Chapter

3

The Nature of Sound

"I'm sorry, you'll have to upgrade to first class if you want to use our ISDN line."

> **"Look with thine ears."**
> ***William Shakespeare's* King Lear *(Act 4, Scene 6)***

Speech technology revolves around the physics of sound. Speech is a particular type of sound and therefore inherits the universal characteristics of *sound* in a scientific sense. Sound is defined as "a vibration in an elastic medium at a frequency and intensity that is capable of being heard by the human ear" (Isaacs et al., 1996). The "elastic medium" we are most concerned with is air. A vibration, such as that caused by the movement of human vocal chords that produces speech, travels through the air, and reaches our ears by physically striking our eardrums. Our ears then send an electronic signal to the brain. That message is perceived as sound. Sound may also be received by a microphone and then electronically processed in any of several ways related to the development of speech software, such as recording or broadcasting. This chapter covers the nature of sound, including its measurement and properties.

Measurement of Sound

Sound travels at a speed of 1130 feet per second (called the speed of sound). It travels through the air in a manner similar to waves traveling across the ocean surface. As sound travels, it contracts and expands, disturbing the air pressure around it. The contractions are called compressions and the expansions are called rarefactions. This is because the contractions compress air particles while the expansions cause the air particles to be more rare in that physical space. These compressions and rarefactions form a sound wave, or *sine wave* named after the trigonometric function. The sine wave is typically represented as a graph showing increases and decreases in atmospheric pressure over time, as shown in Figure 3.1. The center line of the graph represents normal atmospheric pressure. Compressions, the increases in pressure, are called *peaks*. Rarefactions, the decreases in pressure, are called *troughs*. The distance between a peak and a trough is a single wavelength, also known as a *cycle*.

Frequency

The number of cycles a wave completes in one second is called its *frequency.* An example of a 4-cycle wave is shown in Figure 3.2. The shorter a wavelength is, the higher its frequency. The longer the wavelength is, the lower its frequency. Frequency is measured in *hertz,* which means cycles per second. This is abbreviated as Hz. One thousand hertz is one *kilohertz,* abbreviated kHz. The range of sound, or vibrations the human ear is capable of hearing, is between approximately 20 and 22,000 Hz. Vibrations below this range are called *infrasounds* and above it are called *ultrasounds*. The average human voice has a frequency of 5kHz.

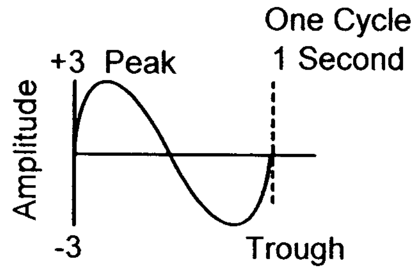

Figure 3.1 A wavelength.

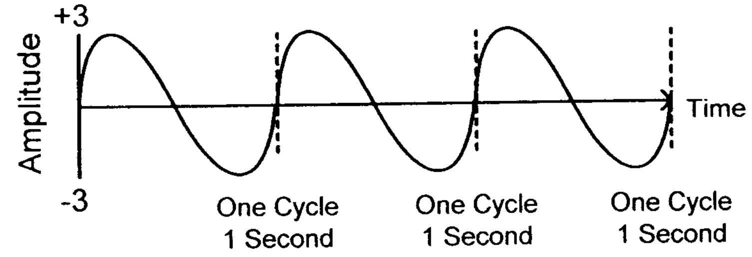

Figure 3.2 A 4-cycle wave.

Sounds Eponymous

The decibel is named after Alexander Graham Bell, the inventor of the telephone. Bell's contemporary Heinrich Rudolph Hertz was the first person to artificially produce electromagnetic waves and is honored by his own scientific unit of measurement.

The pitch of a sound can also be measured. *Frequency* and *pitch* are technically synonymous, but pitch is usually discussed in the context of sounds made by musical instruments and described in units other than kHz. The higher a sound wave's frequency is, the higher its pitch. The lower the frequency, the lower its pitch. From a standpoint of personal experience, humans understand pitch, such as low sounds and high sounds, more easily than frequency. However, computers and electronic devices don't typically think musically, so we need to understand frequency as well. Table 3.1 illustrates ranges of frequencies and the types of musical instruments you would hear in each range. (An audio equalizer for a home stereo is typically divided into similar frequency ranges.)

[redacted]	[redacted]
20-200 Hz	Bass, tuba
200 Hz-10kHz	Saxophones, violins, clarinets, trumpets, guitars, flutes, piccolos
10kHz-20kHz	Above the range of most musical instruments

Table 3.1 Frequency Range of Musical Instruments

Amplitude

Amplitude is the scientific term for the loudness of a sound. The loudness of a sound is determined by the size of its wavelength. In Figure 3.3, sine wave A has a larger wavelength than B; therefore, it is said to have greater amplitude and is perceived as being louder.

A large wavelength has a greater impact on air pressure than a smaller one. Human ears respond to this impact on air pressure by perceiving changes in the intensity of the sound pressure level (SPL). The sound pressure level is the range between the threshold of sound, or softest sound heard by the human ear, and the threshold of pain, or loudest sound we can hear without experiencing pain. However, scientific measurement of the SPL does not accurately reflect human perception of sound because we respond to changes logarithmically, not linearly. Therefore, a more reasonable measuring unit called the decibel, abbreviated dB, is typically used to describe amplitude.

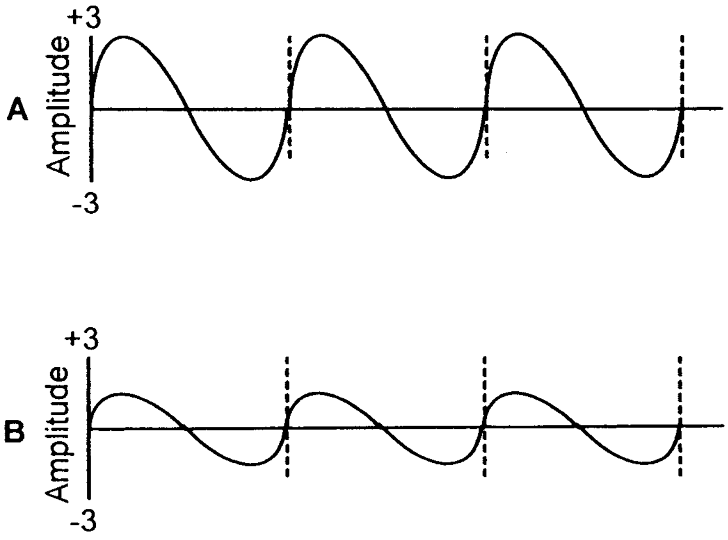

Figure 3.3 A comparison of amplitude.

The decibel is actually a ratio used to describe the change in loudness. Using decibels, the threshold of sound, or absolute quiet, is measured as 0 dB. The threshold of pain is about 130 dB. Table 3.2 lists different decibel levels and the sounds they are associated with.

Decibel Trivia

- A one decibel change in sound pressure is impossible for most people to detect.
- The average human hears loudness differences in 3 dB increments.
- To double the perceived loudness, the sound pressure level must be increased by 10 dB.

Adapted from Sound Check: The Basics of Sound and Sound Systems *by Tony Moscal*

Other Properties of Sound

In addition to frequency and amplitude, sine waves have several other properties that contribute to the human perception of sound. These properties include *phase, timbre,* and *envelope.*

Phase

When two sine waves occur simultaneously, they interact with one another and create a composite sound wave. If the two waves have exactly the same amplitude and frequency, the composite wave will be the sum of the amplitudes—in other words, twice the volume. For example, having two audio speakers instead of one provides double the volume. However, altering the frequency or amplitude of one sine wave would put them "out of phase."

[illegible]	
130 dB	Emergency siren; threshold of pain
120 dB	Jet airplane
110 dB	Jackhammer
100 dB	Heavy thunderstorm
90 dB	Motorcycle
80 dB	Lawn mower
70 dB	Heavy traffic
60 dB	Loud conversation
50 dB	Average office
40 dB	Average household
30 dB	Quiet conversation
20 dB	Whisper
10 dB	A pin dropping
0 dB	Absolute quiet; threshold of hearing

Table 3.2 Decibel Level of Certain Sounds

Timbre

The timbre of a sound is its tonal character, sometimes called its tone color. Most sounds are not pure, or single, sine waves. Most sounds are composite waves comprised of a *fundamental wave* (the sound's lowest frequency) and additional frequencies called *harmonics*. Additional frequencies change the human perception of tonal character and thus the perception of the overall sound. For example, consider a man singing a single musical note. His voice has a specific frequency. If he is joined by a woman singing the note in unison, the composite sound would have the same frequency, but the timbre would be perceived as different.

Envelope

Because sounds occur over time, or for only a specific duration, they are said to have an envelope. The envelope of a sound is a graph that describes the *attack, decay, sustain,* and *release* of the sound. The attack is the point in time when the sound begins. At that point in time, the sound rises from silence to its maximum volume. It then decays slightly and has a period of sustain where the volume is held before it finally releases back into silence. Compare the sound of the single honking of a car horn to that of an ambulance siren. The two sounds could have a similar sine wave, but very different envelopes.

How Humans Hear Sound

Since the definition of sound that we are working with is any "vibration" that "is capable of being heard by the human ear" we should have some familiarity with how sound is actually heard. As discussed earlier in this chapter, any sound wave with a frequency between approximately 20 and 22,000 Hz that strikes our ears is perceived as sound. We previously said that a sound wave reaches human ears and physically strikes the eardrums. Then, our ears send an electronic signal to the brain and the message is perceived as sound. We can delve more deeply into this process by discussing the physiology of hearing.

Anatomically, the human ear has three parts: the outer ear, the middle ear, and the inner ear. The process of hearing begins when a sound wave is physically captured by the outer part of the human ear.

The outer, or external, portion of the ear is shaped in such a way that it funnels sound waves further into the ear. The hole in our outer ear is the beginning of the auditory canal. When a sound wave travels through the auditory canal, it resonates and the sound pressure level is increased by many decibels. At the end of the auditory canal is the eardrum, which marks the beginning of the middle ear.

The eardrum, also called the tympanic membrane, is a flexible membrane that vibrates when struck by a sound wave. A series of three bones, called the ossicles, touches the inside of the tympanic membrane. The ossicles vibrate in sympathetic response to the vibrations of the tympanic membrane. These small bones move in conjunction with one another in a manner similar to levers or engine pistons and carry the sound vibration to the inner ear. The muscles attached to the ossicles function as a guardian to the inner ear. They flex in response to loud, potentially harmful sounds and reduce the amplitude of the sound by as much as 20 dB. However, because this flexing motion has latency of approximately 35 to 150 milliseconds, the ear may not be able to protect itself against extreme explosive or percussive sounds.

The inner ear, formally called the cochlea, is a snail-shaped organ filled with fluid. When the ossicles vibrate, they force the fluid of the cochlea through the coil. This movement transfers vibrations through a membrane called the basilar membrane and the organ of Corti. The organ of Corti contains nerve endings which convert the vibrations into an electronic signal. The electronic signal is a neural impulse, which is carried to the brain by the auditory nerve.

Summary

Any vibration in the air that is capable of being heard by human ears is defined as sound in a scientific sense. In human-computer interaction we are concerned with the impact of speech sounds as well as non-speech sounds. In this chapter we looked at the nature of sound. In the next chapter we look at the nature of language.

WILEY

Designing Effective Speech Interfaces

Susan Weinschenk
Dean T. Barker