This book arose from my experiences in teaching speech science at both undergraduate and graduate levels over the past ten years. When I first started teaching the subject, I was surprised at how intimidated students were by the topic. It soon became clear that a key factor in alleviating their anxiety and promoting a positive learning experience was to demystify the subject by breaking the material into small, logically linked units and by making the organizational links between units of information explicit. Students are thus able to use the basic concepts as a scaffold upon which to extend their scientific knowledge. Careful structuring of the information helps students to grasp the material more easily, to retain the material beyond the next quiz or exam, and to integrate previously learned material with new concepts.

However, although this scaffolding approach facilitates student understanding of the material, it does not help students to appreciate the relevance of the subject. “After all” was the frequently heard comment, “I’m going to be a clinician, not a researcher. I don’t need this stuff to do therapy.” It became clear, over the years, that the way to help students to appreciate the relevance of the speech science material is to explicitly connect it to clinical practice. Once students understand the connection between “science” and “clinic,” the false dichotomy disappears, and students begin to appreciate the necessity of using scientific means of establishing rationales for clinical procedures, evaluating the effectiveness of treatment strategies, increasing clinical accountability, and applying the products of science and technology to clinical practice. Indeed, a discipline can be evaluated on the scope and depth of its scientific basis.

The explosion of computer technology over the past few decades has had an unprecedented effect on all areas of speech science and speech–language pathology. Due to technological advances such as electron microscopy, magnetic resonance imaging, and ultrasonography, basic knowledge of the structures and functions of all systems involved in speech production and perception has expanded tremendously. The application of this scientific
knowledge base to the diagnosis and treatment of communication disorders has resulted in dramatic changes and refinements. Acoustic, aerodynamic, and physiological instrumentations are rapidly becoming accepted tools in hospitals, rehabilitation centers, clinics, schools, and other settings in which human communicative behavior is of interest.

This book takes a systems approach to the scientific study of speech production and speech perception, focusing on the physiological and acoustic generation and measurement of verbal output. Rather than studying scientific concepts in isolation, concepts are explained in relation to the complex interactions of the physiological subsystems of respiration, phonation, articulation and resonance, and audition. This approach provides a framework for discussing the acoustic nature of speech in relation to the human capacity for producing and perceiving speech. In this way, the scientific concepts relate meaningfully to human communicative behavior.

Although a large part of the focus of this book is on information that can be obtained from instrumentation, there are four very important points that students should keep in mind. First, we cannot assume that there is a direct link between underlying anatomical and physiological factors and acoustic or other instrumental data. We make inferences from the data about the functioning of a system, but they are just that—-inferences. Second, the information obtained from instrumentation may be objective, but a subjective component is always involved in the interpretation of the data. Different researchers, teachers, clinicians, and students may interpret data very differently, depending on prior experience, level of sophistication, and so on. Third is the need to be aware of the validity of the information obtained. A degraded or distorted signal can yield acoustic information that is not valid. Furthermore, some kinds of analyses may not be appropriate for particular disorders. For instance, the jitter measure has been found to have decreased reliability when used with people with severe voice problems. Fourth, acoustic and other instrumental data do not replace behavioral and perceptual information, but supplement it.

It is also important to keep in mind that speech production, although a complex process with many levels, is, in reality, one process. Respiration, phonation, and articulation are so closely interwoven that they really cannot be separated. The divisions and separate discussions of these systems are purely for convenience and ease of discussion. For example, subglottal pressure is discussed in the chapter on voice production, but, in fact, subglottal pressure is also influenced by actions of the supraglottal articulators.

As with any subject, not all topics can be included in the book. The domains included under the broad rubric of speech science are numerous, varied, and complex. An author must necessarily select the topics he or she feels are most important. I have selected those that, over my ten years of teaching speech science, have contributed most to students' understanding of speech science as a whole. The book focuses on acoustic and aerodynamic measure-
ment, rather than on techniques such as videendoscopy, stroboscopy, ultrasound, and magnetic resonance imaging. These are, of course, enormously important technologies that students need to be aware of, but they are beyond the scope of this book. Keeping this limitation in mind, the aim of the book is to provide comprehensive and in-depth coverage of speech science. The book also aims to tell a story, that is, to give context and shape to the narrative in a way that logically develops and furthers student understanding and retention of the material.

Overview of Chapters

The chapters are organized in accordance with the systems approach of this book and with the aim of the book, which is first to present more basic information and then to build on this information in a systematic manner. Chapter 2 presents a framework to help students to understand the basic information about the nature of sound. The first part of the chapter focuses on the physics of sound. The vibratory nature of sound is described and explained in detail, focusing on essential concepts of pressure, elasticity, and inertia. We then explore the properties of sounds that make them different from one another, such as frequency and intensity. An important section of Chapter 2 is the presentation of detailed information about waveforms and spectra and the different sorts of information that they provide. Following the presentation of basic acoustic material, a discussion of resonance is presented, with an emphasis on acoustic resonators. Characteristics of acoustic resonators are explored, such as bandwidth, cutoff frequencies, and resonance curves. These concepts are integral to the understanding of the vocal tract as an acoustic resonator, a topic that is covered in depth in Chapter 8.

Chapter 3 introduces students to clinical applications of the concepts presented in Chapter 2. Frequency and intensity variables that are typically used in clinical situations are discussed, with a clear explanation of their contribution in helping to assess and treat various breakdowns in speech production, as well as their importance in documenting clinical effectiveness. Variables such as speaking fundamental frequency, frequency variability, average vocal amplitude level, and dynamic range are highlighted, and the current use of voice range profiles is explained. The chapter concludes by examining some specific disorders in which the measurement of frequency and intensity variables has been helpful in clinical management of such problems as voice disorders and neurological disorders. Although following logically from the material presented in Chapter 2, some instructors may prefer to defer the information until the discussion of phonation in Chapter 8.

Chapter 4 presents a detailed discussion of the respiratory system. Relevant anatomy and physiology help students to understand how respiration
generates the power supply for speech. An important part of the chapter is the detailed discussion of respiratory volumes and capacities, which students often have difficulty in visualizing, but which are crucial to an understanding of the ways in which we use the air supply for speech and other purposes. Another important section covers the differences between vegetative and speech breathing, an understanding of which is essential in clinical intervention of many speech-related disorders. Breathing patterns for speech are highlighted. The volumes, pressures, and flows that are the basis of speech production are delineated, and instrumental ways of measuring these aspects of respiration are presented, including plethysmographs and pneumotachographs for kinematic respiratory analysis.

Chapter 5 focuses on clinical application of the principles discussed in Chapter 4. Respiratory function is discussed in relation to several types of neurological disorders, such as Parkinson’s disease, cerebellar disease, and cerebral palsy. Other problems are also mentioned in which respiratory function may be an issue, including cervical spinal cord injury, voice disorders, and hearing impairment. Following the organization of Chapter 4, important concepts dealing with volumes, pressures, airflows, and chest-wall shape are integrated into the discussion of the clinical management of respiratory problems.

Chapter 6 focuses on the phonatory system, beginning with anatomical and physiological information. Links with respiration are emphasized, including a discussion of phonation threshold pressure, the minimum amount of subglottal pressure needed to set the vocal folds into vibration. The human voice is described in terms of its nearly periodic nature, leading to a discussion of jitter and shimmer, commonly used instrumental measures of vocal function. A detailed exploration of vocal registers, focusing on the physiological and acoustic aspects of modal, falsetto, and pulse registers, is presented. Comparisons are made thereafter between normal and abnormal vocal qualities, such as breathiness and hoarseness. The acoustic and spectral features of these different qualities are described. Acoustic and other ways of measuring registers and quality are presented, such as harmonics-to-noise ratios and Lx waves obtained from electroglottography.

Chapter 7 applies the theoretical principles presented in Chapter 6 to clinical situations that have been described in the research literature. The many ways in which jitter measurement has been utilized with patients with different neurological disorders are described, including amyotrophic lateral sclerosis and Parkinson’s disease. Other medical uses of jitter analysis include evaluating the effects of endotracheal intubation and patient response to chemotherapy for advanced laryngeal cancer and documenting the effectiveness of behavioral voice therapy for certain types of functional voice problems. Similarly, clinical situations in which electroglottography has been used are presented, for example, in documenting the effects of Botox injection on vocal
Chapter 8 concerns the articulatory system and resonance. The articulators of the vocal tract are described, and their contributions to the shaping of the sound wave are presented. To help students understand the way in which the articulators shape the sound wave and the connections between the respira-
tory, laryngeal, and articulatory systems, the articulators of the vocal tract, as well as the vocal folds of the larynx, are conceptualized as a series of valves that open and close to regulate the flow of air through the vocal tract. This leads logically to a discussion of the traditional classification system of conso-
nants and vowels, based on the manner and location of valving of the pul-
monary airflow through the glottis and/or vocal tract. Following the
exploration of the structures and functions of the various vocal tract articula-
tors, Chapter 8 integrates acoustic and structural information in a discussion of the application of resonance to the vocal tract. The characteristics of the vocal tract resonator are described, such as its being a quarter-wave resonator with multiple resonance frequencies, as well as its variable nature. An explanation follows of how the vocal tract filters the glottal sound wave, formalized in Fant’s source-filter theory of vowel production. The commonly used measure of spectrographic analysis is presented, followed by a detailed explanation of the spectral characteristics of vowels and consonants. To put the sounds in the perspective of connected speech, the concepts of coarticulation and supraseg-
mental aspects of speech production are then discussed.

Chapter 9 focuses on the issue of speech intelligibility, emphasizing disor-
ders that affect intelligibility. Problems in speech production are conceptualized in terms of the source-filter theory, with problems affecting the source function, the transfer function, or both. Acoustic measures of speech output, such as mea-
ures of vowel duration, vowel formant measurements, and spectral analysis of consonants, are shown to be important in determining the precise articulatory movements that contribute to reduced intelligibility in dysarthria, hearing im-
pairment, phonological disorders, tracheotomy, and cleft palate. Physiological
measurements of palatometry and glossometry are also discussed in relation to diagnosis and treatment of intelligibility problems.

The focus in Chapter 10 is on the auditory system. Similarly to the other chapters, information about the structure and function of the different parts of
the ear is presented at the beginning of the chapter. Building on previously dis-
cussed concepts of sound transmission, the role of the middle ear in transduc-
ing air pressure vibrations into mechanical vibrations and transmitting these
vibrations to the inner ear is described. The ability of the cochlea of the inner ear to perform a frequency analysis of incoming sounds is also presented. Follow-
ing the description of hearing, the emphasis turns to the perception of speech, the process of recognizing speech sounds and assigning meaning to
them. All the classes of sounds are described in terms of their acoustic patterns that form the basis for phoneme recognition. Concepts central to issues in speech perception, such as categorical perception, multiple cues in perception, and trading relations between acoustic cues, are integrated into the discussion of sound recognition. Research using synthetic speech is described in terms of its role in defining patterns of speech perception. A discussion of currently used instrumental techniques for the diagnosis and treatment of hearing problems focuses on immittance audiometry, otoacoustic emissions, and cochlear implants. The importance of immittance audiometry and otoacoustic emissions in the screening and evaluating of middle and inner ear function, particularly in infants, young children, and hard-to-test individuals, is emphasized. In particular, otoacoustic emissions are used in neonatal screening to increase the chances of early detection of hearing problems.

Chapter 11 turns to the issue of speech perception, which is discussed with reference to hard-of-hearing and deaf individuals, as well as children with recurrent middle ear infections, language and reading disabilities, and phonological deficits. The link between speech perception and higher level linguistic functioning is explored using the research literature.

Chapter 12 presents a discussion of neuroanatomy and neurophysiology relevant to speech production. The chapter begins with descriptions of nerve cell structure and function and continues with examination of cortical and subcortical areas of the brain. The spinal and cranial nerves are described and related to speech production, and the chapter concludes with a discussion of principles of motor control. Chapter 13 then examines clinical applications of the material. Current brain imaging techniques are described, including computerized tomography, magnetic resonance imaging, functional magnetic resonance imaging, positron emission tomography, single photon emission computed tomography, and evoked potentials. Applications of these technologies to various various communication disorders are discussed, including stuttering, Parkinson’s disease, multiple sclerosis, and Alzheimer’s disease.

Finally, Chapter 14 presents a discussion of the nature of models and theories to help students to understand the importance of conceptual and theoretical frameworks in testing ideas about systems and in predicting the behavior of systems under various conditions. A brief description of selected issues in speech production and the models and theories that have been proposed to explain these issues follows the general discussion. Similarly, selected issues in the area of speech perception are identified, followed by a brief description of some of the models and theories proposed to account for these findings.
Sound occurs when a disturbance creates changes in pressure in a gas, such as air, or in a liquid or solid medium. The disturbance is caused by some kind of movement, such as a cup being placed on a table, a book falling on the floor, a tuning fork being struck, or the human vocal folds opening and closing. The pressure changes created by the disturbance are transmitted through the medium and may end up at a listener’s ear, eventually to be perceived as sound. Because the human sound production and perception systems rely primarily on air, we will focus on sound in air, rather than in solids and liquids. Thus, to understand the nature of sound and the means whereby the human
sound producing and receiving systems work, it is essential to understand the behavior of air.

Air is a gas made up of countless numbers of molecules of various chemicals (oxygen, nitrogen, hydrogen, etc.). These molecules of air are not stationary, but constantly move around in random patterns, and at extremely high speeds. This random movement is called **Brownian motion**. As the molecules move around, these particles of air collide with each other and with whatever is in their path—walls, furniture, people, and so on. These collisions produce pressure.

**Air Pressure**

Pressure is a force that acts perpendicularly on a surface. When you sit on a chair, for example, your body exerts a certain amount of downward force on the horizontal surface of the seat, generating a certain amount of pressure. If you were sitting on a sofa, the pressure you exert would be less, because the force would be spread out over a larger area. The force of any pressure, including air pressure, can move objects. For instance, air pressure acting on a tree moves the branches and leaves. Air pressure causes the hanging objects on wind chimes to collide with each other and create a pleasant sound. In the same way, air pressure acting on an eardrum can push it inward or pull it outward.

**Measurement of Air Pressure**

Pressure can be measured in various ways that incorporate the force exerted and the surface area on which the force is acting. When talking about the pressure needed to move the eardrum, for example, the unit of force is the dyne (d), and the unit of area is the square centimeter (cm²). The dyne measures extremely small amounts of force, so it is suitable for measuring the tiny amounts of pressure acting on the eardrum to produce sound. Larger amounts of force over larger surfaces are measured with larger units, such as pounds per square inch (psi), which is a measure deriving from the traditional English system. The pressure of air in your car tires is probably around 30 psi, a much larger amount of pressure over a much larger surface area than the eardrum.

Dynes per square centimeter and pounds per square inch are commonly used measurements of pressure in the fields of speech pathology and audiology. However, more current measurement systems are based on the modernized metric system, the International System of Units (SI). In this system, pressure can be measured using either the **MKS system** or the **cgs system**. MKS and cgs are the abbreviations for three units of measurement: distance, mass, and time. M stands for meters, K for kilograms, and S for seconds. The letter c stands for centimeters, g for grams,
and s for seconds. These two systems are related, but the cgs system uses smaller units for distance and mass than the MKS system. In the cgs system, the unit of measurement for pressure is dynes per square centimeter (dynes/cm²), also called a **microbar**. One dyne/cm² equals 1 microbar. Force in the MKS system is measured in **newtons**, and pressure is measured in newtons per square meter (N/m²), which is also known as a **pascal**. One N/m² is 1 pascal. This is a very large value, so for speech and hearing applications in which minute amounts of pressure are measured the micropascal (µPa) is used. One µPa is equal to one-millionth of 1 Pa. Most older texts use the dynes measure for pressure, although the current measure of choice among scientists is the pascal (or µPa). Both, however, are perfectly acceptable, and these measures are all equivalent. For example, the pressure of air in the atmosphere at sea level is 14.7 psi in the English system and around 1,000,000 dynes/cm² (1,000,000 microbars) in the cgs system.

Pressure can also be described by the amount of force it takes to move a column of liquid such as water or mercury in a tube. In this kind of situation, the unit of measurement is centimeters of water (cm H₂O). If you blow into a tube that is partially filled with water, the force of your breath exerts pressure on the surface of the water, which will therefore be displaced by a certain distance (see Figure 2.1). If you displace the water by 5 cm, you have exerted

![Figure 2.1](https://example.com/figure2_1.png)

*Measuring air pressure in centimeters of water.*
enough pressure on the water to shift it by 5 cm, and the resulting pressure measurement is 5 cm H₂O.

Sometimes the liquid manipulated in this kind of measurement is mercury. In this case, the unit of measurement is millimeters of mercury (mm Hg). Barometers measuring pressure changes in the atmosphere use mercury. Hence the comment often used by weather forecasters that the barometer is rising (indicating that air pressure is increasing) or the barometer is falling (air pressure is decreasing). These different ways of measuring pressure are useful in describing different areas of the speech production process. Table 2.1 provides easy reference to the measurement terms and their units of measure.

Pressures can be generated in different locations and areas and can increase or decrease, depending on the specific circumstances. It is handy, therefore, to have a way of indicating the location or type of pressure. The scientific notation for pressure is $P$. When combined with a subscript, the type or location of pressure is indicated. For example, atmospheric pressure is written as $P_{\text{atmos}}$. At sea level, $P_{\text{atmos}}$ is around 760 mm Hg, or 14.7 psi, or 1,000,000 dynes/cm². At higher altitudes, $P_{\text{atmos}}$ decreases.

Pressure in different locations can be higher or lower than $P_{\text{atmos}}$. For instance, the pressure of the air inside a tire on your car is typically around 30 psi, considerably higher than $P_{\text{atmos}}$. Pressure that is higher than $P_{\text{atmos}}$ is called positive pressure ($P_{\text{pos}}$). If you puncture your tire, the air leaks out, and pressure could decrease below $P_{\text{atmos}}$, say to 7 psi. Pressure below $P_{\text{atmos}}$ is called negative pressure ($P_{\text{neg}}$). Note that $P_{\text{neg}}$ is not the same as a vacuum. A vacuum refers to a total absence of air, and thus a total absence of pressure. The pressures within various locations within our body, such as our lungs ($P_{\text{alveolar}}$), tracheas ($P_{\text{trach}}$), and mouths ($P_{\text{oral}}$), play an important role in producing and perceiving speech.

### Movement of Air

Because air is a gas, it moves in predictable ways. Molecules of air have a natural tendency to equalize, that is, to spread themselves around more or less
Air Pressure, Volume, and Density

Another characteristic of air is that there is an inverse relationship between air volume and pressure and a proportional relationship between air pressure and density. Volume refers to the amount of space occupied in three dimensions; density refers to the amount of mass per unit of volume. As the volume of a particular enclosed space increases, the pressure of the air within that space decreases as long as the temperature remains constant. As the volume of the enclosed space decreases, the pressure of the air increases, given a constant air temperature. This relationship is known as Boyle’s law, after the scientist who discovered it. Figure 2.2 shows this relationship, using the example of a container with a plunger in it. Each container has an equal amount of air, but the plungers are inserted to different depths, changing the volume. The farther a plunger is inserted, the smaller the volume of the container. However, because the amount of air within the container has not changed, the density of the air within the smaller space is increased. When density is increased, pressure is increased as
well, because the molecules collide with the surfaces and with each other more often and more forcefully.

**Sound: Changes in Air Pressure**

We have seen that air molecules move around, creating a relatively steady pressure. This relatively constant pressure that is around us at any particular place or time is called the ambient pressure ($P_{am}$). The $P_{am}$ in the room where you are working could be the same as or different from the $P_{am}$ in another room, or the $P_{atmos}$. For a sound to be generated, the constant $P_{am}$ must be disturbed in some way so as to increase and decrease in a systematic manner. For changes to occur in the ambient air pressure, some force must disturb the air molecules from their usual random patterns of movement. A tuning fork provides an excellent example of the way in which $P_{am}$ is changed by a disturbance and how this change results in sound (see Figure 2.3).

When you strike a tuning fork, its prongs, or tines, are set into vibration; they move back and forth (oscillate) very rapidly. As the tines vibrate, they set up a chain reaction in the air molecules in adjacent areas. As the tine moves...
outward, it pushes against the air molecules closest to it. These molecules are displaced from their original positions and, in turn, push against their neighboring molecules, which push against their neighbors, and so on. When molecules have been displaced in this manner and are approaching and colliding with the next group of molecules, there is an increased density of air in that area. Increased density results in an increase of pressure. Thus, when molecules approach and collide, an area of positive pressure, known as compression, results.

However, the air molecules that have been displaced do not remain in their new positions, but swing back toward their original positions. As molecules down the line are displaced toward their neighbors, molecules that were earlier displaced are already returning to their original positions. However, the molecules returning toward their original positions overshoot their marks and swing farther away to the other side of their original positions. This results in increased distance between the two groups of molecules involved. There is now decreased density of air in the area between the two groups of molecules, resulting in a lower pressure, known as rarefaction.

This process of increasing and decreasing distances between groups of molecules results in increasing and decreasing amounts of density and the
corresponding increases and decreases in air pressure. The changes in pressure continue in a wavelike motion that travels, or propagates, through the air in an ever-widening sphere (see Figure 2.4).

If an ear is in the path of some of these shifting molecules, the compression of air moves the tympanic membrane (eardrum) inward slightly, whereas the rarefaction of air allows the tympanic membrane to move slightly outward. The tympanic membrane is therefore set into vibration through the changes in air pressure arriving at the ear. The vibration of the tympanic membrane sets bones within the middle ear into vibration. The vibration of these bones in turn sets the fluid in the inner ear into vibration, resulting in the stimulation of hair cells (nerve cells) in the inner ear. The triggering of the nerve cells generates a nerve impulse, which is conducted along the auditory pathway to the appropriate areas of the nervous system and is then interpreted by the brain as sound. Thus, the basic nature of sound consists of alternating increases and decreases in $P_{am}$.

Elasticity and Inertia

When air molecules are set into vibration by a disturbance such as a tuning fork being struck, they vibrate a tiny distance around their rest positions, eventually coming to a stop. Once the molecules have been disturbed, two forces in-
teract to keep them swinging back and forth for a while before they settle down again. These forces are \textbf{elasticity} and \textbf{inertia}. Elasticity is a restoring force; it refers to the property of an object to be able to spring back to its original size, form, location, and shape after being stretched, displaced, or deformed. The amount of restoring force depends on the extent to which the object is displaced. \textit{Hooke's law}, which describes elasticity, states that the restoring force is proportional to the distance of displacement and acts in the opposite direction. Thus, the farther an object is displaced from its original location, the stronger the restoring force that pulls it back toward that position. All solid materials possess some degree of elasticity, and air and other gases behave as though they, too, possess elasticity.

After the fork is initially struck and the molecules have been displaced, they start moving back toward their rest positions due to elasticity. However, they do not immediately stop at their rest positions, but overshoot the mark, swinging out farther in the opposite direction. This overshooting is due to inertia. Inertia is a law of physics describing the tendency of matter to remain at rest or to continue in a fixed direction unless affected by some outside force. In the case of the air molecules, the outside force that overcomes the inertia is their inherent elasticity. In other words, due to inertia the molecules continue to overshoot their original positions until the restoring force of elasticity becomes stronger than the inertia and starts to pull the molecules back toward their resting positions again. Hence, the molecules swing back and forth through their original positions due to the interaction of elasticity and inertia (see Figure 2.5). This is handy, because it means that the original disturbance that is the source of the sound (in our example, the striking of the tuning fork) does not have to be reapplied in order for the sound to continue.

The vibration of the air molecules does not, however, last indefinitely. Because of the frictional resistance of the air, each time the molecules move back and forth around their rest positions, they do so with slightly less \textbf{amplitude}. Amplitude refers to the maximum distance away from rest position that the molecule is displaced, which is determined by the amount of energy involved in the movement. The decrease of amplitude, called \textbf{damping}, thus indicates a decrease in the energy of the sound. (We will discuss the concepts of amplitude and energy in more detail later in the chapter.) Damping finally causes the molecules to settle down once again at their original positions. At this point, no further changes in $P_{am}$ occur, and, consequently, no further sound is generated.

It is important to realize that sources of sound cause changes in air pressure not just in one direction, but in all directions. Figure 2.4 shows how increases and decreases in air pressure radiate outward from the source in all directions. You can easily appreciate this fact by walking away from a person who is talking. Although the sound of his or her voice becomes softer as you
move away, you are still able to hear it, at least for a while, no matter in what direction you move.

Wave Motion of Sound

A wave is a disturbance that moves through a medium. Once molecules have been set into vibration, each molecule (or group of molecules) does not, itself, travel long distances to get to a listener’s ear. Rather, the molecules travel only a tiny distance around their rest positions before their motion dies away. What travels toward the listener’s ear is the disturbance, that is, the changes in air pressure caused by molecular vibration. These compressions and rarefactions
can propagate long distances, and these changes in air pressure affect the listener’s ear. Waves, in general, are characterized by small motions of individual particles of the medium, resulting in changes of the medium being transmitted for long distances.

In the case of sound, the wave motion is longitudinal. The individual air molecules move parallel to the direction that the wave is traveling. Wave motion of water, on the other hand, is transverse; that is, the individual molecules of water move up and down at right angles to the direction that the wave is traveling. We are all familiar with the phenomenon of water rippling outward in widening circles when a stone is dropped into a pond. These ripples are the waves, consisting of water molecules moving up and down and transmitting the disturbance in all directions. In a very similar way, when a sound wave is generated by a tuning fork or other vibrating object, the air all around the fork is disturbed, so the changes in pressure radiate from the fork in all directions. The area of compression around the vibrating source is followed by an area of rarefaction, followed by another area of compression, another of rarefaction, and so on, spreading outward in a sphere. The outermost area of the sphere is called the wave front.

The farther these changes in air pressure travel from the source, the more damped they become, because of the relationship of the wave front and distance. The area of the wave front is directly proportional to the square of its distance from the source. Because the total energy of a wave is constant, this means that the amplitude of the wave decreases as it gets farther from the source. In a large lecture hall, for example, the people closest to the speaker hear him or her most loudly, and the farther from the speaker an individual is sitting, the less loudly will the speaker be heard.

Characteristics of Sound Waves

Sound waves are characterized by many different aspects, such as their frequency and period, amplitude, velocity, and wavelength. Many characteristics of a sound wave can be shown on a waveform, which is a graph with time on the horizontal axis and amplitude on the vertical axis. This graph can be used to represent movement over time. For instance, the movement of the tines of the tuning fork can be shown, as can molecular movement generated by the tuning fork. A waveform also shows the corresponding increases and decreases in air pressure. See Figures 2.6 and 2.7.

Frequency and Period  One back and forth movement of the molecule makes up one cycle of vibration. In other words, one cycle of vibration occurs when the molecule moves to a maximum distance away from its original spot, back toward rest position, moves to a maximal point in the opposite direction, and then back again to rest position. Cycles of
vibration, however, are typically thought of in terms of pressure changes, rather than in terms of individual movements of molecules. Acoustically, a cycle of vibration consists of an increase in pressure from \( P_{am} \) (compression), a decrease in pressure to \( P_{am} \), a further decrease in pressure below \( P_{am} \) (rarefaction), and a return to baseline \( P_{am} \). Cycles of vibration are measured in terms of time, typically in seconds. The tines of a tuning fork might vibrate at the rate of 100 cycles per second (cps), causing the surrounding air molecules to vibrate at a rate of 100 cps as well. If this vibration eventually reaches a listener, that person's eardrum will vibrate at 100 cycles per second. The number of cycles per second at which objects (or air) vibrate is called frequency, and the unit of measurement of frequency is hertz (abbreviated Hz). Thus, a tuning fork vibrating at 100 cycles per second has a frequency of 100 Hz. The sound wave produced by the tuning fork, correspondingly, also has a frequency of 100 Hz, and the eardrum in the path of this 100 Hz sound wave would be set into vibration at 100 Hz. Frequency can also be expressed in terms of kilohertz (KHz); thus, 1000 Hz equals 1 kHz, and 2500 Hz equals 2.5 kHz.

We said that the frequency of a sound refers to the rate at which the source and the air molecules vibrate. We can turn this around and think of this rate of vibration in terms of the time it takes for one cycle of

---

**FIGURE 2.6**

*Time movement, molecular movement, and changes in air pressure can all be shown on a waveform.*
Waveforms showing pure tones of different frequencies, amplitudes, and damping.
vibration to occur. For instance, the wave with a frequency of 100 Hz has 100 cycles that occur in a second. Assuming that each cycle in the wave lasts for the same amount of time, we can see that each must take 1/100 second (0.01 second). The time that each cycle takes to occur is referred to as the **period** of the wave, symbolized as $t$. A wave with a frequency of 250 Hz has a period of 1/250 second, or 0.004 second ($t = 0.004 \text{ s}$). Similarly, a wave with a period of 0.002 second (1/500 second) has a frequency of 500 Hz. From these examples you can see that there is a reciprocal relationship between frequency and period. This relationship is expressed by the formula $F = \frac{1}{t}$, where $F$ is frequency and $t$ equals period. If you know the frequency of a wave, you can figure out its period by putting a 1 over the frequency; if you know the period of the wave, you figure out its frequency by putting a 1 over the period (see Figure 2.8).

A wave in which every cycle takes the same amount of time to occur as every other cycle, and in which the extent of the pressure changes (i.e., the amplitude) is equal for all cycles is said to be **periodic**. Perceptually, such a wave would have a musical tone. For example, the vibrating string of a guitar or violin produces a periodic sound wave with a musical tone. However, not all sound waves have cycles lasting the same amount of time. A sound wave might have one cycle that lasts for 0.002 second, the next might take 0.003 second, the following 0.001 second, and so on. A wave in which individual cycles do not take the same amount of time to occur is called **aperiodic**. Perceptually, such a wave sounds like noise. If you

---

**Figure 2.8**

Frequency, period, and wavelength.
clap your hands or hiss through your teeth, you are producing an aperiodic sound wave, or noise.

**Velocity and Wavelength** Sounds are physical phenomena, and they therefore obey physical laws, such as those governing motion and speed or the velocity of objects traveling through a medium. How fast the wave moves depends on the density and elastic properties of the medium through which it is moving. For instance, because water is more dense than air, sound travels about four times as quickly through water as it does through air and even faster than that in some solids such as steel. The speed of sound in air is around 331 m/s at 0°C (32°F), compared to 1461 m/s in water at 19°C. In a steel rod, sound may travel at a speed of 5000 m/s (Dull, Metcalfe, & Williams, 1960).

The speed of sound in liquids and solids is not affected much by temperature. In gases such as air, however, temperature plays an important part in how fast the sound travels. The warmer the air, the more quickly sound is transmitted. In fact, the speed of sound increases at the rate of about 0.6 m/s/°C (Dull et al., 1960). For example, the speed of sound in air at 0°C is 331 m/s, whereas at 20°C (68°F) it is 343 m/s (Durrant & Lovrinic, 1995).

Thus, two aspects of sound are related to time: period and speed. Period is related to the frequency of vibration, which in turn depends on the physical characteristics of the source. In general, the larger and more massive the source, the more slowly it will vibrate, and vice versa. Speed of sound, on the other hand, depends not on the frequency of the sound, but on the characteristics of the medium.

Sound waves not only occur at a certain frequency and are transmitted at a certain speed, but they also travel through space. The measurement of the travel of a sound wave is its **wavelength**. Wavelength refers to the distance in meters or centimeters covered by one complete cycle of pressure change. The wavelength of a sound is measured as the distance covered by the wave from any starting point to the same point on the next cycle. Frequency, period, and wavelength are closely related. The higher the frequency (the more cycles per second), the shorter in duration is the period and the shorter is the wavelength. The lower the frequency (the fewer cycles per second), the longer in duration is the period and the longer is the wavelength.

**Sound Absorption and Reflection** So far in our discussion we have talked about sound being transmitted through the air, without taking into account any objects or boundaries that sound waves might encounter in their travels. In fact, when sound waves come into contact with walls, ceilings, floors, and the like, they may or may not be transmitted through these boundaries. A sound wave that is generated, travels a certain distance, and then hits up against a boundary is called an **incident wave**. Incident waves may be transmitted, absorbed, or reflected.
In the case of a room with thick concrete walls, only a small amount of sound energy from the incident wave may be transmitted through the wall, whereas in a thin-walled room a lot of the incident wave’s energy will probably be propagated through the wall. If all the energy in a sound wave is not transmitted, some portion of the sound not transmitted may be absorbed and some may be reflected.

Absorption is basically the damping of a wave, with diminishing changes in air pressure due to friction. Materials differ in the amount of sound energy they absorb. Typically, materials that are hard or dense and/or have smooth surfaces do not absorb much of the energy of the sound waves coming into contact with them. Materials that are soft and porous and/or have rough surfaces absorb a lot of the sound energy. Different materials, therefore, are used for different acoustic purposes. For instance, a special material, acoustic tile, is often put on ceilings specifically to absorb sound and prevent it from being transmitted. Often walls and ceilings in speech and hearing laboratories are acoustically treated in this way to absorb sound and reduce noise from the outside environment.

In reflection, some portion of the sound that is not transmitted or absorbed bounces back from the surface of the boundary and travels in the opposite direction of the incident wave. Similarly to absorption, the amount of reflection depends on the type of surface. A hard, smooth surface will reflect more sound than a soft or rough surface. This phenomenon is very similar to how a mirror works. A mirror is specially treated to have a hard, smooth surface. It therefore does not transmit or absorb light waves, but reflects them back into the environment, allowing the image to be seen.

Constructive and Destructive Interference In general, physical objects cannot occupy the same space at the same time. Sound waves, however, can, because areas of high and low pressures can combine. Suppose that a tuning fork tuned to 100 Hz produces an incident sound wave that travels through the air, comes up against a wall, and is then reflected back toward the tuning fork. The tuning fork, in the meantime, continues to generate new incident waves 100 times per second. Thus, incident and reflected waves are combining with each other at any instant in time and space. This combining of waves is known as interference. The air pressure changes forming these two waves can interfere with each other in various ways. If the areas of compression and rarefaction of the two waves combine at exactly the same time and the same moment in space, the amplitude of the resulting wave will be doubled. This happens because when two areas of high pressure combine the resulting pressure is higher still. When two areas of low pressure combine, the pressure is further low-
ered at that point. This produces greater deviations from normal $P_{am}$ and therefore increased amplitude of the wave. Interference that results in increased amplitude is called constructive interference (see Figure 2.9).

If the areas of compression of one of the waves combine at exactly the same time with an area of rarefaction of the other wave, the amplitude of the resulting wave will be decreased, which is known as destructive interference.

Theoretically, it is possible for two sound waves with the same frequency to combine such that each compression of one wave is matched exactly with the corresponding rarefaction of the second wave; the resulting sound will be completely damped. This happens when an area of high pressure combines with an area of low pressure, causing the air molecules to equalize themselves in such a way that an area of normal pressure ($P_{am}$) results. Because sound consists of changes in air pressure, normal $P_{am}$, by definition, cannot be a sound.

Waves of different frequencies can also combine, and their areas of compression and rarefaction will not line up exactly. The amplitude of the resulting wave will therefore not be doubled, nor will the sound be completely eliminated. Rather, the amplitude of the sound will be changed in complex ways. This relative timing of areas of high and low pressure in waves is called phase.

Sound waves that combine and interfere with each other can affect the way that you perceive the sound. For instance, a sound can experience reverberation, meaning that it lasts slightly longer because of the interference. This happens when a reflected sound wave arrives at your ear slightly delayed in time compared with the arrival of the incident wave at the same point. The duration of the delay depends on the

\[ \text{F I G U R E 2.9} \]

Constructive and destructive interference.
distance between the reflective surface and your ear. The sound wave must travel to the reflective boundary and return to your ear, delaying its arrival by a fraction of a second. You only hear one sound, but this sound is extended in duration because of the reflected wave arriving in time to keep your eardrum vibrating longer than it would with just the incident wave vibrating it. However, if the distance between your ear and the reflective surface is substantial, then the delay will be perceptually noticeable. In fact, in this case the reflected sound wave is heard as a separate sound, in other words, an echo. With an echo, the incident wave vibrates your eardrum, but your eardrum has time to settle down and stop vibrating before it is set into vibration once again by the reflected wave.

Reverberation can be desirable because it can increase the intensity of the incident sound reaching a listener. However, too much reverberation can interfere with communication by making the phonemes blend together and become garbled. This is an issue that has become very prominent over the last few years in educational settings. Many classrooms are overly reverberant because of the uncarpeted floor and bare walls that create multiple reflections of sounds. The more reflective the room, the longer the incident sound and reflections take to be absorbed and damp. This can hinder students in their understanding of the instructor’s speech, which is particularly problematic for individuals with hearing disorders or other types of learning difficulties.

**Pure Tones and Complex Waves** When molecules of air are set into vibration by an object such as a tuning fork, the way in which they vibrate has a certain regular, predictable pattern. Pendulums on grandfather clocks and playground swings move in the same way. We will use the example of a swing to illustrate this motion. When you push a swing, it moves away from you to some maximum distance. As it approaches this maximum point, it slows down and stops for an instant. It then reverses direction and starts moving back toward you. As it does so, its speed increases, and it reaches its maximum speed as it passes over its rest position. As it continues to move toward you, its speed decreases until the swing stops for an instant, and then, once again, it reverses direction. This regular, smooth, back and forth movement with its characteristic pattern of acceleration through the rest position and deceleration at the endpoints of the movement is called **simple harmonic motion** (SHM). Simple harmonic motion of an object such as a tuning fork generates the same smooth and even pattern of vibration of the air molecules around it, which in turn vibrate the tympanic membrane in simple harmonic motion. An object vibrating in SHM produces a sound wave that has only one frequency, called a **pure tone**. Perceptually, such a sound is heard as a rather thin, clear tone.

Other kinds of sounds, however, are characterized by waves that consist of more than one frequency, called complex waves (see Figure 2.10). Complex...
Sinusoids of different frequencies combine to form a complex wave. The F0 of the complex wave is the same as the lowest frequency sinusoidal wave.

**FIGURE 2.10**

*Pure tones combining to form a complex wave.*
waves are much more common than pure tones and occur when sounds of different frequencies combine and interfere with each other in various ways. The interference results in a more complex vibration of the air molecules. One can imagine air molecules vibrating in SHM as walking back and forth in a regular even manner, whereas the molecules in a complex wave do a more complex “dance” involving all kinds of different movements as they vibrate around their rest positions. Both walking and dancing are patterned movements, but the patterns are more complicated and varied in dancing than in walking. Corresponding to the more complex molecular movement in complex sound waves, the tympanic membrane also vibrates in a more complex manner.

A complex sound is defined as a wave consisting of two or more frequencies. There are two types of complex sounds, periodic and aperiodic. Periodic complex sounds consist of a series of frequencies that are systematically related to each other. The lowest frequency of the sound is the \textit{fundamental frequency} ($F_0$), and the frequencies above the fundamental are called \textit{harmonic frequencies}, or just harmonics. The harmonics in a complex periodic sound are whole-number multiples of the fundamental frequency. For example, if the $F_0$ of a complex periodic wave is 100 Hz, the harmonics will be 200 Hz, 300 Hz, 400 Hz, 500 Hz, and so on. A complex periodic wave with an $F_0$ of 300 Hz will have harmonics of 600 Hz, 900 Hz, 1200 Hz, and so on. A complex periodic wave has a musical tone and sounds richer and more resonant than a pure tone wave. In fact, the more harmonics in a sound wave, the more resonant it will sound, and vice versa. Most musical instruments produce sounds that are periodic and complex.

The harmonics in a complex periodic sound can be identified through a process called \textit{Fourier analysis}. Jean-Baptiste Fourier (1768–1830) was a French mathematician who showed that any complex wave can be represented by the sum of its component frequencies as well as their amplitudes and phases. Fourier analysis is a highly complex mathematical procedure, which is typically performed these days by computer.

Aperiodic complex sounds also consist of two or more frequencies, but the frequencies in this kind of sound are not systematically related to each other. Rather, a broad range of frequencies make up the sound. For example, an aperiodic complex sound could contain all frequencies between 100 and 5000 Hz. Another aperiodic sound might include frequencies from 2000 to 4000 Hz. Such waves sound like noise, with no musical tone, such as steam escaping from a radiator or the sound of applause. There are two kinds of aperiodic complex sounds, differentiated on the basis of their duration. \textit{Continuous} sounds are able to be prolonged, whereas \textit{transient} sounds are extremely brief in duration. The steam hissing out from the radiator is continuous, whereas the sound made by a person hitting his or her hand on a desk is transient.
Speech as a Stream of Complex Periodic and Aperiodic Waves

We are now in a position to understand the basic acoustic nature of speech. When we produce speech sounds, we are actually producing different kinds of complex periodic and complex aperiodic sounds. All human sounds are complex, due to the nature of the source of the sounds. (This will be discussed in detail in later chapters.) Vowels are complex periodic sounds that have a musical tone; voiceless consonants are complex aperiodic sounds that are either transient (such as the stops /p/, /t/, and /k/) or continuous (such as the fricatives /f/, /s/, and /h/) and that sound like noise. Voiced stop and fricative consonants are a combination of periodic and aperiodic complex sound waves.

Visually Depicting Sound Waves: Waveforms and Spectra

Sound waves are invisible and intangible; the pressure changes are miniscule and cannot be seen. This could be a problem in understanding and working with sound, as sound is, by its nature, fleeting and insubstantial. Fortunately, there are graphic ways of representing sound waves that are extremely useful in helping to visualize the nature and characteristics of various sounds. One type of visual display is the waveform, and another is called a spectrum.

Waveforms, as we have seen, are graphs that show time along the horizontal axis and amplitude along the vertical axis. The amplitude can represent the amount of whatever is being graphed. For instance, if a pen were attached to the tine of a tuning fork in such a way that as the tine moved the pen moved with it and traced a line on graph paper, the vertical axis of the resulting waveform would represent the distance that the tine vibrated around its rest position, and the horizontal axis would represent the time that the tine was moving. Now imagine that we attach a pen to an individual molecule so that when the molecule vibrated around its rest position a similar waveform would result. Now the waveform does not depict the motion of the tine of the tuning fork, but the amplitude of the motion of the molecule of air over time. When dealing with sound, typically what is represented on the waveform is not the tuning fork or individual molecule motion, but the changes in air pressure that result from molecular motion. So an acoustic waveform shows the amplitude of air pressure changes over time.

If we drew a line at around the midlevel of the graph, it would represent normal $P_{am}$, or baseline pressure. When the line goes above baseline, it represents an increase in pressure, that is, compression, and the height of the line at any point represents the amount, or magnitude, of increase. Similarly, when the line goes below baseline, it represents a decrease in pressure, or rarefaction, and again the depth of the line at any point represents the magnitude of decrease.

A waveform is useful in showing many different aspects of a sound. For instance, by counting the peaks in the waveform, we can calculate the frequency
of the wave. Or, by measuring the time of each wave cycle, we can tell the period of the wave. Also, because the vertical axis measures the magnitude of pressure changes, it is easy to visualize the relative amplitude of the wave. Figure 2.7(d), for example, shows a wave that is damping.

Shape is another aspect of sound that is visible from a waveform (see Figure 2.11). A smoothly varying shape, a **sinusoid**, tells us that the wave is a pure tone, vibrating in SHM. If all the cycles in the wave repeat themselves in
a predictable fashion, the wave is periodic. If the cycles look different and take different amounts of time to occur, the wave is aperiodic. If the amplitude of the wave is decreasing over time, then the sound is damping. If the cycles in the wave repeat themselves in a regular way, but the shape of the wave is not sinusoidal (the cycles look more irregular in shape), then the wave is depicting a periodic complex sound. In this case, by counting the largest peaks we are able to determine the $F_0$ of the complex periodic sound. What cannot be seen from a waveform, however, are the harmonics of a complex sound.

To visualize harmonics, we need another kind of graph, a **line spectrum** (see Figure 2.12). A line spectrum also has a horizontal and vertical axis, but in this case the horizontal axis represents frequency, starting with low frequencies to the left, with frequency increasing to the right. The vertical axis represents amplitude, but rather than depicting amplitude of pressure changes, the amplitude in a line spectrum shows the amount of acoustic energy at each harmonic frequency of the sound. Each frequency in the wave, including the $F_0$, is represented by a vertical line, the height of which shows the amplitude of that specific frequency.

What is not evident in this kind of spectrum is time. The frequencies shown on a line spectrum are those that are present in a sound at one particular instant of time. Think of the difference between waveforms and spectra (singular: spectrum) in terms of a layer cake. A waveform would correspond to the entire uncut cake: You could determine the overall shape, the color of the frosting, the height of the cake, and the like, but you could not tell, without cutting the cake, what the inside was like. The line spectrum is analogous to one slice of cake. You can judge the internal section of the slice: its color, how many layers, whether there is jam or frosting between the layers, and so on, but from just the one slice, you could not tell about the overall shape of the cake.

You could also think of a spectrum as a snapshot of a person at one particular instant of time; a waveform corresponds more to a video, in which changes over time can be seen as the person walks around, performs different actions, and the like. A line spectrum can show equally well whether a sound is a pure tone (in which case it would have just one line) or a complex sound (more than one line). However, we would not be able to judge if the sound were damping or changing in other ways over time, because the overall amplitude of the sound over time is not shown.

A line spectrum is not used to represent complex aperiodic sounds, because these sounds are characterized by broad bands of frequencies. Rather than drawing individual vertical lines extremely close to each other, as would be appropriate for aperiodic complex sounds, we draw what is called the **envelope** of the wave as a horizontal line that is understood to connect all the component frequencies in the sound. This kind of spectrum is known as a **continuous spectrum**. As with periodic complex sounds on a line spectrum, the height of the
line at any frequency represents the amount of acoustic energy at that frequency. What cannot be seen from a continuous spectrum is the duration of the sound, so we cannot tell if the sound is continuous or transient. Figure 2.13 shows a waveform and the
Waveform of a transient aperiodic sound and corresponding spectrum

Waveform of a continuous aperiodic sound and corresponding spectrum

**Figure 2.13**
Waveform of a transient aperiodic sound and corresponding spectrum.
corresponding spectrum of a transient and a continuous complex aperiodic sound wave.

It is very important to distinguish between waveforms and spectra, because the information they provide is very different. Figure 2.14 shows that a flat horizontal line on a waveform depicts silence, because the pressure is constant over time. However, a flat horizontal line on a spectrum would depict an aperiodic complex sound containing all frequencies in a certain range.

**Attributes of Sounds**

Although all sounds are generated by changes in air pressure, individual sounds are very different from one another. A high-pitched siren sounds very different from a low-pitched fog horn. A discreet whisper during class sounds different from a football cheer. The husky voice of a person with laryngitis sounds very different from that of a trained singer. We shall now explore the attributes of sounds that result in these kinds of differences, including frequency and pitch and amplitude, intensity, and loudness.
Frequency and Pitch

_Frequency_ refers to the rate at which an object vibrates and is measured in hertz (Hz). Frequency is an objective measurement of a physical phenomenon. _Pitch_, on the other hand, is a psychological event. Pitch is how we perceive the sensation of sound as being high or low on a musical scale. You could say that pitch is the perceptual counterpart of frequency. Whereas frequency is measured in hertz, pitch is measured in _mels_. The mel scale is a perceptual, subjective scale that was constructed by having people subjectively decide whether tones of different pitches were higher or lower than others. A tone of 1000 Hz was selected as the standard tone and was called 1000 mels. A tone that was subjectively determined to be twice as high as this standard tone was called 2000 mels, one that was determined to be half as low was called 500 mels, and so on. Frequency and pitch are related. The rate at which an object vibrates determines how high or low the sound is perceived. In general, the faster the rate of vibration (the higher the frequency) is, the higher pitched the sound will be perceived. The slower the rate of vibration (the lower the frequency) is, the lower pitched the sound will be perceived. However, this relationship is not linear, as we will see later.

The frequency of a vibrating object depends on its physical characteristics, such as its overall size, length, thickness, and density; the material of which it is made and its stiffness; and so on. Typically, the larger the object or the more massive it is, the more slowly it will vibrate. A larger tuning fork with thicker tines will vibrate more slowly than a smaller fork with thinner tines. There are ways, however, of changing the rate at which an object vibrates. The three most important determinants of frequency are the length of the vibrating object, the mass of the object, and the tension of the object. If any one of these three characteristics is changed, the frequency will change in predictable ways. In terms of length, the longer the vibrating source, the more slowly it vibrates. Therefore, if we increase the length of whatever is vibrating, its frequency will decrease. For example, imagine that the chains of a swing hanging from a frame are a certain length so that, when the swing is pushed, it will swing back and forth (i.e., one cycle of vibration) exactly twice per second, resulting in a frequency of 2 Hz. By inserting more links in the chains, the length is increased, and the swing will move back and forth more slowly, perhaps at 1.5 Hz. Conversely, by removing links from the chain, the length is decreased and the swing now speeds up to, say, 2.5 Hz.

Mass is also important in determining the frequency at which an object will vibrate. The more massive an object, the more slowly it vibrates, and vice versa. For instance, the swing in our example vibrates at 2 Hz when no one is sitting on it. When a child sits on the swing, this adds mass, and the swing now moves back and forth more slowly, perhaps only once per second (1 Hz). The
greater the mass is of the person sitting on the swing, the more the decrease in the rate of vibration. In the same way, decreasing the mass of a vibrating object will increase its rate of vibration.

Finally, the tension, or stiffness, of a body plays an important part in determining its frequency of vibration. The more tense or stiff, the quicker is the rate of vibration. The looser, or more relaxed the object, the slower is the rate of vibration. This principle can be easily seen with a rubber band. If a rubber band is held loosely at either end and plucked, it will vibrate at a certain rate. If the band is stretched more tightly, it becomes stiffer and more tense, and its frequency increases when it is plucked. Note, however, that when the elastic band was stretched it became longer, and, as we have just seen, objects that are longer generally vibrate more slowly than shorter ones. But, although the elastic band is longer when it is stretched, its mass per unit of area has been decreased, so although it is longer, it is now skinnier as well. In other words, there is the same amount of material in the elastic band, but it has been stretched out over a greater distance, and its cross-sectional area is less. The increased tension on the band and its decreased mass per unit of area cause it to vibrate more quickly. This interaction between length, mass, and tension to determine frequency is extremely important in the production of voice, as we will see in later chapters.

**Human Range of Hearing**  
The range of frequencies that humans are capable of perceiving is around 20 to 20,000 Hz. Frequencies below this range are called **subsonic**, and frequencies above this range are **supersonic**. This range is really very limited. Some animals can perceive sounds well below 20 Hz, such as elephants, who make use of these subsonic sounds in their communication. Some birds, such as pigeons and chickens, can also perceive exceedingly low-pitched sounds that would be inaudible to humans (Gill, 1995). On the other hand, dogs can hear sounds well above 20,000 Hz, which is why dog whistles actually work: The whistle emits an extremely high-frequency sound that is inaudible to humans, but audible to dogs. Bats, too, are well known for their use of very high-frequency sounds to locate objects in space. Humans tend to hear best those frequencies that are in the middle of their range, around 1000 Hz (1 kHz) to 4000 Hz (4 kHz). Sounds that are above or below this midrange are not as easily perceived by the human auditory system. Conveniently, most speech sounds fall within this range.

**Amplitude and Intensity**

**Amplitude**  
The term *amplitude* refers to either the amount of motion of a vibrating object or the amount of pressure change generated by the motion of the object. With reference to molecular movement, amplitude refers to the dis-
tance that the molecules are displaced from their rest positions during vibration. When talking about sound, it is more common for amplitude to refer to the measurement of the pressure changes that constitute sound. Like frequency, amplitude is a measurement of a physical phenomenon—the pressure changes that occur in the air as sound is propagated through it. Because amplitude is related to pressure, it is typically measured in dynes per square centimeter (dyne/cm²) in the cgs system or micropascals (µPa) in the MKS system.

**Intensity**

To understand intensity, we first need to appreciate the concepts of energy, work, and power. Scientifically speaking, energy is the capacity of an individual or an object to perform work. Work is defined as a push or a pull that moves an object a certain distance. Energy and work, although related, are not identical. If you push as hard as you can against a building, you have probably expended energy, but no work has been done (unless, of course, you have actually managed to move the building). Energy is measured in ergs in the cgs system or joules (J) in the MKS system. One erg is the amount of work done when a force of 1 dyne displaces an object by 1 centimeter (Denes & Pinson, 1993). One joule is a force of 1 newton acting through a distance of 1 meter.

Power refers to the amount of energy expended in a given time and is measured in watts (W). One watt equals 1 joule per second or 10 million ergs per second. Power is easy to understand in terms of your own level of energy and how quickly you expend it. If you take a leisurely bike ride, you expend energy at a certain rate. Perhaps you can go on for several hours before fatiguing. If you are practicing for a race and you ride as fast and hard as you can, you may expend the same amount of energy in a much shorter amount of time. Even though you used up the same amount of energy for each ride, the powers would be different because of the different rates at which you expended the energy.

Intensity refers to power (i.e., the amount of energy expended in a second) measured over a particular area, usually square meters or square centimeters. Thus, the unit of measurement of intensity is watts per square centimeter (W/cm²). You can think of intensity as the amount of energy or power required to generate a certain output, whether the output is sound or some other form of energy, such as light. For instance, light bulbs are available that produce different intensities of output (with light, the output is measured in units called lumens). A light bulb that uses less energy produces a less intense light than one that uses more energy. A 40 W light bulb uses 40 W to generate a certain light output. Such a light bulb gives off a less bright light than a 75 W light bulb. The same is true of sound intensity. If your stereo speakers use 10 W of power, you will hear a certain loudness of sound. If you have really
large speakers designed to produce 30 W, the sound output will be much greater.

It is clear that there must be some kind of relationship between amplitude and intensity, which are physical measurements, and loudness, which is how your ears perceive the intensity generated by the degree of pressure change. The greater the amplitude and intensity, the louder the sound that is heard, and vice versa. Similar to frequency, this relationship is not linear. A sound must be 10 times more intense before it becomes twice as loud, but it must be 100 times as intense before it becomes three times as loud (Dull et al., 1960). A perceptual scale for loudness was developed in a similar way to the mel scale for pitch. A 1000 Hz tone of a particular intensity served as the reference for loudness. The scale that resulted from matching sounds of varying frequencies and intensities to the 1000 Hz tone is called the **phon** scale.

Amplitude and intensity (sound power) are related to each other. The greater the amplitude, the greater the intensity that is generated in the sound wave. However, intensity increases much more rapidly than does amplitude. Mathematically, intensity is the square of amplitude. In other words, if a sound wave has a certain amplitude with a corresponding intensity level and the amplitude of that sound wave is doubled, the intensity of the wave will not be doubled; it will be increased by the square of the increase in amplitude, so the intensity will be quadrupled (increase of $2^2$). If the amplitude is increased by a factor of 5, the intensity will be increased by a factor of $5^2$ (i.e., 25). Thus, it takes only small increases in air pressure changes to generate much larger increases in sound intensity.

**Decibel Scale** The **decibel (dB) scale** is designed to measure sounds in a way that takes into account the amplitudes and intensities of sounds in relation to how we perceive sounds. The decibel scale is named after Alexander Graham Bell. It is abbreviated with a lowercase d, which stands for deci, or one-tenth, and an uppercase B, for Bell. A decibel is thus one-tenth of a bel, which is a large unit of measurement of sound intensity based on the logarithm of a ratio.

The human auditory system is sensitive to an enormous range of intensity levels. From the softest sound that a person can hear to the loudest sound, which produces a feeling of pain in the ears, is a range of around 1 trillion intensities. Trying to deal with huge numbers like this on a **linear scale** would be unwieldy and confusing. The decibel scale is **logarithmic**, which has the effect of compressing the trillion intensities into a scale with far fewer levels. This compression occurs because of the essential difference between linear and logarithmic scales. A **linear scale** is one in which units are the same distance from each other, and units can be added or subtracted (see Figure 2.15). For example, a ruler is a linear measure. A ruler has distances marked off on it, typically in inches (in.), centimeters (cm), and millimeters (mm). A distance of, say, 10 mm between...
Each step in a linear scale represents an equal interval or increase.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

Each step in this base 10 logarithmic scale represents an increase of ten times over the previous number.

\[
\begin{align*}
10^1 &= 10 \\
10^2 &= 100 \\
10^3 &= 1,000 \\
10^4 &= 10,000 \\
10^5 &= 100,000 \\
10^6 &= 1,000,000
\end{align*}
\]

**Figure 2.15**

Linear versus logarithmic scale.

Two points on the ruler is ten times greater than a distance of 1 mm. Put another way, the distance between 2 and 3 in. is exactly the same as the distance between 8 and 9 in., which is exactly the same as the distance between 4 and 5 in. Thus, successive units are always the same distance from each other. Temperature scales are also examples of linear scales, with the distance between units indicating equal increments of temperature increase or decrease.

Logarithmic scales, on the other hand, contain units that increase by greater and greater amounts as we go up the scale. These units cannot be added or subtracted because they are not equal. A logarithmic scale has several components. The first is a base, such as 2 or 10. A logarithmic scale with a base of 2 indicates that each successive unit increases by a factor of 2. With a base of 10, each unit increases by a factor of 10. The number of increases on a logarithmic scale is indicated by the exponent, or power (not to be confused with power as it relates to energy). A base is raised to some power. A base of 10 raised to the first power, or with an exponent of 1, is \(10^1\). 10 multiplied by 1 is 10. A base of 10 raised to the second power equals \(10^2\). In this case you multiply the base by itself 2 times. 10 times 10 is 100. Similarly, 10 raised to the third power is \(10^3\). Multiply base 10 by itself 3 times, and you get 10 times 10 times 10, equaling 1000. Particularly with base 10, the exponents are often used by themselves with the understanding that this is the power to which 10 must be raised in order to equal the original number. This exponent is then called the log. For example, the log of 100 is
2, the log of 1000 is 3, the log of 10,000 is 4, and so on. Thus, whereas linear scales increase successively by some equal amount, logarithmic scales increase by successively greater amounts. Therefore, a huge number of units of intensity (a trillion or so) on a linear scale becomes condensed on a logarithmic scale to around 140 units (Durrant & Lovrinic, 1995).

Aside from being a logarithmic scale, the decibel scale is also a ratio scale. A ratio reflects a relationship between quantities. For example, a ratio of males to females of 2:1 indicates that there are two males for every female. The ratio that the decibel scale measures is the relationship between the amplitudes or the intensities of two sounds. Why is it necessary to make these kinds of comparisons? Because amplitude and intensity are physical measures, we could measure the precise amplitude or the precise intensity of any sound. However, this would not give us a particularly meaningful value. Say that a sound were measured to have an amplitude of 0.045 dyne/cm². This is not very revealing information, in terms of how loud you actually perceive the sound to be. We need some way of comparing this sound to a sound that has a known amplitude and intensity level so that we can tell if it is higher or lower in intensity. In other words, we need a standard sound that can serve as the basis of comparison for all other sounds. This is the basis for the decibel scale.

The decibel scale is a ratio scale; it compares any target sound with a standard reference sound. This standard reference sound has a specific amplitude of 20 µPa (0.0002 dyne/cm², or 0.0002 microbar, µbar) and a specific intensity of 10⁻¹² W/m² (10⁻¹⁶ W/cm²). In the perceptual domain, a sound with this amplitude and intensity indicates the softest sound of a particular frequency that a pair of normal human ears can hear 50 percent of the time under ideal listening conditions. Perceptually, this is known as the threshold of hearing. (Baken [1996] pointed out that the human threshold of hearing at 0.0002 µbar is equivalent to 0.000000204 cm H₂O, demonstrating the extraordinary sensitivity of the human auditory system.) On the decibel scale, a sound of this level is indicated by 0. Keeping in mind that the decibel scale is a ratio scale, 0 dB does not mean that there is silence, but that the sound in question has the same intensity and amplitude as the standard reference sound. It is also possible to have a negative number on the decibel scale; this just means that the sound in question has an amplitude and intensity less than that of the standard reference sound.

The formula to derive the intensity or amplitude of a target sound (I₁ or P₁) in relation to a standard reference intensity or amplitude (I₀ or P₀) is based on the logarithm of a ratio, or the bel. The formula for intensity is \( N(\text{bels}) = \log_{10} \frac{I_1}{I_0} \). This means that the number of bels is equal to a logarithmic scale of base 10 on which the target intensity level (I₁) is divided by the reference intensity level (I₀). However, because the bel is a large unit, most applications in acoustics use the deci-
bel, which, as noted previously, is one-tenth of a bel. Thus, the formula for intensity in decibels (dB) is \(N(dB) = 10 \log_{10} \frac{I_1}{I_0}\). Say, for example, the target intensity \(I_1\) is 100 times the reference intensity \(I_0\). We know that the log of 100 is 2. Multiplying 2 by 10 gives 20, so a sound that is 100 times as intense as the reference sound has a value of 20 dB. In the same way, a sound that is 1000 times as intense as the reference sound will have a dB value of 30 (the log of 1000 is 3; multiply 3 by 10), a sound that is 10,000 times as intense as the reference will be equal to 40 dB, and so on. The standard reference for intensity is \(10^{-12} \text{ W/m}^2\). Substituting this value for \(I_0\), intensity level is expressed as the formula: \(IL(dB) = 10 \log_{10} \frac{I_1}{10^{-12}} \text{ W/m}^2\).

A similar formula is used for amplitude, or sound pressure level (SPL). The only difference is that instead of multiplying the log by 10, it is multiplied by 20, because of the relationship between intensity and amplitude (recall that intensity is the square of pressure). The formula for SPL is \(SPL(dB) = 20 \log_{10} \frac{P_1}{P_0}\). We know that the reference for amplitude is 20 micropascals, so \(SPL(dB) = 20 \log_{10} \frac{P_1}{20 \mu Pa}\). It is very important to keep in mind, however, that the difference in the equations for intensity and pressure does not mean that a sound with an IL of 30 dB has an SPL of 60 dB.

The decibel unit is dimensionless unless it is anchored to a referent. In other words, if you say that a sound has a decibel level of 32, even though we know this number is in comparison to 0 dB the information is meaningless, because the number lacks a referent. By analogy, if someone told you that Sue is twice as tall as Mary, you still would not know how tall either of them was, despite the comparison between them. Therefore, it is important when using the decibel scale to specify whether amplitude or intensity is being measured. When amplitude is measured, the units on the decibel scale are referenced to sound pressure level (SPL). For intensity, the units are referenced to intensity level (IL). Thus, it is clear that a sound of 32 dB SPL has a certain amplitude in relation to the standard amplitude reference of 20 \(\mu Pa\); a sound of 32 dB IL has a certain intensity in relation to the standard intensity referent of \(10^{-12} \text{ W/m}^2\).

Zero on the decibel scale means that there is a one-to-one ratio of the two sound intensities or amplitudes of the sounds being compared. A unit of 1 dB corresponds to an intensity ratio of about 1.26:1. That is, the higher intensity is 26 percent greater than the lower one (Denes & Pinson, 1993). Thus, a 1 dB step in intensity corresponds to about a 26 percent change. Perceptually, the decibel is the smallest change in sound intensity that an individual with normal hearing can perceive (Durrant & Lovrinic, 1995). Each step in the decibel scale corresponds to a more or less equal increase in a person’s perception of loudness, even though the actual pressure and power differences increase dramatically (Borden, Harris, & Raphael, 1994). Since the decibel scale is logarithmic, a 1 dB step at the threshold of hearing will be a very tiny change. At the intensity level for normal conversational speech, however, which is around 60 dB.
IL, this 26 percent intensity change for 1 dB is 1 million times greater than the 1 dB change at the threshold of hearing.

Since intensity is the square of amplitude, a 100-fold increase in intensity corresponds to a 10-fold increase in amplitude. A 10,000-fold increase in intensity corresponds to a 100-fold increase in amplitude (Denes & Pinson, 1993). Because of this correspondence between intensity and amplitude, the same decibel level can refer to values of both intensity and amplitude. Decibel IL always equals decibel SPL as long as equivalent reference pressures and intensities are used (Speaks, 1992). For example, 60 dB IL equals 60 dB SPL, 78.5 dB IL equals 78.5 dB SPL, and so on. By analogy, a pound of feathers equals a pound of lead, despite the different materials of which they are composed.

Very specific mathematical relationships have been worked out between intensity, amplitude, and the decibel scale. In terms of intensity, any doubling (or halving) of sound power results in an increase (or decrease) of 3 dB. Increasing (or decreasing) intensity by a factor of 10 corresponds to an increase (or decrease) of 10 dB. In terms of amplitude, a doubling (or halving) of sound pressure corresponds to an increase (or decrease) of 6 dB. A ten-fold change in sound pressure corresponds to a change of 20 dB. Also note that doubling (or halving) sound power increases (or decreases) both IL and SPL by 3 dB, whereas doubling (or halving) sound pressure increases (or decreases) both IL and SPL by 6 dB.

<table>
<thead>
<tr>
<th>TABLE 2.2</th>
<th>Familiar Sounds and Sound Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOUND</td>
<td>SOUND LEVEL (dB SPL OR IL)</td>
</tr>
<tr>
<td>Threshold of hearing</td>
<td>0</td>
</tr>
<tr>
<td>Normal breathing</td>
<td>10</td>
</tr>
<tr>
<td>Rustle of leaves</td>
<td>20</td>
</tr>
<tr>
<td>Very soft whisper</td>
<td>30</td>
</tr>
<tr>
<td>Quiet residential community</td>
<td>40</td>
</tr>
<tr>
<td>Department store</td>
<td>50</td>
</tr>
<tr>
<td>Normal conversation</td>
<td>60</td>
</tr>
<tr>
<td>Inside moving car</td>
<td>70</td>
</tr>
<tr>
<td>Loud music from radio</td>
<td>80</td>
</tr>
<tr>
<td>City traffic</td>
<td>90</td>
</tr>
<tr>
<td>Subway train</td>
<td>100</td>
</tr>
<tr>
<td>Loud thunder</td>
<td>110</td>
</tr>
<tr>
<td>Amplified rock and roll band</td>
<td>120</td>
</tr>
<tr>
<td>Machine gun fire at close range</td>
<td>130</td>
</tr>
<tr>
<td>Jet engine at takeoff</td>
<td>140</td>
</tr>
<tr>
<td>Space rocket at blast-off</td>
<td>180</td>
</tr>
</tbody>
</table>

Advantages of the Decibel Scale  One important advantage of using the decibel scale is that huge ranges of intensities are condensed, because of the logarithmic nature of the scale, into around 140 units on the decibel scale. Another advantage is that the relationship between the decibel scale and absolute values of pressure and intensity is very similar to the physiological function of the human auditory system. A change in intensity that is just barely able to be perceived near our hearing threshold is produced by a 1 dB change in the stimulus. As the sound intensity level increases, the intensity change that produces a just perceivable change in loudness continues to be about 1 dB, although the absolute intensity change increases. Table 2.2 provides examples of decibel levels that correspond to the perception of some familiar sounds.

Auditory Area  With a knowledge of frequency, amplitude, and intensity, we are now in a position to examine the human range of hearing in terms of both frequency and intensity. We noted that humans can perceive frequencies from around 20 to 20,000 Hz. The human auditory system is also equipped to perceive an enormously wide range of intensities. However, the intensities that humans are sensitive to depend on the frequency of the sound. The human auditory system is more sensitive to sounds in the midrange of frequencies than to those that are very low or very high. Frequencies in the midrange can be perceived when they are less intense, whereas a very low or very high frequency sound needs to have much more intensity in order to be perceptible to humans. A sound in the middle of the range, say 2000 Hz, requires an intensity of around 11 dB to be just audible (threshold of hearing), whereas a sound with a frequency of 125 Hz requires an intensity of 47.5 dB. According to the ANSI Standard (1969) for audiometers, Table 2.3 shows frequencies and intensities at the threshold of hearing for normal ears.

Although the threshold of hearing changes depending on the frequency of the sound, any frequency with an intensity of 130 dB will cause a sensation of
pain in the ear. This level is known as the **threshold of pain**. Thus, some frequencies have a wider range of intensities at which they can be perceived, whereas others have a more limited range between the threshold of hearing and the threshold of pain. For example, a person with normal hearing would be able to just barely perceive a 1000 Hz tone with an intensity of 7.5 dB and would feel pain in the ear when that tone had an intensity of 130 dB. However, a tone with a frequency of 125 Hz would be just barely audible at an intensity of 47.5 dB, but would still cause pain at an intensity of 130 dB. People with hearing impairment are sensitive to smaller ranges of intensities. For example, a hearing-impaired individual may be only able to detect a 1000 Hz tone when it has an intensity of 60 dB and may find the sound painfully loud at an intensity of 90 dB.

### Resonance

**Free and Forced Vibration**

Vibration of an object can occur freely or by force. An object that is vibrating freely does so without interference at a rate determined by its physical characteristics, including its mass, tension, and stiffness. Whenever this particular object is set into vibration, it will always vibrate at its own specific frequency. This is free vibration, and the frequency at which the object vibrates is called its **natural or resonant frequency** (RF). Free vibration demonstrates the natural response of a vibratory system (Durrant & Lovrinic, 1995). For example, a swing, when pushed, will move back and forth at a certain rate, say twice per second. Its RF is thus 2 Hz, and this is the rate at which the swing will always move as long as its physical characteristics remain constant.

Forced vibration refers to the fact that the vibrations from one object can set another object into vibration if the RFs of both objects are reasonably close to each other. You have probably had the experience of hearing or seeing the pictures hanging on your walls rattle when a car with a particularly loud stereo goes by. If the vibrations generated by the stereo are close to the RF of the wall, the wall will be set into vibration, causing the rattling of the pictures. Or maybe some part in your car starts to rattle in sympathy with the engine noise as you turn the key in the ignition. Here, too, the vibration of the engine may be close in frequency to some other part of your car, which is then forced to vibrate in sympathy with the engine noise. Another example of resonance is portrayed in cartoons in which a singer hits a particularly high note and suddenly a wine glass shatters. This, too, is an example of forced vibration. The glass has its own specific RF. When the singer hits a note that is close to the
glass’s RF, the glass is set into vibration. If the amplitude of vibration is great enough, the glass shatters as it vibrates.

The phenomenon of how the vibration generated by one object forces another object into vibration can be explained using the example of two tuning forks with identical frequencies. When one fork is struck and vibrates and is then brought close to the other fork, the vibrations from fork 1 push against fork 2, and eventually set it into vibration. When tuning fork 2 begins to vibrate, the resulting sound from both tuning forks is louder than the sound generated by tuning fork 1 alone. Using a swing as an analogy, if you push a swing very gently, the first push will hardly move the swing from its rest position. If you keep giving gentle pushes, after a time the swing will start to move, and the more pushes you give it, the greater distance back and forth the swing will move. The timing of the pushes is very important, however. To get the swing to move, the push has to be timed with the movement of the swing toward you, when it has reached its maximum amplitude, and just before it starts swinging back to its rest position. Timed like this, the swing’s amplitude will increase. If you push the swing as it is still moving toward you, its amplitude will be decreased rather than increased. If you push the swing just as it is passing over its rest position on its way toward you, it will stop. In this analogy you are acting as tuning fork number 1, and the swing corresponds to tuning fork number 2.

Essentially, what happens with the tuning forks is constructive interference. Because the two tuning forks have the same frequency, each small vibration arrives at exactly the right moment so that areas of compression and rarefaction combine constructively and the amplitude of the resulting wave increases. Tuning fork 2 has not created the original sound, just as in the swing analogy the swing has not created the original movement. Instead, tuning fork 2 is vibrating in response to the vibrations of tuning fork 1. In the analogy, the swing is set into vibration in response to your pushing.

Another example of forced vibration is the situation in which we set the tuning fork into vibration and then press its stem against a table top. In this case, the fork forces the table top to vibrate at the fork’s RF, even though the RF of the table top is different from that of the fork. Since the table has a much larger vibrating area than the tuning fork, forcing the table top to vibrate in sympathy increases the amplitude of the resulting sound wave.

In resonance, an object is forced to vibrate in response to the vibrations of another object. In the example, tuning fork 1 creates the original sound, and tuning fork 2 resonates the sound, making it louder. Tuning fork 1 supplies the driving or applied frequency, and tuning fork 2 is the resonator. The closer the RF of the driving force is to the RF of the resonator, the greater will be the amplitude of the response of the resonator.
Types of Resonators

There are two types of resonators, mechanical and acoustic. For a mechanical resonator, the actual object itself is set into vibration, for instance, tuning fork number 2 or the table in our examples above. Another type of resonator is a container filled with air. Such a container is an acoustic resonator and is enormously important in the production of speech.

Acoustic Resonators

A volume of air enclosed in a container can resonate. When a sound wave is applied to the air in the container, the air is compressed and rarefied. Because air has elasticity, the air inside the container pushes the compressed air out again. If another sound wave reaches the container at the same time that the compressed air is being pushed out, constructive interference causes an increased amplitude of the sound wave, as long as the applied frequency is close to the RF of the enclosed air. The RF of the air-filled container depends to a great extent on its volume. A smaller volume of air resonates at higher frequencies, whereas a larger volume of air resonates at lower frequencies.

Many musical instruments are acoustic resonators in which the air is set into vibration through the action of some other vibration. A good example is a guitar. The body of the guitar is a container with a round hole in the middle, filled with air. The strings are stretched across the body of the guitar. When you pluck the string, you create a sound by setting the string into vibration. This sets air particles in the vicinity of the string into vibration, causing pressure changes to spread in all directions. Some of the pressure changes go into the body of the guitar through the hole and force the air within the body, as well as the body itself, to vibrate. The air vibrates with the greatest amplitude at frequencies close to the frequency at which the string is vibrating. The string creates the driving frequency, and the air in the body resonates at this frequency. Due to constructive interference, a much louder sound is created. If the string were not stretched across the guitar, but you just held it and plucked it, a very soft sound would result. Forcing the air in the guitar to vibrate with greatest amplitude at the frequency of the string results in a much greater loudness.

Acoustic Resonators as Filters

One reason that acoustic resonators are so important for speech is that they act as filters by filtering some frequencies out of a sound, while allowing others to remain. Let us use as an illustration of acoustic filtering a tube that is perfectly cylindrical along its length. This tube is an acoustic resonator because it is filled with air, and it has a RF of, say, 500 Hz. If you were at one end of this tube and another person were at the other end, and you whistled at different
pitches, with each pitch having the same intensity, the other person would hear some pitches more loudly and some less loudly. The frequencies closest to the tube’s RF of 500 Hz would be heard the loudest. The farther from 500 Hz the frequency of your whistle is, the more softly the other person would hear the sound. There might even be some pitches that the person would not hear at all.

In this example the effect of resonance is to amplify those frequencies that are closest to the tube’s RF and damp or attenuate those frequencies that are farther away from its RF. The whistle close to 500 Hz is similar to the RF of the tube, and so the sound is amplified. The farther away the frequencies of the whistle to the RF of the tube, the more the sound is attenuated or damped. Thus, the resonator acts as a filter by amplifying and transmitting those frequencies close to its own RF and attenuating or preventing frequencies farther away from its own RF from being transmitted. This filtering property of acoustic resonators is one of the fundamental ways in which we produce different sounds.

**Bandwidth** Not all acoustic resonators are perfectly symmetrical, like the tube in our example. Some containers are irregularly shaped or even change their shape. The shape and other physical characteristics of the container, such as whether it is closed at both of its ends, open at both ends, closed at one end and open at the other, and so on, determine the bandwidth of the resonator (see Figure 2.16). Bandwidth refers to the range of frequencies that a resonator will transmit. A symmetrical tube like the one in our example will only transmit a narrow range of frequencies. Our tube could have a bandwidth of 100 Hz. This means that it would transmit those frequencies within 50 Hz of its RF on either side. Frequencies between 450 and 550 Hz would be transmitted and amplified, whereas those frequencies below 450 Hz and above 550 Hz would be damped. This kind of resonator is said to be sharply or narrowly tuned.

A sharply tuned system responds slowly to the driving frequencies. In other words, the amplitude of an applied vibration grows slowly until it reaches its greatest level. A narrowly tuned resonator is also lightly damped. In other words, once it has been forced into vibration, the vibrations take a relatively long time to fade away.

Resonators that are more complex and irregular in shape tend to have wider bandwidths. An irregularly shaped container with an RF of 500 Hz might have a bandwidth of 400 Hz. Thus, frequencies between 300 and 700 Hz would be transmitted, whereas those below 300 Hz and above 700 Hz would be attenuated. Such a resonator is more broadly tuned. A broadly tuned system will respond very quickly to the applied frequencies, but the vibrations will also fade more quickly. A broadly tuned resonator is heavily damped. Broadly tuned systems are common in speech and hearing applications and
include the diaphragms of microphones, earphones, and loudspeakers, as well as our eardrums and vocal tracts.

**Cutoff Frequencies** Resonant systems seldom have a clearcut point above which frequencies are amplified and below which they are attenuated. Instead, frequencies are transmitted with increasingly less efficiency as the driving frequency becomes farther removed from the RF of the system, until the amount of acoustic energy that is transmitted is so small as to be basically nonexistent.

A numerical value has been developed to describe the point at which the resonant system is considered to be unresponsive. This is the point where the intensity transmission is reduced by one-half and is known as the **cutoff**
frequency. Remember that a reduction in intensity of one-half is equivalent to a decrease of 3 dB. Therefore, the frequency at which the intensity is 3 dB less than the peak intensity of the RF is the cutoff frequency and is also, therefore, called the 3 dB down point. Another way the cutoff frequency can be stated is in percentage form. The 3 dB cutoff corresponds to a percentage of 70.7. In other words, any frequency within the resonator’s bandwidth will generate an output whose amplitude will be at least 70.7 percent of the amplitude of the vibrations caused by the frequency closest to the RF of the resonator.

**Resonance Curves** The way in which a resonator vibrates in response to any applied frequency can be described by a graph known as a resonance (or filter) curve. This curve is also called the *transfer function* of a resonant system. If we apply different frequencies to a resonator and each frequency has the same amplitude, the resonator will be forced into vibration by each of these applied frequencies. However, the applied frequencies closest to the RF of the resonator will cause the largest vibrations. The sounds that are used to set a resonator in motion are known as the input to the resonator. The way in which the resonator vibrates in response to these sounds is known as its output for a given input, and it is this input–output relationship that is shown on a resonance curve. Another way of stating this relationship is that a resonance or filter curve shows the frequency response of a resonant system.

As Figure 2.17 shows, the greatest amplitude of response occurs at the RF of the system. Frequencies far from the RF have been filtered, so their
amplitudes are less than the amplitude at the RF. Thus, this resonance curve describes how the amplitudes of the various component frequencies of the sound wave are affected by the resonator or, to put it slightly differently, how the spectrum of a sound is changed when the sound is resonated with this particular system. It is important to keep in mind that this curve is not a sound wave, but describes the frequency response, or transfer function, of the resonator.

**Parameters of a Filter** All resonators, or filter systems, have certain characteristics. These include the natural or resonant frequency, the upper cutoff frequency, the lower cutoff frequency, the bandwidth, and the attenuation or rejection rate.

The natural frequency is also called the **center frequency** \( (F_c) \). This is the resonant frequency of the system that results in the greatest amplitude of vibration of the resonator. The center frequency depends on the physical characteristics of the resonator, such as its length and shape. In the example earlier, the center frequency of the tube was 500 Hz.

The **upper cutoff frequency** \( (F_u) \) is the frequency above \( F_c \) at which there is a 3 dB less amplitude of response of the resonator than that at \( F_c \). The **lower cutoff frequency** \( (F_l) \) corresponds to the frequency below \( F_c \) where the amplitude is decreased by 3 dB. Because a 3 dB decrease in amplitude corresponds to a halving of amplitude, both \( F_u \) and \( F_l \) are also called the 3 dB down points, or the **half-power points**.

The bandwidth, or **passband**, refers to the frequencies between \( F_u \) and \( F_l \), that is, the range of frequencies that the resonator will transmit. The bandwidth can be broad or narrow, depending on the physical characteristics of the resonator. The rate at which the resonator's amplitude of response is attenuated is known as the **attenuation rate**. (It is also referred to as the **roll-off rate**, the **rejection rate**, or the **slope**.) This parameter describes how rapidly the resonator decreases in its amplitude of response to different frequencies and is measured in decibels per octave. Slopes can range from shallow to steep. A slope of less than 18 dB/octave would be considered fairly shallow. A filter with an attenuation rate between 18 and 48 dB/octave is moderately steep, while filters with cutoffs greater than 90 dB/octave are extremely steep (Rosen & Howell, 1991).

**Types of Filters** Different types of filters are suitable for performing different types of functions. Three kinds of filters are commonly encountered in speech pathology and audiology. First is a **low-pass filter**, which passes acoustic energy below a specific upper cutoff frequency. Acoustic energy above the \( F_u \) is attenuated at a particular rejection rate. Frequencies below the cutoff frequency are passed through the system well, while those above the cutoff frequency are attenuated. Second is a...
**high-pass filter**, which passes energy above a designated lower cutoff frequency. Energy below $F_l$ is rejected at a particular attenuation rate, while acoustic energy above $F_l$ is transmitted through the system. Third, a **band-pass filter** passes energy in a particular range of frequencies between an $F_l$ and an $F_u$. Energy outside this range is rejected at a specific attenuation rate. A band-pass filter is a combination of low-pass and high-pass filtering. The low-pass filter transmits energy below the $F_u$, while the high-pass filter transmits energy above the $F_l$. The vocal tract is an example of a band-pass filter.

### Summary

- Sound consists of increases and decreases in air pressure caused by the movement of a source, such as a tuning fork.
- Sound waves are characterized by different dimensions of frequency, period, wavelength, amplitude, and intensity.
- Sound waves can consist of one frequency (pure tone) or many frequencies (complex waves).
- Sound waves can be visually depicted on waveforms and spectra.
- Amplitude and intensity of sounds can be measured conveniently on the decibel scale.
- Resonance involves forced vibration in which an object or container of air is set into vibration by the action of another vibration.
- Acoustic resonators may be sharply or broadly tuned, with different center frequencies and upper and lower cutoff frequencies.

### Review Exercises

1. Use an example other than a tuning fork to explain how sound is generated. Describe the forces involved in generating, maintaining, and damping vibration.
2. Define and explain the terms *reverberation, absorption, reflection, and interference.*
3. Discuss the following statement: “Speech is a stream of periodic and aperiodic complex sounds.”
4. Describe the decibel scale and explain its major advantage in measuring sound intensity.
5. Discuss the concept of resonance and explain how an acoustic resonator acts as a filter.
An understanding of the basic nature of sound and of the important dimensions of sound, including frequency and intensity, allows speech-language pathologists and audiologists to apply this knowledge in the diagnosis and treatment of many different communication problems. Norms have been developed for many acoustic variables related to human voice and speech pro-
duction, including various aspects of frequency and intensity. Norms are available for different age groups, including infants and toddlers, school-age children, and adults at different stages of life. These objective normal values form a scientifically based body of knowledge against which the voice and speech characteristics of individuals with various communication disorders can be compared, such as functional and organic voice disorders, neurological problems, and speech problems related to hearing impairment.

These kinds of comparisons are invaluable in clinical situations. They help to refine the process of differential diagnosis by providing information about aspects of speech production that may not be detected or processed by human ears, but that offer important clues about a patient’s condition. Objective measures are vital also to validate a clinician’s perceptual judgments of disordered voice and speech. In addition, they offer a starting point for rehabilitation and provide an objective means of assessing the patient’s progress. Objective acoustic measures not only supplement the clinician’s therapeutic skills, but also strengthen clinician accountability, which is a crucial issue in health care management. More and more, third-party payers such as insurance companies and HMOs will only pay for benefits if it can be shown that the treatment is effective. Objective measures are invaluable in providing this kind of supporting information about the patient’s progress.

However, a word of caution is in order. Although norms, in general, provide a valuable basis of comparison, abnormal data do not always indicate a pathology, even when they exceed typical values. It is important, therefore, for clinicians and researchers to critically evaluate the information obtained and to supplement it with information from other sources.

Vocal Frequency and Amplitude

When people talk, their vocal folds open and close to produce vibration, resulting in a complex periodic sound with a certain fundamental frequency ($F_0$) and amplitude. As with any complex periodic sound, the $F_0$ is perceived as the pitch of the sound, and the amplitude is perceived as its loudness.

Vocal fold vibration gives rise to numerous aspects of vocal $F_0$ and amplitude that can be measured by way of acoustic instrumentation. A commonly used instrument in university speech and hearing clinics, rehabilitation centers, schools, and hospitals is the Visi-Pitch, made and distributed by Kay Elemetrics. This is a computer-based transducer fitted with acoustic hardware and software. A person speaks into an attached microphone, which changes, or transduces, the acoustic signal into corresponding electrical signals. The hardware and software digitize the signal, converting it into a format that the computer can process. The Visi-Pitch calculates $F_0$ and relative amplitude over time,
and displays these on a monitor. Variables related to \( F_0 \) and amplitude can then be determined, such as the average and range. Other instruments are also commercially available for acoustic measurement of voice \( F_0 \) and amplitude.

**Frequency Variables**

\( F_0 \) variables that are very commonly used in the clinical situation include average \( F_0 \), \( F_0 \) variability and range, and maximum phonational frequency range.

**Average Fundamental Frequency** The \( F_0 \) measured in a particular task, such as sustaining a vowel, reading aloud, or conversational speech, is averaged over the speaking time of that task. When average \( F_0 \) is measured in an oral reading or conversational speech task, it is often referred to as the person’s speaking fundamental frequency (SFF). Many speech scientists have examined the average \( F_0 \) in different age groups and across both sexes. Table 3.1 shows some of the values for average \( F_0 \) that have been derived using various types of acoustic instrumentation.

Looking at the table, you can see a systematic pattern of \( F_0 \) values that vary according to age level and sex. Infants in the first several years of life have a very high \( F_0 \), from around 350 to almost 500 Hz. In musical terms, this is from around the F an octave above middle C to the B almost two octaves above middle C. This high \( F_0 \) is the result of the infant’s very short and thin vocal folds, which have a very rapid rate of vibration. As the baby

---

**Table 3.1**

**Average \( F_0 \) for Different Age Groups Reported in the Literature**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Age Group</th>
<th>Males</th>
<th>Females</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robb &amp; Saxman (1985)</td>
<td>11–25 (mos)</td>
<td>357</td>
<td>357</td>
<td></td>
</tr>
<tr>
<td>McGlone &amp; Shipp (1971)</td>
<td>13–23 (mos)</td>
<td>443</td>
<td>443</td>
<td></td>
</tr>
<tr>
<td>Eguchi &amp; Hirsh (1969)</td>
<td>3</td>
<td>298</td>
<td>298</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>286</td>
<td>286</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>289</td>
<td>289</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5–6</td>
<td>241</td>
<td>231</td>
<td>AA</td>
</tr>
<tr>
<td></td>
<td>5–6</td>
<td>249</td>
<td>248</td>
<td>H</td>
</tr>
<tr>
<td>Weinberg &amp; Bennett (1971)</td>
<td>5</td>
<td>252</td>
<td>248</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>247</td>
<td>247</td>
<td></td>
</tr>
<tr>
<td>Wheat &amp; Hudson (1988)</td>
<td>6</td>
<td>219</td>
<td>211</td>
<td>AA</td>
</tr>
<tr>
<td>Ferrand &amp; Bloom (1996)</td>
<td>3–6</td>
<td>256</td>
<td>246</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7–10</td>
<td>237</td>
<td>253</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 3.1 continued

<table>
<thead>
<tr>
<th>AUTHOR(S)</th>
<th>AGE GROUP</th>
<th>MALES (HERTZ)</th>
<th>FEMALES (HERTZ)</th>
<th>RACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorenson (1989)</td>
<td>6</td>
<td>251</td>
<td>296</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>288</td>
<td>258</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>229</td>
<td>251</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>221</td>
<td>266</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>220</td>
<td>229</td>
<td></td>
</tr>
<tr>
<td>Pederson et al. (1990)</td>
<td>8.6–12.9</td>
<td>256</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13–15.9</td>
<td>248</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16–19.8</td>
<td>241</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bennett (1983)</td>
<td>8.2</td>
<td>234</td>
<td>235</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.2</td>
<td>226</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.2</td>
<td>224</td>
<td>228</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.2</td>
<td>216</td>
<td>221</td>
<td></td>
</tr>
<tr>
<td>Fitch (1990)</td>
<td>21–26</td>
<td>109</td>
<td>210</td>
<td></td>
</tr>
<tr>
<td>Kent (1994)</td>
<td>20.3</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47.9</td>
<td>123</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>73.3</td>
<td>119</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>136</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hollien &amp; Shipp (1972)</td>
<td>20–29</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30–39</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40–49</td>
<td>107</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50–59</td>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60–69</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70–79</td>
<td>132</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80–89</td>
<td>146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Pinto &amp; Hollien (1982)</td>
<td>18–25</td>
<td>229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgan &amp; Rastatter (1986)</td>
<td>20–24</td>
<td>228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pegoraro Krook (1988)</td>
<td>20–29</td>
<td>195</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30–39</td>
<td>195</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40–49</td>
<td>191</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50–59</td>
<td>182</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60–69</td>
<td>181</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70–79</td>
<td>188</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80–89</td>
<td>188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown et al. (1996)</td>
<td>21–33</td>
<td>136</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20–22</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40–54</td>
<td>128</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>42–50</td>
<td>186</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>69–87</td>
<td>134</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65–89</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honjo &amp; Isshiki (1980)</td>
<td>69–85</td>
<td>162</td>
<td>177</td>
<td></td>
</tr>
</tbody>
</table>

Values for postlingual children and adults are for conversational speech or oral reading. Some studies included race as a factor: white (W), African American (AA), Hispanic (H). Unless otherwise stated, ages are in years.

**Note:** Some values have been rounded to the nearest whole number.
grows, his or her vocal folds lengthen and thicken, with a corresponding decrease in $F_0$. From around age 3 to 10 years, both males and females have an average $F_0$ of approximately 270 to 300 Hz. After puberty, the average $F_0$ for males drops markedly, while that for females remains essentially the same or may decrease slightly. Again, these $F_0$ patterns are related to growth factors. At puberty, the male larynx enlarges considerably, and the vocal folds become longer, thicker, and more massive, with a corresponding dramatic drop in $F_0$. A girl’s larynx and vocal folds also enlarge somewhat during puberty, but not to the same extent as a boy’s. By age 20, the average $F_0$ for males is around 120 Hz, while for females the average $F_0$ is approximately 100 Hz higher than males, at 220 Hz.

Average $F_0$ remains fairly stable for adult males and females until the sixth and seventh decades, when males’ average $F_0$ typically increase due to age-related degenerative changes that occur within the larynx, including a thinning of the vocal folds. Because objects that are thinner and less massive vibrate more quickly than those that are more massive, the thinning of the male’s vocal folds produces an increase in vocal $F_0$. Older women, on the other hand, tend to develop more massive vocal folds due to hormonal changes. Thus, the $F_0$ of older women tends to decrease with age.

These age- and gender-related values for average $F_0$ are vitally important in clinical situations. First, they tell us that we cannot judge $F_0$ the same way for young children and adults, or for younger adults and older adults, or for adult males and females. Second, they provide an objective basis for evaluating the appropriateness of an individual’s pitch level. That is, is it normal or too high for their age and gender or too low for their age and gender? So, if we are interested in judging if a person’s pitch level is adequate, why not just listen to her or him and rate the pitch according to what we hear? The answer is that subjective, perceptual methods of assessing pitch are unreliable. Although $F_0$ and pitch are related, how we perceive pitch also depends on the interaction of frequency, intensity, and other properties of the sound. In addition, the human auditory system responds more easily to some frequencies than others. For example, pitch changes at lower frequencies are usually perceived much more easily than changes at higher frequencies. Raising $F_0$ from 100 to 200 Hz results in a much greater change in perceived pitch than going from 3000 to 3100 Hz (Baken, 1996). Thus, a pitch level that is heard as abnormal may be due to the speaker’s actual $F_0$, vocal intensity, or other factors. Therefore, to make a clinical decision about whether or how to treat an individual with a pitch disorder, objective acoustic measurement of the person’s $F_0$ is critical. Quantifying a speaker’s $F_0$ levels and comparing his or her SFF to established norms for speakers of similar age and sex will help the clinician decide whether the perceived abnormality really does result from an $F_0$ problem, or whether (and what) other vocal factors are involved.
Frequency Variability  People constantly change their $F_0$ levels as they speak to reflect different emotions, different types of accenting and stress of syllables, and different grammatical constructions. These $F_0$ changes contribute to the overall melody, or prosody, of speech. For instance, the sentence “Peter’s going home” can be said either as a declarative statement or as a question. As a declarative, the $F_0$ level drops at the end of the utterance, whereas for a question the $F_0$ level rises at the end of the utterance. The prosody of a sentence also is influenced by the mood of the speaker. There are likely to be many more $F_0$ changes, and more extensive changes, when the individual is wildly excited that Peter’s going home than when the speaker is depressed by Peter’s plans. Acoustically, these $F_0$ changes correspond to frequency variability. A certain amount of frequency variability is desirable in a speaker’s voice, depending on the individual’s age, sex, social situation, mood, and so on. This variability is something that speakers of a particular language in a particular culture intuitively recognize. Too much or too little frequency variability sounds wrong and can indicate a functional, organic, or neurogenic voice problem.

$F_0$ variability is measured in terms of standard deviation (SD) from the average $F_0$. Standard deviation is a statistical measure that reflects the spread of scores around the average score, so standard deviation of $F_0$ reflects the spread of $F_0$ around the average $\bar{F}_0$. When this variability is measured in hertz, it is called $F_0$SD. $F_0$SD in normal conversational speech is around 20 to 35 Hz. $F_0$SD is likely to increase when the speaker is excited or agitated. Sometimes $F_0$SD is converted to semitones. When the frequency variability is discussed in semitones rather than hertz, it is called pitch sigma. Pitch sigma for normal speakers during conversation should be around 2 to 4 semitones for both males and females (Colton & Casper, 1996).

Another measure of $F_0$ variability is the range, which is the difference between the highest and lowest $F_0$ in a particular sample of speech. The range can be expressed in hertz, or it can be converted to semitones and octaves. Table 3.2 displays $F_0$ range information for different age groups.

An interesting trend emerges from the numbers in Table 3.2. The infants (11 to 25 months) have by far the greatest range of frequencies. This is not surprising, because their vocalizations include a wide variety of nonwords, such as squeaks and squeals and cries. From around age 3 years, when children have mastered speech, to just before puberty, the range of $F_0$ used in normal conversational speech is around 150 to 200 Hz. The range decreases further in the adult years. From approximately age 7 years on, females tend to use a wider range of $F_0$ than males. This may be a sociocultural rather than a physical phenomenon. Ferrand and Bloom (1996) compared $F_0$ ranges in various age groups of boys and girls. The 7- to 10-year-old boys in their study had a narrower range than the 3- to 6-year-old boys. The range of the 7- to 10-year-old girls, however, was as wide
as for younger girls. The authors suggested that boys, a few years before puberty, start to imitate the prosodic patterns of adult males, who tend to use fewer pitch levels than females and who also, in general, avoid dramatic pitch shifts.

$F_0$ variability, whether expressed in semitones, $F_0$SD, or $F_0$ range, is an important indicator of normal or disordered speech. In the clinic situation, it is very common for a person with a voice or speech disorder to demonstrate a reduced $F_0$ range. This is particularly true in many neurological problems, such as vocal fold paralysis or Parkinson’s disease. $F_0$SD is also often used diagnostically to determine how well a person is able to control her vocal fold vibration. When a speaker prolongs a vowel, there should be very little $F_0$ variability in her production, because the goal is to produce the vowel with as steady an $F_0$ as possible. In this type of task, $F_0$SD should be low, around 3 to 6 Hz. A figure higher than this may indicate that the speaker has difficulty in controlling the frequency aspects of vocal fold vibration.

**Maximum Phonational Frequency Range**   Maximum phonational frequency range (MPFR) refers to the complete range of frequencies that an individual can generate, as opposed to the $F_0$ variability measure, which refers to the range of $F_0$ a person generally uses in connected speech. MPFR has been defined as the range of frequencies from the lowest tone that the person can sustain to the highest, including *falsetto*, which refers to a very high range of

### Table 3.2

<table>
<thead>
<tr>
<th>Authors</th>
<th>Age Group</th>
<th>Males (Hertz)</th>
<th>Females (Hertz)</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robb &amp; Saxman (1985)</td>
<td>11–25 (mos)</td>
<td>1202</td>
<td>1202</td>
<td></td>
</tr>
<tr>
<td>Ferrand &amp; Bloom (1996)</td>
<td>3–4</td>
<td>202</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5–6</td>
<td>214</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7–8</td>
<td>158</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9–10</td>
<td>151</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5–6</td>
<td>204</td>
<td>190</td>
<td>AA</td>
</tr>
<tr>
<td></td>
<td>5–6</td>
<td>185</td>
<td>164</td>
<td>H</td>
</tr>
<tr>
<td>Benjamin (1981)</td>
<td>Younger adults</td>
<td>64</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Older adults</td>
<td>78</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Fitch (1990)</td>
<td>21–26</td>
<td>46</td>
<td>77</td>
<td></td>
</tr>
</tbody>
</table>

Some authors provide race information for white (W), African American (AA), and Hispanic (H) groups. Age is in years unless otherwise stated.

Notes:
(a) Some values were derived by subtracting minimum from maximum frequencies.
(b) Some values have been rounded to the nearest whole number.
frequencies (Baken, 1996). MPFR is often measured in semitones or octaves. A range of around 3 octaves is normal for young adults, while older adults may show a decrease in range. In terms of hertz, the lowest frequency that adult males are able to produce is around 80 Hz, and the highest is in the 700 Hz range (Colton & Casper, 1996). Adult females produce, on the whole, a low $F_0$ of approximately 135 Hz, with a high level that can reach over 1000 Hz. Trained singers, of course, are able to produce an even higher $F_0$ than this. Kent (1994) provided normative MPFRs in semitones for individuals from age 8 years to late adulthood. The lowest value of 24.3 semitones was demonstrated by older men in poor health; the greatest range of 41.4 semitones was shown by teenaged boys. Most values fell between these extremes, in the range of 30 semitones, or 2.5 octaves.

The MPFR is a useful measure because it reflects both the physiological limits of a speaker’s voice (Colton & Casper, 1996) and the physical condition of the person’s vocal mechanism and basic vocal ability (Baken, 1996). In fact, it has been shown that, for normally speaking adults, neither age nor sex greatly affects MPFR. What does affect this value is physical condition. Older subjects in good health tend to have larger MPFRs than younger subjects in poor physical condition. Thus, a clinician may suspect that a voice problem exists in a speaker who has a reduced MPFR.

### Amplitude and Intensity Variables

Similarly to frequency variables, the voice amplitude that a person can generate is often an important indication of the normalcy or pathology of the vocal apparatus. Several common measurements of amplitude include the average amplitude level, amplitude variability, and dynamic range. Amplitude variables are typically measured in dB SPL (0.0002 dyne/cm²). It is important to keep in mind that the voice amplitude that a person can generate depends strongly on the vocal $F_0$. We will talk more about this relationship between frequency and amplitude when we discuss the voice range profile.

#### Average Amplitude Level

Like average $F_0$, average amplitude refers to the overall level of amplitude during a speech task such as oral reading, conversation, or sustaining a vowel. Perceptually, this corresponds to the loudness that the individual generates during the speech activity. Amplitude level varies depending on the speaker’s situation. A person’s average amplitude level will be much lower during a soft conversation in a classroom and much higher when cheering at a football game. Normal conversational speech usually ranges between 65 and 80 dB SPL, with an average SPL in the general range of 70 dB for adult males and females (Baken, 1996). Children seem to use similar amplitude levels as adults. Amplitude levels do not depend on age as much as do frequency levels,
although there is some evidence that amplitude may be decreased slightly in older individuals.

Clinically, a reduced amplitude level is strongly indicative of speech disorder, particularly those resulting from neurologic disease. Parkinson’s disease, for example, is typically characterized by very low vocal amplitude, because the afflicted person has great difficulty in opening his or her vocal folds widely enough and closing them. Also, weak vocal amplitude is a large part of the problem in *alaryngeal speech*, that is, speech produced without a larynx after an individual has the larynx removed due to cancer. Obtaining acoustic information in these kinds of cases is vital for diagnosis and intervention purposes. Similar to the measurement of average frequency, measurement of average amplitude provides a basis of comparison for the patient’s voice before, during, and after therapy. Identifying the specific amplitude levels that a patient is able to generate for various speech tasks can help the clinician determine whether a particular treatment strategy is producing the desired effect or whether a different strategy should be tried.

**Amplitude Variability** During any conversation, amplitude varies depending on the speaker’s mood, feelings, the message he or she is conveying, the stress and accenting of syllables and words, and so on. Titze (1994) remarked:

> Loudness variation is an important part of phrasing in speech. It serves as a punctuation, the setting apart of words, phrases, sentences, and paragraphs. In addition, loudness variation is used to emphasize . . . or to get attention. In poetry, rhetoric, and song, large variations in intensity are used to dramatize emotions, to convince someone of a point of view, or simply to entertain. (p. 246)

As with \( F_0 \), amplitude variability is expressed as a standard deviation, measured in dB SPL. Standard deviation of amplitude for a neutral, unemotional sentence is around 10 dB SPL. The greater the speaker’s level of excitement or enthusiasm, the greater the variability is likely to be. Because amplitude variability is such an important marker of emphasis, emotion, and the like, a reduced ability to vary loudness can be highly upsetting. Not much research exists regarding amplitude variability in different age groups. However, many different communication disorders are characterized by such reduction in amplitude variability, including those resulting from organic and neurologic causes. In general, people who have reduced average amplitude levels also have problems with amplitude variability. As with frequency, the ability to vary amplitude is what imparts dynamism and interest to a speaker’s voice. Individuals who lack this ability, for whatever reason, tend to sound flat and monotonous.
Dynamic Range  Dynamic range is similar to phonational range. It relates to the physiological range of the vocal amplitudes that a speaker can generate, from the softest phonation that is not a whisper to the loudest shout. A normal adult female should be able to produce a minimum level of around 50 dB and a maximum of approximately 115 dB SPL; the figures for normal adult males should be slightly higher (Coleman, Mabis, & Hinson, 1977). Depending on frequency, the minimum dynamic range should be around 30 dB SPL. A restricted dynamic range may prevent a person from using stress and emphasis patterns appropriately, reducing the flexibility of spoken language. Although amplitudes above and below the 60 to 80 dB range are not normally used for conversational speech, the ability to raise one’s voice is important for special occasions that demand higher levels of loudness (Sulter, Schutte, & Miller, 1995). The dynamic range depends on the $F_0$ produced and tends to be greatest for $F_0$ in the midrange and less for $F_0$ that is much lower or much higher.

Voice Range Profile  The voice range profile (VRP, also called a phonetogram or $F_0$ SPL profile) is a graph that plots a person’s phonational range against his or her dynamic range. Dynamic range is plotted on the vertical axis in dB SPL, and $F_0$ is plotted on the horizontal axis in hertz. See Figure 3.1.

---

**Dynamic range** refers to the range of vocal amplitudes a speaker can generate from the softest phonation that is not a whisper to the loudest shout, measured in dB.

---

**Voice range profile.**
To generate the profile, the speaker phonates a vowel at various $F_0$ levels. At each $F_0$ level, the individual produces the softest and loudest sounds that she can. This results in two contours. The upper contour shows maximum intensity at each selected frequency, and the lower contour shows minimum intensity at each frequency, giving a good indication of normal frequency and amplitude relationships in the human voice.

VRPs have a characteristic shape, reflecting some physiological quirks of the human voice production system. The shape is roughly similar to an oval, with narrower endpoints and a more expanded mid-portion. This shape results from the typical relationship between the range of frequencies and amplitudes that we can generate. Humans have a far greater dynamic range in the middle of the frequency range, whereas the dynamic range shrinks considerably at very high or very low frequencies. In the middle of the frequency range, an individual can generally vary the intensity by 20 to 30 dB, whereas at the ends of the frequency range the person might only be able to vary intensity by a few decibels (Titze, 1994).

Another feature of the shape is the dip that is usually seen in the maximum intensity upper contour at around 390 Hz for men and 440 Hz for women (Sulter et al., 1995). This dip reflects a drop in maximum intensity that occurs when speakers without any voice training go from their usual conversational speech frequency range to the much higher falsetto range. Trained singers show a considerably reduced dip, indicating a smoother transition between these ranges.

The VRP can be thought of as a snapshot of vocal fold behavior at one moment in time (Aerainer & Klingholz, 1993). This picture gives very useful information for several reasons. First, a VRP can help to determine the physiological limits of any individual’s voice, because the dynamic and phonational frequency ranges are directly related to the person’s ability to control the vocal folds (Sulter et al., 1995). A person who has difficulty in achieving normal frequency and amplitude ranges will demonstrate a constricted or compressed VRP, with the upper and lower contours closer together than normal. Second, this kind of graph can show the impact of treatment or surgical intervention on an individual’s voice. An expanded VRT after treatment would show graphically that the patient’s phonational and/or dynamic range has increased.

Useful clinical information can also be obtained by determining where the graph falls in terms of frequency. For example, Behrman, Agresti, Blumstein, and Sharma (1996) presented the case of an 11-year-old girl with a voice problem whose presurgery frequency range on the VRP extended from 440 to 987 Hz. Pictorially, her VRP was shifted to the right, to the higher frequencies. Table 3.1 shows that average $F_0$ for an 11-year-old girl should be somewhere
between 220 and 250 Hz. Presurgery, this child was unable physiologically to achieve the normal frequency levels for her age and sex, and her pitch was correspondingly abnormally high. Postsurgery, her frequency range on the VRP shifted to the more normal values of 196 to 494 Hz, shown as a shift to the left on the VRP. Perceptually, too, her pitch level sounded much more appropriate for an 11-year-old girl. In this case, the VRP provided compelling objective evidence that the surgery had been successful.

Another use of VRPs is to compare the vocal characteristics of different groups of speakers. In comparison to adults, for instance, children demonstrate a somewhat compressed VRP. Children’s upper contours are lower and their lower contours are higher than those of adults. This is a graphical and objective demonstration that children are not able to generate as high or as low amplitudes as adults, most likely due to the different structural and physiological characteristics of the vocal folds in children and adults (McAllister, Sederholm, Sundberg, & Gramming, 1994). VRPs also show differences between trained and untrained voices. Sulter and colleagues (1995) looked at differences in VRP features in males and females with and without vocal training. Untrained women obtained an average $F_0$ range of around 157 to 1223 Hz. Vocally trained women obtained a range from around 128 to 1320 Hz, showing a wider range at both the lower and higher ends of the frequency range. Vocally trained men obtained an average range from around 74 to 785 Hz, whereas their untrained counterparts had a range from 86 to about 688 Hz. Plotting this information on a VRP showed that not only do trained singers have wider phonational ranges than nonsingers, but they are also able to phonate at lower intensities over almost the entire frequency range.

**Breakdowns in Control of Vocal Frequency and Amplitude**

So far we have been discussing in a theoretical way the various dimensions of sounds, including those of $F_0$ and amplitude. These measurements are extremely valuable when dealing with breakdowns in the control of frequency and amplitude, as often occurs in voice and neurological disorders.

**Voice Disorders**

Patients with voice disorders are very commonly seen in hospitals, clinics, private practice, nursing homes, preschools, and elementary and secondary schools. Individuals of any age, including infants, can exhibit voice disorders.
Voice disorders can result from numerous and varied causes, including vocal abuse, organic problems such as benign tumors or cysts on the vocal folds, neurological problems such as stroke or progressive degenerative diseases (e.g., Parkinson’s disease), trauma such as car accidents and gunshot wounds that affect the larynx, and numerous others. Voice disorders result in problems related to frequency and amplitude. In terms of frequency, speakers may use a range of $F_0$ that is too low or too high for their age, sex, and build; may produce a restricted range of $F_0$ (i.e., speak in a monotone); may demonstrate frequency breaks, in which the voice shifts $F_0$ involuntarily and abruptly; or may demonstrate diplophonía, which occurs when each vocal fold vibrates at a slightly different rate, resulting in the perception of two pitches simultaneously. Often a combination of these frequency control problems exists. Amplitude problems include using a habitual amplitude level that is too high or too low, producing a restricted range of amplitudes (i.e., monoloudness), or having sudden involuntary and inappropriate amplitude changes. Often, frequency and amplitude problems occur simultaneously in various combinations.

Before sophisticated computerized equipment was widely available, clinicians used to rely solely on their subjective, perceptual impressions of the person’s voice in making a diagnosis of the voice problem. However, with current instrumentation it is not only important but easy to supplement this perceptual information with more objective and quantifiable information about the acoustic characteristics of the speaker’s voice. This kind of information is invaluable for making finely tuned diagnoses about the problem, detecting early changes in speech and voice that are not apparent perceptually, making intelligent decisions about treatment options, and assessing the outcomes of treatment. Treatment options and results can be compared and validated. For instance, some kinds of voice problems related to emotional stress are often treated by teaching the individual to reduce tension in the larynx. However, until recently, there was little evidence that this kind of treatment actually works. To determine the effectiveness of tension reduction treatment techniques, Roy, Bless, Heisey, and Ford (1997) audiorecorded 25 patients pre- and posttreatment and analyzed these voice samples both perceptually and acoustically in terms of $F_0$. The patients’ $F_0$ values in connected speech changed from the pretreatment session to the posttreatment sessions, reflecting a decrease in laryngeal tension. These acoustic data provide objective support for the effectiveness of this particular kind of voice therapy.

In addition, the precise measurement of $F_0$ can help the clinician to make decisions about voice parameters that are difficult to verify perceptually. For example, a person with vocal nodules typically demonstrates a voice that sounds low-pitched, hoarse, and breathy. Before acoustic instrumentation was commonly used, many clinicians treated this problem by having the patient use a higher pitch level. Often, however, the low pitch that is heard does not
correspond to a lower than normal \( F_0 \). In other words, the patient’s \( F_0 \), when objectively measured, is actually within normal limits, but is perceived as lower due to the influence of other factors, such as rate of speech and the hoarseness itself. Because the speaker’s \( F_0 \) is within normal limits, treatment focusing on changing the \( F_0 \) is not indicated. Thus, this acoustic information signals the clinician to choose a more appropriate treatment strategy based on the physiological functioning of the vocal folds. This kind of precise measurement, then, can help the clinician to avoid implementing ineffective treatment plans.

Other uses of \( F_0 \) measurement help in making medically related decisions. For instance, Orlikoff, Kraus, Harrison, Ho, and Gartner (1997) used \( F_0 \) as an indication of how a cancerous tumor on the vocal folds responded to chemotherapy. They measured “comfortable” \( F_0 \) plus variability in \( F_0 \) during speaking situations in patients with advanced laryngeal cancer before each of three cycles of chemotherapy. They found a link between the reduction or elimination of the cancerous growth and the \( F_0 \) measures. \( F_0 \) variability increased when the growth was reduced, indicating that the patients had increased their vocal flexibility. These frequency measures were helpful in documenting both the extent of the laryngeal impairment resulting from the cancer and the effectiveness of the chemotherapy.

Frequency and intensity characteristics have been used as a yardstick for successful voice restoration in people who have had their larynx removed due to cancer and who use different methods of alaryngeal speech (speech produced without a larynx). One such type of voice production is called esophageal speech. This is voice produced by vibrating a certain part of the esophagus, rather than the vocal folds in the larynx. Speech produced in this way typically is much lower in \( F_0 \) and amplitude than normally produced speech.

Clinicians have traditionally judged the effectiveness of esophageal speech partially in terms of its frequency and amplitude. Historically, speech pathologists have characterized proficient esophageal speech as having a higher \( F_0 \) and higher amplitude level than poor esophageal speech. Slavin and Ferrand (1995) tested this assumption by acoustically analyzing the speech of esophageal speakers who were all judged perceptually as being highly proficient speakers. The average frequency for the entire group of speakers was 69 Hz, considerably lower than the normal values for adult speakers presented in Table 3.1. However, Slavin and Ferrand (1995) found that these individuals clustered into four groups, with each group characterized by a different frequency and amplitude profile. For instance, individuals in one group had higher than average mean \( F_0 \) and greater frequency variability than the group mean. Speakers in a different group had lower mean \( F_0 \) and less frequency variability. A third group of speakers obtained an \( F_0 \) around 69 Hz, but these individuals had relatively high amplitude levels (around a 70 dB SPL). Understanding that different patterns of frequency and amplitude characterize individual esophageal speakers can allow clinicians to be more flexible in choosing rehabilitation strategies geared
Neurological Disorders

A voice or speech problem can often be the first sign of a more generalized neurological disorder. This was shown in a classic study on voice and speech symptoms in patients with Parkinson’s disease (PD) by Logemann, Fisher, Boshes, and Blonsky in 1978. They found that 89 percent of these individuals showed a voice problem. Indeed, almost half of this number had shown a voice problem as the first sign of the neurological disease. Other neurological disorders that tend to affect speech and voice production include amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS), Huntington’s chorea, and many others. Strokes, brain tumors, and traumatic brain injury also contribute to voice and speech problems. Over the past ten to fifteen years, much information has been collected regarding the acoustic characteristics of voice that often result from these disorders. This information is used to plan, carry out, and evaluate different types of therapy regimes.

In 1975, Darley, Aronson, and Brown made voice history by conducting a large-scale study to investigate the voice and speech characteristics of patients with various kinds of neurological problems. They characterized these problems in terms of perceptual parameters of voice and speech, such as problems with pitch (too high, too low, monotone) and loudness (too loud, too soft, monoloud), and derived different profiles of vocal characteristics for different disorders. See Table 3.3.

Recently, researchers have begun to supplement these perceptual dimensions of vocal function with more objective information. For example, acoustic measures can provide objective validation of perceptual judgments of voice pitch and loudness characteristics.

<table>
<thead>
<tr>
<th>TABLE 3.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual Characteristics Related to Pitch and Loudness in Some Types of Dysarthria</td>
</tr>
<tr>
<td>TYPE OF DYSARTHRIA</td>
</tr>
<tr>
<td>Ataxic</td>
</tr>
<tr>
<td>Flaccid</td>
</tr>
<tr>
<td>Hyperkinetic</td>
</tr>
<tr>
<td>Hypokinetic</td>
</tr>
<tr>
<td>Spastic</td>
</tr>
</tbody>
</table>

Source: Based on Darley, Aronson, and Brown’s (1975) classification.
analysis of $F_0$ and amplitude has revealed that patients with PD tend to show higher than normal $F_0$, lower standard deviations of frequency and amplitude, and decreased phonational and dynamic ranges (Gamboa et al., 1997). In addition, speakers with neurological problems are less able to use $F_0$ effectively to distinguish between declarative and interrogative sentences. Normal speakers in a study by LeDorze, Ouellet, and Ryalls (1994) showed an average difference of 83 Hz between the final syllables in the same utterance produced as a declarative and as an interrogative sentence. Subjects with neurological problems produced an average final syllable difference of only 25 Hz.

These kinds of acoustic data support the perceptual impressions of restricted pitch and loudness ranges, which are the common complaint of patients with PD and other neurological disorders. Furthermore, these measures have the added advantage of quantifying the precise degree of loss of frequency and intensity range compared to normal. Acoustic analysis of $F_0$ and amplitude can also serve to detect early changes in voice production due to neurological disease, even before such changes can be heard perceptually. This has been shown in patients suffering from amyotrophic lateral sclerosis (ALS). In this disease, the patient’s voice becomes progressively weaker. $F_0$ and amplitude levels decrease over time, and eventually the person becomes completely unable to produce voice and speech. Research has demonstrated that patients with ALS who sound perceptually normal have a much smaller frequency range than normal, 16.4 semitones compared to 22.7 semitones (Silbergleit, Johnson, & Jacobson, 1997). This smaller frequency range may be an indication of early signs of laryngeal weakness affecting the laryngeal muscles. Knowing that weakness is present despite the normally sounding voice, clinicians can offer intervention at early stages of the disease in order to maintain the patient’s vocal function for as long as possible.

Another important factor in ALS that has been revealed by acoustic analysis is that the vocal characteristics of patients are not uniform, but vary greatly from patient to patient. Measurement of mean $F_0$, standard deviation of $F_0$, plus $F_0$ and amplitude contours during sustained phonation and production of phrases show that, although changes in $F_0$ seem to be present consistently in patients with ALS, some speakers have lower than normal $F_0$ levels while others are higher than normal. In addition, some but not all individuals with ALS exhibit a reduced range of $F_0$ during connected speech (Strand, Buder, Yorkston, & Olson Ramig, 1994). This information is critical in planning effective therapy tailored to each individual’s particular voice and speech defects. With this knowledge, therapy procedures can be planned that focus on normalizing or maintaining vocal function as much as possible for the individual. Another benefit is that the same instrumentation used to obtain diagnostic data can then be used for visual feedback and motivation during therapy.
Clinical Study

Background
Jaime is a speech–language pathologist in an elementary school, which is part of a large school district. An important part of her job is to screen all entering kindergartners in the school district for any speech or language problems. She recently acquired a Visi-Pitch (Kay) and is eager to implement a new screening protocol that she developed for identifying possible voice disorders in the children.

Acoustic Measures
Jaime’s protocol includes several $F_0$ and intensity variables, including average $F_0$, SFF, and average intensity. For average $F_0$, the child is asked to hold the vowel /a/ for 5 seconds; SFF is taken from a picture description task. Average intensity is also taken from these activities. During these activities, Jaime also listens carefully to the children’s voices and rates them perceptually.

After screening 150 children (87 girls and 63 boys), she is able to obtain means for these different acoustic measures. Jaime did not think it was necessary to separate the data for the girls and boys and obtained all her information from the averages for all the children. The average $F_0$ for /a/ for all the children was approximately 248 Hz; SFF yielded an average of 242 Hz. Average intensity was 59 dB. Based on these normative values that she derived, Jaime was able to identify several children whose means fell outside of these ranges. For example, Peter obtained an $F_0$ of 224 Hz for /a/, 215 Hz for his SFF, and an average intensity of 77 dB. This was consistent perceptually with his low-pitched and somewhat loud voice. Emma’s values were 258 Hz for /a/, 254 Hz for SFF, and 56 dB for average intensity. Jaime was surprised about the $F_0$ measures, because perceptually, Emma’s voice sounded extremely low-pitched for her age. Emma was also hoarse and complained that her throat was sore. Both of these children were referred to an ENT for a laryngeal exam.

Clinical Questions
1. What are the possible advantages and disadvantages of using this kind of instrumentation in conjunction with traditional perceptual methods of screening?
2. What may be the advantages and/or disadvantages of Jaime’s using her own normative data to make comparisons, rather than using values from reported literature?
3. Should the values of these two children have been compared to the single group averages that Jaime derived from the screenings? Why or why not?
4. What could account for the lack of consistency between the perceptual and acoustic results in Emma’s case?

**Summary**

- Frequency measures typically used in clinical settings include average $F_0$, speaking $F_0$, $F_0$ range, and maximum phonational frequency range.
- Typically used amplitude measures include average amplitude, amplitude variability, and dynamic range.
- The voice range profile plots a person’s maximum phonational frequency range against his or her dynamic range at different frequencies and acts as a snapshot of phonatory behavior.
- Frequency and amplitude variables related to voice use are used in clinical situations to make diagnostic decisions, supplement perceptual judgments of voice, and assess outcomes of treatment.
- Neurologic diseases are often characterized by problems in the control of vocal frequency and amplitude.

**Review Exercises**

1. Identify three clinical situations in which the measurement of frequency and amplitude variables might be useful, and explain why.
2. Discuss the rationale for the development of separate average $F_0$ and $F_0$ variability norms for men, women, and children.
3. Explain why the voice range profile can be considered as a snapshot of phonatory behavior.
4. What are two major advantages in measuring frequency and/or amplitude variables in individuals with voice problems?
5. Explain how frequency and amplitude measures can serve to detect early laryngeal changes in neurological diseases.