford. 213.

Bickerton, D. 1995. *Language and Human Behavior.* University of Washington Press, Seattle, WA.

Bigand, E., and B. Tillmann. 2005. Effect on context on the perception of pitch structures. In *Pitch*. C.J. Plack, O.J. Oxenham, R.R. Fay and A.N. Popper, eds. Springer-Verlag, New York.

Bilsen, F.A., and J.L. Goldstein. 1974. Pitch of dichotically delayed noise and its possible spectral basis. *J. Acoust. Soc. Am.* **55**:292.

Binder, J.R. 1999. In *Functional MRI*. C.T.W. Moonen and P.A. Bandettieri, eds. Springer, Berlin. 393.

Blood, A.J., and R.J. Zatorre. 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. USA* **98**:11818.

Borchgrevink, H.M. 1982. Prosody and musical rhythm are controlled by the speech hemisphere. In *Music, Mind, and Brain*. M. Clynes, ed. Plenum Press, New York, 151.

Bradshaw, J.L., and N.C. Nettleton. 1981. The nature of hemispheric specialization in man. *Behavioral and Brain Sci.* **4**:51.

Bredberg, G., H.H. Lindemann., H.W. Ades., R. West., and H. Engstrom. 1970. Scanning electron microscopy of the organ of Corti. *Science* **170**:861.

Bregman, A.S. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, Cambridge, Massachusetts (with CD with auditory demonstrations).

Bregman, A.S., and J. Campbell. 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. *J. Experim. Psychol.* **89**:244.

Brodal, A. 1969. *Neurological Anatomy*. Oxford University Press, London.
Camalet, S., Th. Duke., F. Jülicher., and J. Prost. 2000. Auditory sensitivity provided by self-tuned critical oscillations of hair cells. *Proc. Natl. Acad. Sc. USA* **97**:3183.
Cohen, M.R., and I.E. Drabkin. 1948. *A Source Book in Greek Science*. McGraw-Hill Book Company, Inc., New York.
Corso, J.F. 1957. Absolute judgments of musical tonality. *J. Acoust. Soc. Am.* **29:**138.
Cremer, L. 1984. *Physics of the Violin*. MIT Press, Cambridge, Massachusetts.
Dallos, P. 1992. The active cochlea. *J. Neurosci*. **12**:4575.
Damasio, A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Hartcourt, Inc., New York.
Damaske, P. 1971. Heat-related two-channel stereophony with loud-speaker reproduction. *J. Acoust. Soc. Am.* **50**:1109.
Davis, A. 1962. Advances in the neurophysiology and neuroanatomy of the cochlea. *J. Acoust. Soc. Am.* **34**:1377.
de Boer, E. 1983. No sharpening? A challenge to cochlear mechanics. *J. Acoust. Soc. Am.* **73**:567.
de Cheveigné, A. 2005. Pitch perception models. In *Pitch*. C.J. Plack, O.J. Oxenham, R.R. Fay, and A.N. Popper, eds. Springer-Verlag, New York.
Denenberg, V.H. 1981. Hemispheric laterality in animals and the effects of early experience. *Brain and Behavioral Sci*. **4**:1.
d'Errico, F., C. Henshilwood., G. Lawson., M. Vanhaeren., A.-M. Tillier., M. Soressi., F. Bresson., B. Maureille., A. Nowell., I.A. Lakarra., L. Backwell., and M. Julien. 2003. The emergence of language, symbolism and music—an alternative multidisciplinary perspective. *J. World Prehistory* **17**(1):2.
Deutsch, D., ed. 1982a. *The Psychology of Music*. Academic Press, New York.
Deutsch, D. 1982b. Organizational processes in music. In *Music, Mind, and Brain*. M. Clynes, ed. Plenum Press, New York, 119.
Deutsch, D. 1996. *Musical Illusions and Paradoxes*. CD, Philomel Records, Inc., La Jolla.
Deutsch, D. 2004. Guest editorial. *Music Percept.* **21**:285–287.
Dolan, R.J. 2002. Emotion, cognition and behavior. *Science* **298**:1191.
Egan, J.P., and H.W. Hake. 1950. On the masking pattern of a simple auditory stimulus. *J. Acoust. Soc. Am.* **22**:622.
Feeney, M.F. 1997. Dichotic beats of mistuned consonances. *J. Acoust. Soc. Am.* **102**(4):2333.
Flanagan, J.L. 1972. *Speech Analysis, Synthesis and Perception*, 2nd ed., Springer-Verlag, New York.
Fletcher, H., and W.A. Munson. 1933. Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Am.* **5**:82.
Fletcher, N.H., and T.D. Rossing. 1998. *The Physics of Musical Instruments*. Springer-Verlag, New York.
Friedlander, F.G. 1953. On the oscillations of a bowed string. *Cambridge Philos. Soc. Proc*. **49**:516.
Gazzaniga, M.S. 1970. *The Bisected Brain*. Meredith, Des Moines.
Gelfand, S.A. 1990. *Hearing*. Marcel Dekker, New York.
Geschwind, N. 1972. *Language and the brain. Sci. Am.* **226**(4):76.
Globus, A., M.R. Rosenzweig, E.L. Bennett., and M.C. Diamond. 1973. *J. Comp.Physiol. Psych*. **82**:175.
Goldberg, J.M., and P.B. Brown. 1969. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanics of sound localization. *J. Neurophysiol*. **32:**613.
Goldstein, J.L. 1970 Aural combination tones. In *Frequency Analysis and Periodicity Detection in Hearing*. R. Plomp, and G.F. Smoorenburg, eds. A. W. Suithoff, Leiden. 230.
Goldstein, J.L. 1973. An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* **54**:1496.

Gray, P.M., B. Krause, J. Atema, R. Payne, C. Krumhansl, and L. Baptista. 2001. The music of nature and the nature of music. *Science* **291:**52.

Hall, D., and A. Askenfelt. 1988. Piano string excitation V: Spectra for real hammers and strings. *J. Acoust. Soc. Am.* **83**:1627.

Halpern, A.R. 2001. Cerebral substrates of musical imagery. In *The Biological Foundations of Music*. R.J. Zatorre and I. Peretz, eds. Annals of the New York Academy of Sciences, New York. 179.

Han, C.J., C.M. O'Tuathaigh, L. van Trigt, J.J. Quinn, M.S. Fanselau, R. Mongeau, C. Koch, and D.J. Anderson. 2003. Trace but not delay fear conditioning requires attention and the anterior cingulated cortex. In *Proc. Natl. Acad. Sci. USA* **100**:13087.

Hartmann, W.M. 1975. The electronic music synthesizer and the physics of music. *Am. J. Phys*. **43**:755.

Hartmann, W.M. 1993. On the origin of the enlarged melodic octave. *J. Acoust. Soc. Am.* **93**:3400.

Hartmann, W.M. 1996. Pitch, periodicity, and auditory organization. *J. Acoust. Soc. Am.* **100**(6):3491–3502.

Hartmann, W.M. 2005. *Signal, Sound and Sensation*. Springer-Verlag, New York.

Hebb, D. 1949. *Organization and Behaviour*. Wiley and Sons, New York.

Herholz, K. 2004. *NeuroPET: Positron Emission Tomography in Neuroscience and Clinical Neurology*. Springer Verlag, Berlin.

Hinton, G.E. 1992. How neural networks learn from experience. *Sci. Am.* **267**(3):145.

Hohne, K.-H. 2001. *VOVEL-MAN 3D Navigator: Brain and Skull*. Springer, Berlin.

Hosokawa, T., D.A. Rusakov, T.V.P Bliss, and A. Fine. 1995. *J. Neurosci.* **15**(8):5560.

Houtsma, A.J.M. 1970. Perception of musical pitch. *J. Acoust. Soc. Am.* **48**:88(A).

Houtsma, A.J.M., J.L. Goldstein 1972. Perception of musical intervals: Evidence for the central origin of the pitch of complex tones. *J. Acoust. Soc. Am.* **51**:520.

Hudspeth, A.J. 1985. The cellular basis of hearing: The biophysics of hair cells. *Science* **230**:745.

Hudspeth, A.J. 1989. How the ear′s works work. *Nature* **341**:397.

Huron, D. 2001. Is music an evolutionary adaptation? In *The Biological Foundations of Music*. R.J. Zatorre and I. Peretz, eds. Annals of the New York Academy of Sciences, New York. 43.

Hutchins, C.M., and F.L. Fielding. 1968. Acoustical measurements of violins. *Phys. Today* **21**(7):34.

Iverson, P., and C.L. Krumhansl. 1993. Isolating the dynamic attributes of musical timbre. *J. Acoust. Soc. Am.* **94**:2595.

Jansson E., N.-E. Molin, and H. Sundin. 1970. Resonances of a violin body studied by hologram interferometry and acoustical methods. *Physica Scripta* **2**:243.

Johnston, I. 2003. *Measured Tones: The Interplay of Physics and Music*. Institute of Physics Publishing, Bristol.

Johnstone, B.M., R. Patuzzi, and P. Sellick. 1983. Comparison of basilar membrane, hair cell and neural responses. In *Hearing - Psychological Bases and Psychophysics*. R. Klinke and R. Hartmann, eds. Springer-Verlag, Berlin. 46.

Johnstone, B.M., R. Patuzzi, and G.K. Yates. 1986. Basilar membrane measurements and the traveling wave. *Hearing Res*. **22**:147.

Kachar, B., W.E. Brownell, W.E. Altschuler, and J. Fex. 1986. Electrokinetic shape changes of cochlear outer hair cells. *Nature* **322**:365.

Kameoka, A., and M. Kuriyagawa. 1969. Consonance theory Part II: Consonance of complex tones and its calculation method. *J. Acoust. Soc. Am.* **45**:1460.

Keller, J.B. 1953. Bowing of violin strings. *Comm. Pure Appl. Math*. **6**:483.

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

Kennedy, H.J., A.C. Crawford, and R. Fettiplace. 2005. Force generation by mammalian hair bundles supports a role in cochlear amplification. *Nature* **433**:880.

Kiang, N.Y.-S., T. Watanabe, E.C. Thomas, and L.F. Clark. 1965. *Discharge Patterns of Single Fibers in the Cat′s Auditory Nerve*. MIT Press, Cambridge, Massachusetts.

Kimura, D. 1963. Right temporal lobe damage. *Arch. Neurol*. **8**:264.

Klein, W., R. Plomp, and L.C.W. Pols. 1970. Vowel spectra, vowel spaces, and vowel identification. *J. Acoust. Soc. Am.* **48**:999.

Koch, C. 2004. *The Quest for Consciousness: A Neurobiological Approach*. Roberts and Co., Englewood, Colorado.

Koelsch, S. 2005. Ein neurokognitives Modell der Musikperzeption. *Musiktherapeutische Umschau* **26**:365.

Kohonen, T. 1988. *Self-Organization and Associative Memory*, 2nd ed. Springer Verlag, Berlin.

Küppers, B.O. 1990. *Information and the Origin of Life.* MIT Press, Cambridge.

Lerdahl F., and R. Jackendorff. 1983. *A Generative Theory of Tonal Music*. MIT Press, Cambridge.

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

Licklider, J.C.R. 1959. Three auditory theories. In *Psychology: A Study of a Science*, Vol. 1. S. Koch, ed. McGraw-Hill, New York.

Lin, J.Y., and W.M. Hartmann. 1998. The pitch of a mistuned harmonic: Evidence for a template model. *J. Acoust. Soc. Am.* **103**(5):2606.

Lipps, T. 1905. *Psychologische Studien*. Durr′sche Buchhandlung, Leipzig.

Marr, D. 1982. *Vision*. W.H. Freemann and Co, San Francisco.

Matthews, M.V., and J. Kohut. 1973. Electronic stimulation of violin resonances. *J. Acoust. Soc. Am.* **53**:1620.

Melcher, J.R., T.M. Talavage, and M.P Harms. 1999. In *Functional MRI*. C.T.W. Moonen and P.A. Bandettieri, eds. Springer-Verlag, Berlin. 393.

Mersenne, M. 1636. *Harmonie Universelle*. Cramoisy, Paris (reprinted 1975, Editions du CNRS, Paris).

Meyer, M. 1900. Elements of a psychological theory of melody. *Psych. Rev*. **7**:241.

Milner, B. 1967. Brain mechanisms suggested by studies of temporal lobes. In *Brain Mechanisms Underlying Speech and Language*. C.H. Millikan and F.L. Darley, eds. Grune and Stratton, New York.

Milner, B., L. Taylor, and R.W. Sperry. 1968. Lateralized suppression of dichotically presented digits after commisural section in man. *Science* **161**:184.

Miyashita, Y. 2004. Cognitive memory: Cellular and network machineries and their top-down control. *Science* **306**:435.

Molino, J.A. 1973. Pure-tone equal-loudness contours for standard tones of different frequencies. *Percept. Psychophys*. **14**:1.

Molino, J.A. 1974. Psychophysical verification of predicted interaural differences in localizing distant sound sources. *J. Acoust. Soc. Am.* **55**:139.

Moonen, C.T.W., and P.A. Bandettieri, eds. 1999. *Functional MRI.* Springer-Verlag, Berlin.

Moore, B.C.J. 1973. Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.* **54**:610.

Morell, V. 1996. Setting a biological stopwatch. *Science* **271**:905–906.

Papçun, G., S. Krashen, D. Terbeek, R. Remington, and R. Harshman. 1974. Is the left hemisphere specialized for speech, language and/or something else? *J. Acoust. Soc. Am.* **55**:319.

Patterson, B. 1974. Musical dynamics. *Sci. Am*. **231**(5):78.

Penfield, W., and L. Roberts. 1959. *Speech and Brain Mechanisms*. Princeton University Press, Princeton, NJ.

Peretz, I. 2001a. Music perception and recognition. In *The Handbook of Cognitive Neuropsychology*. B. Rapp, ed. Psychology Press, Hove, UK. 521.

Peretz, I. 2001b. Brain specialization for music. In *The Biological Foundations of Music*. R.J. Zatorre and I. Peretz, eds. Annals of the New York Academy of Sciences, New York. 153.

Peretz, I., and R.J. Zatorre. 2005. Brain organization for music processing. *Annu. Rev. Psychol*. **56**:89.
Pierce, J.R. 1983. *The Science of Musical Sound*. Scientific American Books, W.H. Freeman, San Francisco.
Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. William Morrow, New York.
Plack, C.J., and O.J. Oxenham. 2005. The psychophysics of pitch. In *Pitch*. C.J. Plack, O.J. Oxenham, R.R. Fay, and A.N. Popper, eds. Springer-Verlag, New York. 7.
Plack, C.J., O.J. Oxenham, R.R. Fay, and A.N. Popper, eds. 2005. *Pitch*. Springer-Verlag, New York.
Plomp, R. 1964. The ear as a frequency analyzer. *J. Acoust. Soc. Am.* **36**:1628.
Plomp, R. 1965. Detectability threshold for combination tones. *J. Acoust. Soc. Am.* **37**:1110.
Plomp, R. 1967a. Beats of mistuned consonances. *J. Acoust. Soc. Am.* **42**:462.
Plomp, R. 1967b. Pitch of complex tones. *J. Acoust. Soc. Am.* **41**:1526.
Plomp, R. 1970. Timbre as a multidimensional attribute of complex tones. In *Frequency Analysis and Periodicity Detection in Hearing*. R. Plomp and F.G. Smoorenburg, eds. A.W. Suithoff, Leiden. 397.
Plomp, R. 1976. *Aspects of Tone Sensations*. Academic Press, New York.
Plomp, R., and M.A. Bouman. 1959. Relation between hearing threshold and duration for tone pulses. *J. Acoust. Soc. Am.* **31**:749.
Plomp, R., and W.J.M. Levelt. 1965. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**:548.
Plomp, R., and H.J.M. Steeneken. 1973. Place dependence of timbre in reverberant sound fields. *Acustica* **28**:50.
Premack, D. 2004. Is language key to human intelligence? *Science* **303**:318–320.
Rabinovich, M., R. Huerta., and G. Laurent. 2008. Transient dynamics for neural processing. *Science* **321**:48.
Raiford, C.A., and E.D. Schubert. 1971. Recognition of phase changes in octave complexes. *J. Acoust. Soc. Am.* **50**:559.
Rakowski, A. 1971. Pitch discrimination at the threshold of hearing. *Proc. 7th Int. Congr. Acoust. Budapest.* **3**:373.
Rakowski, A. 1972. Direct comparison of absolute and relative pitch. *Proc. Symp. Hearing Theory*, IPO, Eindhoven.
Ratliff, F. 1972. Contour and contrast. *Sci. Am.* **226**(6):91.
Reinicke, W., and L. Cremer. 1970. Application of holographic interferometry to vibrations of the bodies of string instruments. *J. Acoust. Soc. Am.* **48**:988.
Richards, A.M. 1977. Loudness perception for short-duration tones in masking noise. *J. Speech Hearing Res*. **20**:684.
Ritsma, R.J. 1967. Frequencies dominant in the perception of the pitch of complex sounds. *J. Acoust. Soc. Am.* **42**:191.
Roederer, J.G. 1978. On the relationship between human brain functions and the foundations of physics, Science, and technology. *Found. Phys*. **8**:423.
Roederer, J.G. 1984. The search for a survival value of music. *Music Percept*. **1**:350.
Roederer, J.G. 2005. *Information and its Role in Nature*. Springer-Verlag, Heidelberg.
Roederer, J.G. 2009. Biological conditions for the emergence of musical arts in a civilization of intelligent beings. In *Between Worlds: The Art and Science of Interstellar Message Composition*. D. Vacoch, ed. The MIT Press, Boston, (in press).
Rose, J.E., J.F. Brugge, D.J. Anderson, and J.E. Hind. 1969. Some possible neural correlates of combination tones. *J. Neurophys*. **32**:402.
Ruggero, M.A., and N. Rich. 1991. Application of a commercially manufactured Doppler shift velocimeter to the measurement of basilar membrane motion. *Hearing Res*. **51**:215.
Ruggero M.A., N.C. Rich, S.S. Narayan, and L. Robles. 1997. Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.* **101**:2151.

Sachs, M.B., and P.J. Abbas. 1974. Rate versus level functions for auditory-nerve fibers in cats: Tone-burst stimuli. *J. Acoust. Soc. Am.* **56**:1835.

Saunders, F.A. 1946. The mechanical action of instruments of the violin family. *J. Acoust. Soc. Am.* **17**:169.

Savage, W.R. 1977. *Problems for Musical Acoustics*. Oxford University Press, New York.

Scharf, B. 1983. Loudness adaptation. In *Hearing Research and Theory,* Vol. 2. J.V Tobias and E.D. Schubert, eds. Academic Press, New York.

Schellenberg, E.G., and S.E. Trehub. 2008. Is there an Asian advantage for pitch memory? *Music Percept*. **25**:241.

Schelleng, J.C. 1973. The bowed string and the player. *J. Acoust. Soc. Am.* **53**:26.

Shannon, C.E., and W.W. Weaver. 1949. *The Mathematical Theory of Communication*. Univ. Illinois Press, Urbana.

Siebert, W.M. 1970. Frequency discrimination in the auditory system: Place or periodicity mechanisms? *Proc. IEEE* **58**:723.

Small, A.M. 1970. Periodicity pitch. In *Foundations of Modern Auditory Theory*. J.V. Tobias, ed. Academic Press, New York. 1.

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

Smoorenburg, G.F. 1972. Audibility region of combination tones. *J. Acoust. Soc. Am.* **52**:603.

Sokolich, W.G., and J.J. Zwislocki. 1974. Evidence for phase oppositions between inner and outer hair cells. *J. Acoust. Soc. Am.* **55**:466.

Stevens, S.S. 1955. Measurement of loudness. *J. Acoust. Soc. Am.* **27**:815.

Stevens, S.S. 1970. Neural events and the psychophysical law. *Science* **170**:1043.

Stevens, S.S., I. Volkmann, and E.B. Newman. 1937. A scale for the measurement of psychological magnitude pitch. *J. Acoust. Soc. Am.* **8**:185.

Sundberg, J., ed. 1992. *Gluing Tones*. Royal Swedish Academy of Music (with compact disc), Stockholm.

Terhardt, E. 1971. Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. *Proc. 7th Int. Congr. Acoust. Budapest.* **3**:621.

Terhardt, E. 1972. Zur Tonhöhenwahrnehmung von Klängen, I, II. *Acustica* **26**:173, 187.

Terhardt, E. 1974. Pitch, consonance and harmony. *J. Acoust. Soc. Am.* **55**:1061.

Terhardt, E. 1998. *Akustische Kommunikation. Grundlagen mit Hörbeispielen*. Springer-Verlag, Berlin.

Terhardt, E., and H. Fastl. 1971. Zum Einfluss von Störtönen und Störgeräuschen auf die Tonhöhe von Sinustönen. *Acustica* **25**:53.

Terhardt, E., G. Stoll, and M. Seewann. 1982. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. Acoust. Soc. Am.* **71**:679.

Terhardt, E., and M. Zick. 1975. Evaluation of the tempered tone scale in normal, stretched and contracted intonation. *Acustica* **32**:268.

Tramo, M.J., P.A. Cariani, B. Deglute, and L.D. Braida. 2001. Neurobiological foundations for the theory of harmony in Western tonal music. In *The Biological Foundations of Music*, R.J. Zatorre and I. Peretz, eds. Annals of the New York Academy of Sciences, New York. 92.

Trehub, S.E. 2001. Musical predispositions in infancy. In *The Biological Foundations of Music.* R.J. Zatorre and I. Peretz, eds. Annals of the New York Academy of Sciences, New York. 1.

Tsao, D.Y., W.A. Freiwald, R.B.H. Tootell, and M.S. Livingstone. 2006. A cortical region consisting entirely of face-selective cells. *Science* **311**:670.

van Noorden, L.A.P.S. 1975. *Temporal Coherence in the Perception of Tone Sequences*. Institute for Perception Research, Eindhoven (with a phonographic demonstration record).

von Békésy, G. 1960. *Experiments in Hearing*. McGraw Hill Book Company, New York.

von Helmholtz, H. 1877. *On the Sensations of Tone.* English translation A.J. Ellis, 1954. Dover Publications, New York.

Walliser, K. 1969. Über die Abhängigkeit der Tonhöhenempfindung von Sinustönen von Schallpegel, von überlagertem drosselndem Störschall und von der Darbietungsdauer. *Acustica* **21**:211.

Ward, W.D. 1970. Musical perception. In *Foundations of Modern Auditory Theory*. J.V. Tobias, ed. Academic Press, New York. 405.

Whitlock, J.R., A.J. Heynen, M.G. Shuler, and M.F. Bear. 2006. Learning induces long-term potentiation in the hippocampus. *Science* **313**:1093.

Wightman, F.L. 1973. The pattern-transformation model of pitch. *J. Acoust. Soc. Am.* **54**:407.

Yost, W.A., and C.S. Watson, eds. 1987. *Auditory Processing of Complex Sounds*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Zatorre, R.J., and I. Peretz, eds. 2001. *The Biological Foundations of Music*. Annals of the New York Academy of Sciences, New York.

Zwicker, E., and H. Fastl. 1999. *Psychoacoustics*, 2. Aufl. Springer-Verlag, Berlin.

Zwicker, E., G. Flottorp, and S.S. Stevens. 1957. Critical bandwidth in loudness summation. *J. Acoust. Soc. Am.* **29**:548.

Zwicker, E., and B. Scharf. 1965. A model of loudness summation. *Psych. Rev*. **72**:3

Zwislocki, J.J. 1965. Analysis of some auditory characteristics. In *Handbook of Mathematical Psychology*. R.D. Luce, R.R. Bush, and E. Galanter, eds. Wiley, New York.

Zwislocki, J.J. 1969. Temporal summation of loudness: An analysis. *J. Acoust. Soc. Am.* **46**:431.

Zwislocki, J.J., and W.G. Sokolich. 1973. Velocity and displacement responses in auditory-nerve fibers. *Science* **182**:64.

# Index

# About the Author

Juan G. Roederer, Professor Emeritus at the University of Alaska-Fairbanks, is a space scientist of international reputation. Italian born, raised in Austria, and educated in Argentina, he received a PhD in physics from the University of Buenos Aires. From 1956-1966 he was professor of physics at that university. In 1967 he and his family emigrated to the United States, where he became professor of physics at the University of Denver. In 1977 he was appointed director of the world-renowned Geophysical Institute of the University of Alaska in Fairbanks, a post he held until 1986. Since then he has taught and conducted research at that university. He served two U.S. presidents as chairman of the United States Arctic Research Commission, and for many years held leading offices in several international scientific organizations. Between 1997 and 2003 he was Senior Adviser of the International Centre for Theoretical Physics in Trieste, Italy.

Professor Roederer is author of over 250 articles in scientific journals on space physics, science policy, psychoacoustics and informatics, and he has written five

university-level textbooks, three of which were translated into foreign languages. He is a corresponding member of the Academies of Science of Austria and Argentina and of the Academy of Sciences for the Developing World. He received four NASA awards for his collaboration in the "Galileo" mission to Jupiter, and was among 100 leading geophysicists worldwide who received the medal "100 Years in Geophysics" from the former Soviet Academy of Sciences.

His close and active relation with music—he studied organ with the late maestros Héctor Zeoli in Buenos Aires and Hans Jendis in Göttingen—prompted Roederer in the early 1970s to write a syllabus for his course "Physics of Music" at the University of Denver, of which the first edition of this book was an offspring. Between 1973 and 1985 he organized the international workshops on the *Neuropsychological Foundations of Music* at the Carinthian Music Festival in Ossiach, Austria, credited for having launched the interdisciplinary approach to the study of music perception. In 2007 he gave the Opening Lecture at the congress *Mozart and Science* in Baden/Vienna, Austria.