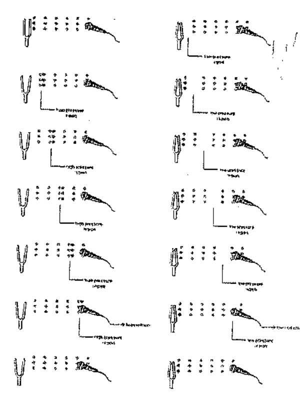
CHAPTER 4

Fundamentals of Digital Audio

TABLE OF CONTENTS



CHAPTER

4

KEY CONCEPTS

- Sound wave
- Frequency and pitch
- Digitizing sound
- Sampling rate and bit depth of digital audio
- Nyquist's theorem
- Dynamic range
- Audio file size optimization
- MIDI

GENERAL LEARNING OBJECTIVES

In this chapter, you will learn

- The common terms in audio.
- The properties of sound waves.
- The basic steps of digitization—sampling and quantization—in digital audio.
- The effects of sampling rate and bit depth on digital audio quality.
- The meaning of *decibel*.
- The common file types for digital audio.
- The general strategies for reducing digital audio file sizes.
- The difference between digital audio recording and MIDI.

4.1 INTRODUCTION

Sound is an integral part of our everyday sensory experience. It is also an analog phenomenon. As is the case with most of our everyday sensory experiences, we seldom think of sound in its microscopic sense. However, in order to understand digital audio, you need to recognize the fundamental nature of sound as waves—the physics of sound. In this chapter we will explain and interpret the graphical depiction of sound as a ***waveform***. A waveform serves as a means for us to "see" the information that we hear by providing quantitative properties of a sound, such as its amplitude and frequency.

4.2 THE NATURE OF SOUND WAVES

Sound is a wave that is generated by vibrating objects in a medium such as air. The vibrating objects can be the vocal cords of a person, a guitar string, or a tuning fork. Plucking a guitar string in the air causes the string to move back and forth. This creates a disturbance of the surrounding air molecules. When the string moves in one direction, it causes the air molecules to compress into a smaller space, raising the air pressure slightly in that region. The air molecules under higher pressure in that region then push the other air molecules

surrounding them, and so fo
it creates a gap between the
region, causing the surround
air molecules propagates, ra
air pressure—forming a sour
it causes the eardrums to mo
ognizes the changing air pre
the movement of the string,
in turn, causes higher pressu
pressure region of the wave.
is higher and the sound is lo

SOUND AS A MEC

Because the propagation of
ticle interactions, a sound w
of this property is that a sou
The motion of a particle
type of wave is characterize
particles that propagates, no

If you place a microphon
change will be detected by th
changes of pressure in the pro
as changes of electrical sign
cally with the changes in air
(Figure 4.1). The vertical axis
trical signals caused by the so
A vibrating guitar string
ing periodic changes of air p
wave reaching the recorder a
sound wave can be represent

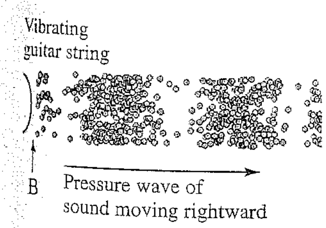


Figure 4.1 A pressure sound

surrounding them, and so forth. When the vibrating string moves in the reverse direction, it creates a gap between the string and the molecules. This lowers the air pressure in that region, causing the surrounding air molecules to move into that area. The displacement of air molecules propagates, radiating away from the string and causing periodic changes of air pressure—forming a sound wave. When this compression wave reaches your eardrums, it causes the eardrums to move back and forth. This sends a signal to your brain, which recognizes the changing air pressure as a sound. The harder you pluck the string, the greater the movement of the string, which causes greater displacement of the air molecules. This, in turn, causes higher pressure in the high-pressure region and lower pressure in the low-pressure region of the wave. When the string is plucked harder, the amplitude of the sound is higher and the sound is louder.

SOUND AS A MECHANICAL WAVE

Because the propagation of a sound wave in a medium relies on the mechanism of particle interactions, a sound wave is characterized as a mechanical wave. The implication of this property is that a sound wave does not propagate in a vacuum.

The motion of a particle in a sound wave is parallel to the direction of the wave. This type of wave is characterized as a longitudinal wave. Notice that it is the *motion* of the particles that propagates, not the particles themselves.

If you place a microphone in the path of the sound wave, the periodic air-pressure change will be detected by the recorder and converted into varying electrical signals. The changes of pressure in the propagating sound wave reaching the recorder are thus captured as changes of electrical signals over time. The sound wave can be represented graphically with the changes in air pressure or electrical signals plotted over time—a waveform (Figure 4.1). The vertical axis of the waveform represents the relative air pressure or electrical signals caused by the sound wave. The horizontal axis represents time.

A vibrating guitar string in the air causes the string to move back and forth, causing periodic changes of air pressure. The changes of pressure in the propagating sound wave reaching the recorder are captured as changes of electrical signals over time. The sound wave can be represented graphically with the changes in air pressure or electrical

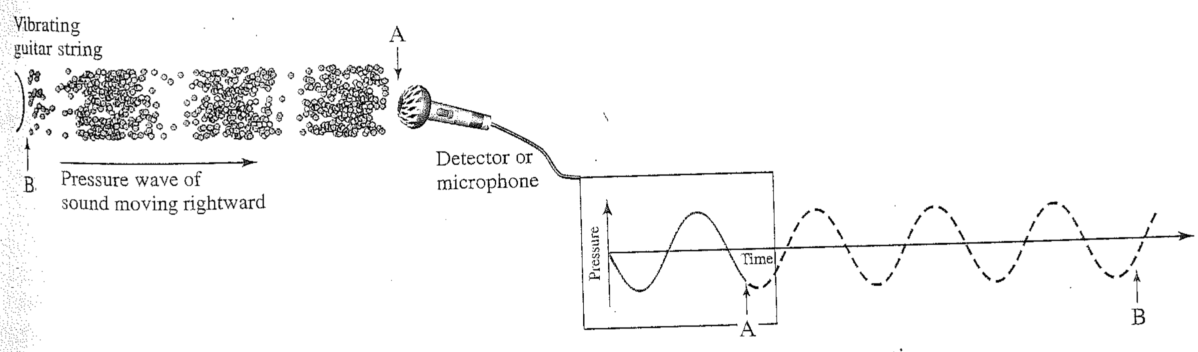


Figure 4.1 A pressure sound wave from a guitar string and its graphical representation.

signals plotted over time—a waveform. Be careful *not* to interpret the waveform as a representation of the sound wave in *space*. The picture on the left in Figure 4.1 is a picture of the air molecules in space. On the right is a graph over *time*. The air-pressure information at point B is *not* where time is zero on the waveform graph. Instead, the pressure information of point B has not yet propagated to the microphone; it would have been after point A has been recorded—about three more cycles to the right of point A on the waveform.

A waveform is a graphical representation of the pressure–time (not pressure-space) fluctuations of a sound wave. A sound wave propagates in space. The waveform matches the pressure changes in time at a fixed location. The crests correspond to the high pressure (compression of air molecules), and troughs correspond to the low pressure (rarefaction). The horizontal axis is time. However, looking at an illustration of a longitudinal pressure wave of sound placed side by side with its waveform can mislead you to think of the horizontal axis of the waveform as distance if you are not careful. Remember that the horizontal axis of a waveform is time, not distance. Let's re-emphasize two key points about a waveform:

Sound as a Pressure Wave An illustration of a sound wave as a longitudinal air-pressure wave.

1. Be careful *not* to interpret sound wave as a wave that has crests and troughs, as in a transverse wave. Sound wave is a longitudinal wave, in which the particles of the medium (such as air molecules) move back and forth, not up and down, in the direction of the wave propagation.
2. Be careful *not* to interpret the waveform as a representation of the sound wave in *space*. Instead, the waveform graph represents the pressure changes over *time*.

Besides visualization of the pressure oscillation of the sound wave over time, a waveform can also give us information about the pitch and loudness of the sound. The following sections discuss how these two properties are measured and derived from the waveform.

4.2.1 Frequency and Pitch

A sound wave is produced by a vibrating object in a medium, say air. No matter what the vibrating object is, it is vibrating or moving back and forth at a certain frequency. This causes the surrounding air molecules to vibrate at this same frequency, sending out the sound-pressure wave. The ***frequency*** of a wave refers to the number of complete back-and-forth cycles of vibrational motion of the medium particles per unit of time. The common unit for frequency is ***Hertz (Hz)*** where the unit of time is 1 second.

$$1 \text{ Hz} = 1 \text{ cycle/second}$$

The period of a wave is the time for a complete back-and-forth cycle of vibrational motion of the medium particles. Shown in Figures 4.2a and b are two simple sine wave waveforms. If the tick mark on the horizontal axis marks the first second, then the frequency of the wave in Figure 4.2a has a frequency of 2 Hz, because it completes two cycles within 1 second. In Figure 4.2b, the wave has a frequency of 4 Hz.

Two notes that are an octave apart correspond to sound waves whose frequencies are in a ratio of 2:1.

Sound frequency is related to the ***pitch*** of the sound. Higher frequencies correspond to higher pitches. Generally speaking, the human ear can hear sound ranging from 20 Hz to 20,000 Hz.

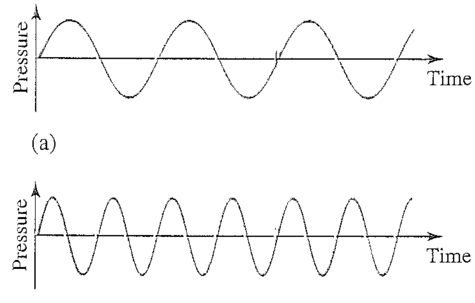


Figure 4.2 Simple waveforms representing two different frequencies: (a) lower frequency, (b) higher frequency.

4.2.2 Sound Intensity and Loudness

Sound intensity is related to the perceived ***loudness*** of a sound, but the two are not exactly the same. Sound intensity is often measured in ***decibels (dB)***. A decibel is based on a ratio of a louder sound to a softer one. By definition,

$$\text{Number of decibels} = 10 \times \log (I_1/I_{\text{ref}})$$

(Equation 1)

where I_1 and I_{ref} are the two sound intensity values in comparison.

$$\text{Number of decibels} = 20 \times \log (V_1/V_{\text{ref}})$$

(Equation 2)

where V_1 and V_{ref} are the magnitudes of two electrical voltages or currents in comparison.

Notice that the decibel is not an absolute unit. It is an expression of a ratio of two values. More precisely, it is a logarithm of the ratio of two values. The general implication of this is that doubling the sound intensity means an increase of about 3 decibels. Or, a louder sound that causes twice the magnitude of the electric voltages or currents as a softer sound is about 6 decibels higher than the softer one. Why 3 and 6? Let's plug some numbers into the previous equations.

Say that you have a sound whose pressure wave produces an electrical signal V_1, and this pressure is double the pressure of some reference sound (V_{ref}). This means

$$V_1 = 2 \times V_{\text{ref}}$$

Plugging this relationship into Equation 2, we get:

$$\begin{aligned}\text{Number of decibels} &= 20 \times \log (2 \times V_{\text{ref}}/V_{\text{ref}}) \\ &= 20 \times \log (2) \\ &\cong 20 \times 0.3 \\ &= 6\end{aligned}$$

Similarly, plugging numbers into Equation 1 for doubling the sound intensity gives you 3 decibels. It may seem that this explanation presents more mathematics than you really need to know in order to work with digital audio editing programs. However, in many audio editing programs, the audio amplitude is measured in decibels. In addition, 3 and 6 decibels are given as preset values in amplification filters. Understanding what decibels mean and their relationship to audio signals helps you create predictable results in audio editing.

DECIBELS AND BELS

The unit called a *bel* was defined by scientists at Bell Labs to compare two power values. The unit was named after Alexander Graham Bell. By definition,

$$\text{Number of bels} = \log (P_1/P_0)$$

where P_1 and P_0 are the two power values in comparison. For sound, these can be considered the sound intensity.

A decibel (dB) is 1/10th of a bel (i.e., 1 bel equals 10 decibels.) So

$$\text{Number of decibels} = 10 \times \log (P_1/P_0)$$

Because power equals voltage times current, this relationship leads to the following (which we present without going into the mathematical derivation).

$$\text{Number of decibels} = 20 \times \log (V_1/V_0)$$

where V_1 and V_0 are the two voltage or amplitude values in comparison.

The threshold of hearing is the minimum sound-pressure level at which humans can hear a sound at a given frequency. It varies with frequency. Generally, 0 dB refers to the threshold of hearing at 1,000 Hz. Note that 0 dB does not mean zero sound intensity or the absence of a sound wave.

The threshold of pain is about 120 decibels, representing a sound intensity that is 1,000,000,000,000 (or 10^{12}) times greater than that of 0 decibel.

LOUDNESS VERSUS SOUND INTENSITY

The loudness of a sound is a subjective perception, but sound intensity is an objective measurement. Thus, loudness and sound intensity are not exactly the same properties.

To measure loudness, a 1,000-Hz tone is used as a reference tone. The volume of the reference tone is adjusted until it is perceived by listeners to be equally as loud as the sound being measured. Sound intensity, on the other hand, can be measured objectively by auditory devices independent of a listener.

The age of the listener is one of the factors that affect the subjective perception of a sound. The frequency of the sound is also a factor because of the human ear's sensitivity to different sound frequencies. The loudness of sound (as perceived by human ears) is only roughly proportional to the logarithm of sound intensity. However, in general, the higher the sound intensity, the louder the sound is perceived.

4.3 ADDING SOUND WAVES

A simple sine wave waveform represents a simple single tone—single frequency. When two or more sound waves meet, their amplitudes add up, resulting in a more complex waveform (Figure 4.3). The sound we perceive every day is seldom a single tone. The waveforms representing speech, music, and noise are complex waveforms that result from adding multiple waveforms of different frequencies. For example, Figure 4.4 shows a waveform of the spoken word "one."

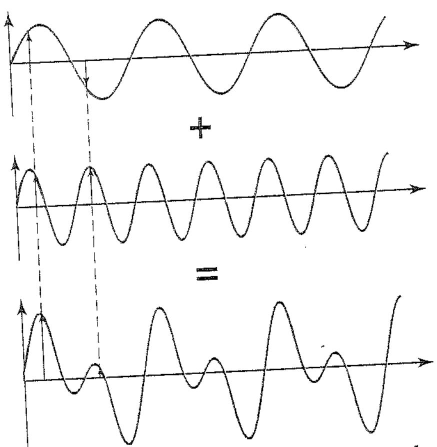
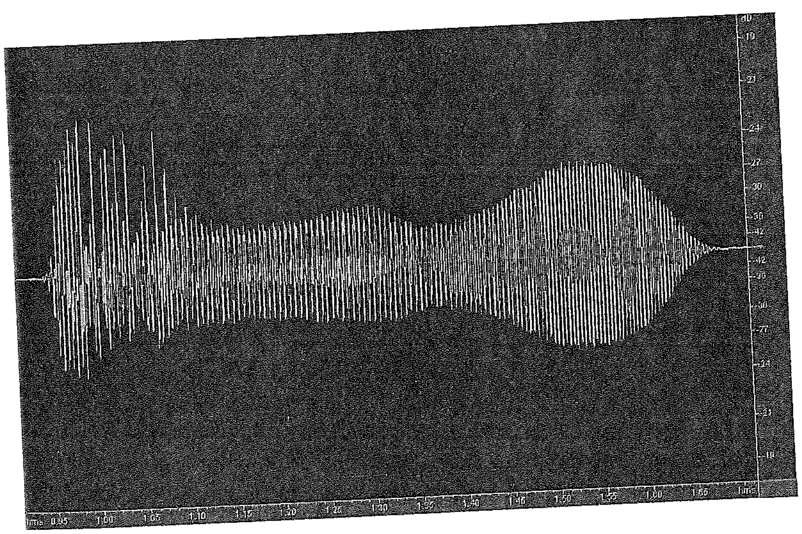
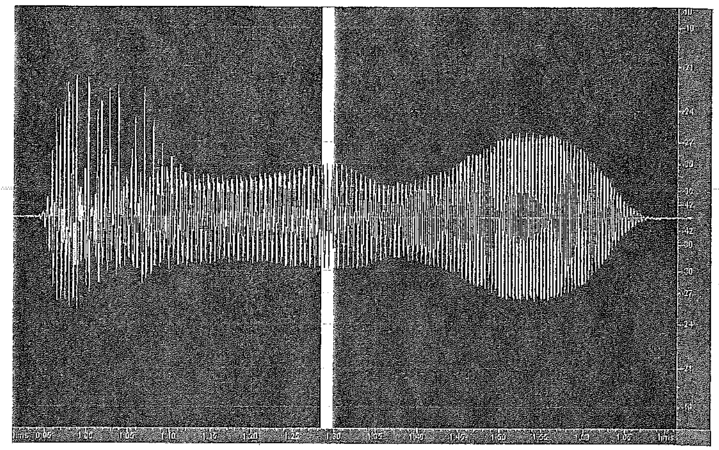


Figure 4.3 Addition of two simple sine wave waveforms results in a more complex waveform.

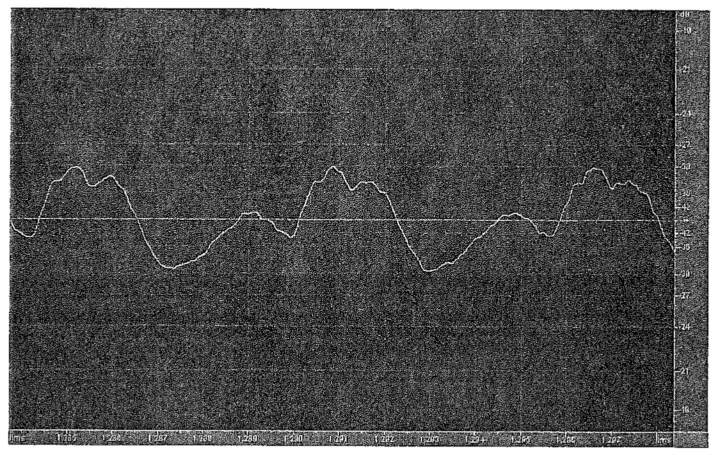


(a)

Figure 4.4 (a) A waveform of the spoken word "one." *(continued)*



(b)



(c)

Figure 4.4 *(continued)* (b) The highlighted segment to be magnified. (c) The zoomed-in view of the segment of the waveform.

The mathematical basis of the Fourier transform is discussed in the CS Module.

DECOMPOSING SOUND

When we record a sound, such as the spoken word "one" in Figure 4.4, the waveform recorded is a complex one. Can a complex wave be decomposed into its simple component parts—the different sine waves that make up the complex wave? Yes! One of the mathematical methods to accomplish this decomposition is called the Fourier transform.

But why would you want to decompose a complex wave into simple sine waves? The frequency of a simple sine wave can be determined easily. When you want to remove certain sounds that can be characterized by a range of frequencies, such as low-pitched noise, you can apply filters using the Fourier transform to selectively remove these unwanted sounds from a complex sound. These filters are available in many digital audio-processing programs and are used for breaking down a sound to remove unwanted frequencies.

4.4 DIGITIZING SOUND

A sound wave is an analog phenomenon. In a sound wave, the amplitude changes continuously over time. Like the process of digitizing any analog information, the process of digitizing sound involves sampling and quantizing an analog sound wave.

4.4.1 Step 1: Sampling

In the *sampling* step, the sound wave is sampled at a specific rate into discrete samples of amplitude values. The higher the sampling rate, the higher the accuracy in capturing the sound. However, a higher sampling rate will generate more sample points, thus requiring more storage space and processing time. To give you a feel for the sampling rate for digital sound, the sampling rate for CD-quality audio is 44,100 Hz (i.e., 44,100 samples per second).

> **Sampling and Quantizing in Digital Audio** An interactive tutorial illustrating the concepts of sampling and quantizing in digitizing a sound wave.

To keep the illustration simple and clear, our examples will use very low sampling rates. These rates are too low to be practical for digitizing sound in real life, but they are simple enough to demonstrate the point. Figure 4.5a represents a continuous sound wave signal. When we digitize the sound wave, we take discrete samples. Figure 4.5b shows a sampling rate of 10 Hz (that is, taking 10 samples of the pressure per second). Figure 4.5c shows a simple reconstruction of the wave by keeping the pressure value a constant between sample points. As you see, changes—crests and troughs—between any two sample points are missed.

> The common reconstruction of an analog wave from the discrete sample points is not done by keeping the sample values constant between sample points, as shown in Figure 4.5c. Instead, it is usually done by interpolation of the sample points using mathematical algorithms to regenerate a smooth curve. However, no matter what technique is used to reproduce the sound wave, the number of sample points, and thus the sampling rate, is the limiting factor for the accuracy of the reconstruction.

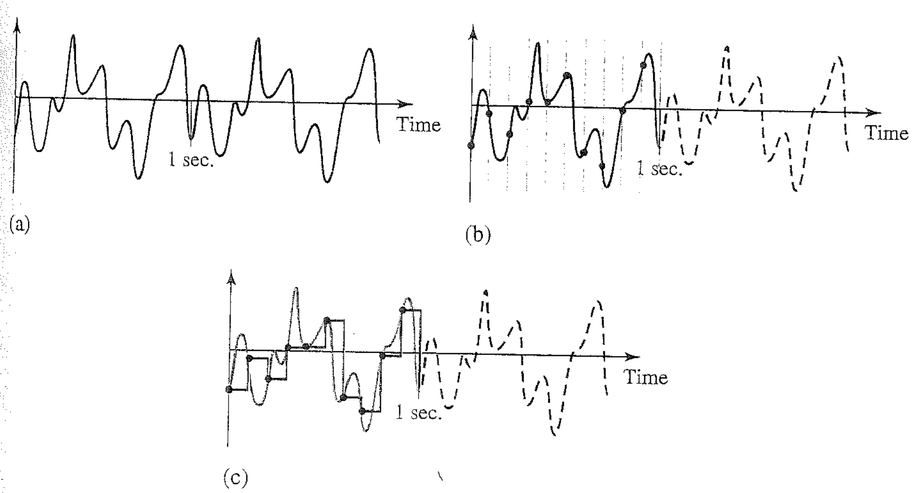


Figure 4.5 (a) A theoretical continuous sound wave signal. (b) Ten samples of the pressure are taken per second—a sampling rate of 10 Hz. (c) A simple reconstruction by keeping the pressure value a constant between sample points.

What if we raise the sampling rate to 20 Hz? Figure 4.6a shows a sampling rate of 20 Hz, and Figure 4.6c shows a simple reconstruction of the wave. The reconstructed wave now looks closer to the original wave than the one sampled at 10 Hz does. As you see, a higher sampling rate can increase the accuracy of the reproduction of a sound wave.

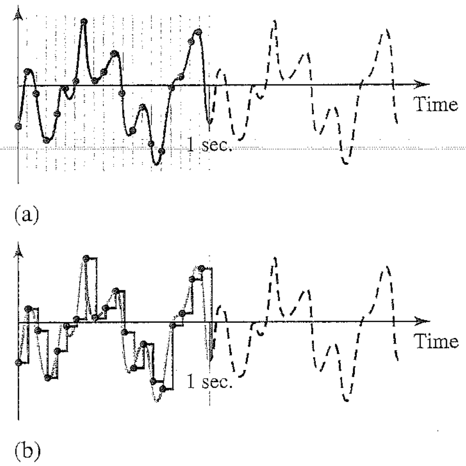


Figure 4.6 (a) A sampling rate of 20 Hz. (b) A simple reconstruction of the sound wave.

Because a higher sampling rate will produce more sample points, the file size of the resulting file will be larger. As in the 10-Hz example, changes—crests and troughs—between the sample points are still missed. This is an inherent limitation of the discreteness of digital data. No matter how high the sampling rate you adopt, there will always be missed points between those discrete sample points.

SAMPLING RATE VERSUS AUDIO FREQUENCY

Be careful not to confuse the sampling rate with the audio frequency. Both the sampling rate and the audio frequency are measured in Hertz (Hz), but they are not the same thing. The audio frequency relates to the pitch of the sound. The higher the frequency, the higher the pitch. The sampling rate refers to the number of samples taken per second for a sound wave. The sampling rate is a characteristic of the digitization process, but the audio frequency describes the sensory characteristics of the sound we perceive. The same sampling rate may be used for digitizing sounds of different audio frequencies. Conversely, the same sound may be digitized using different sampling rates to create different digital audio files.

4.4.2 Step 2: Quantizing

In the ***quantizing*** step, each of the discrete samples of amplitude values obtained from the sampling step will be mapped and rounded to the nearest value on a scale of discrete levels. Therefore, the more levels available in the scale, the higher the accuracy in reproducing the sound. For digital audio, having more levels means higher ***resolution***. However, higher resolution will require more storage space. The number of levels in the scale is expressed in ***bit depth***—the power of 2. For example, an 8-bit audio allows $2^8 = 256$ possible levels in the scale. To give you a feel of the bit depth for digital sound, CD-quality audio is 16-bit (i.e., $2^{16} = 65{,}536$ possible levels in quantizing the samples of amplitude values).

For demonstration purposes, we use 3-bit as an example, which is a scale of eight discrete levels. Again, 3-bit is not a practical bit depth used in real-life applications. The sampled data are mapped to the nearest level on the scale (Figure 4.7). Some samples may deviate more from their original amplitudes. With a low bit depth, data with different original amplitudes may be quantized onto the same level—for example, note the quantized sample points of the last six sample points in Figure 4.7b.

Audio Examples (a Short Musical Sound and a Sound Effect) with Various Combinations of Sampling Rate and Bit Depth Audio examples of a short musical sound and an explosion sound effect with various sampling rates (44,100 Hz, 22,050 Hz, and 11,025 Hz) and bit depths (8 bit and 16 bit) are used to demonstrate the impact of sampling rate and bit depth on (i) the audio quality and (ii) the file size. Review questions on the impact of sampling rate and bit depth, the audio's waveforms, and frequency spectral views are also available.

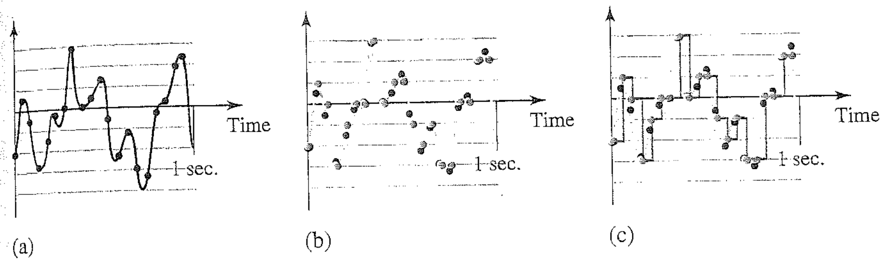


Figure 4.7 Quantizing with 3-bit resolution.

4.5 DYNAMIC RANGE

In the quantization step, a scale of discrete levels of amplitude values is used to map the sample points. The range of the scale, from the lowest to highest possible quantization values in the scale, defines the ***dynamic range*** for digitizing audio. In the quantization example of the previous section, where a scale of eight levels (3-bit samples) is used, the lowest level of the scale is placed at about the lowest amplitude of the sound wave and the highest level at about the highest point of the sound wave. The remaining six levels are equally spaced in between these two levels. This scale is extended to include the highest- and lowest-amplitude values of the sound wave. That is, none of the sample points is outside of this range. The scale, shown in Figure 4.8, covers a full amplitude range of the sound wave.

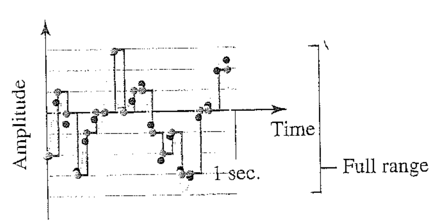


Figure 4.8 Quantization using a scale range equal to the full amplitude range of the sound wave.

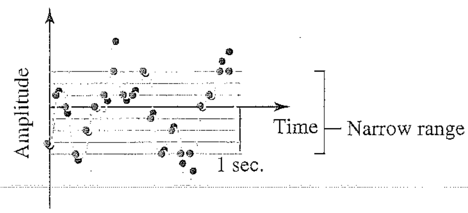


Figure 4.9 Quantization using a scale range narrower than the full amplitude range of the sound wave.

If the dynamic range is smaller than the full amplitude range of the sound wave (Figure 4.9), some data will be lost. The digitized sound wave will be "chopped" off at the limit of the range, causing clipping of the sound wave. Clipping is an undesirable effect because of loss of data. The clipped amplitude values are not recoverable. However, with a reduced dynamic range, the accuracy can be improved for the data within the range. This is an advantage, especially if most of the sample points are within a smaller middle region of the range. By sacrificing a small number of highest- and lowest-amplitude values, the accuracy of the majority of the sample points can be improved. In the simple example shown in Figure 4.9, you see that (with the same bit depth) reducing the dynamic range actually allows more subtle changes to be distinguishable (Figure 4.10). Instead of being quantized to the same value, some sample points can now be set apart on different levels.

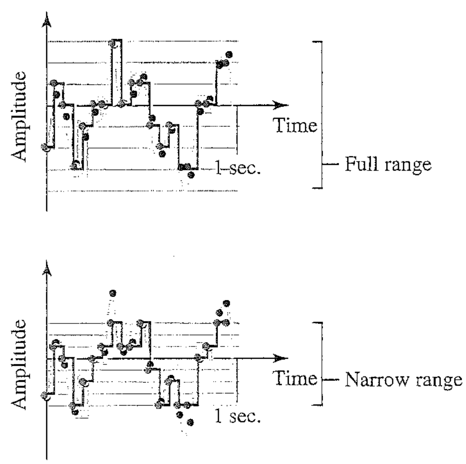


Figure 4.10 The quantized data (shaded in blue) in which subtle changes become more distinguishable in the reduced dynamic range than in the full range.

The dynamic range of a sound system refers to the range of the highest and lowest sound levels that the system can reproduce. In such context, a larger dynamic range is considered better because the sound system can then allow you to input audio components with wider ranges, and it can output the audio with minimal distortion.

What if you extend the dynamic range to more than the amplitude range of the sound wave? Then you will get the opposite result of reducing the dynamic range—the accuracy will be lost. As you see in Figure 4.11, the amplitude of the sound wave is now within only six levels of the range. So, although this extended dynamic range has eight quantization levels available, only six levels are utilized for this sound wave.

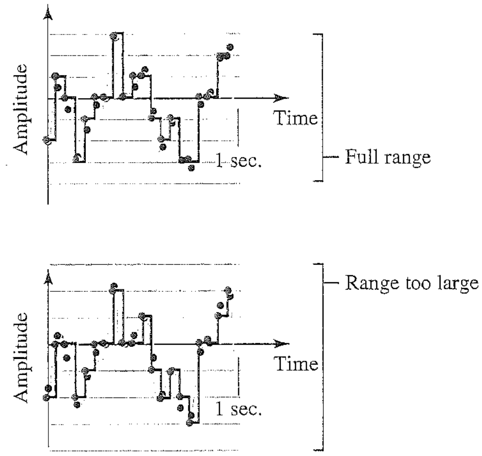


Figure 4.11 Comparing quantization using two ranges of scale: the full amplitude range of the sound wave, and one wider than the full amplitude range.

4.6 FILE SIZE, FILE COMPRESSION, AND FILE TYPES OF DIGITAL AUDIO

Higher sampling rate and bit depth always deliver better fidelity of a digitized file—whether it is an image or audio. So the recurring question may come up again: If higher sampling rate and bit depth give better fidelity and less degradation, then why shouldn't we always choose the maximum possible values? Well, the answer to this question is always the file size. Large file sizes require more storage space, longer processing time, and longer transfer time. Especially if the digital media files are created for use on the Internet, the network transfer time of a file is often a more important consideration than the storage space, which is becoming less expensive. The larger the file size, the longer it will take to download to another person's computer.

In addition to the sampling rate and bit depth, do not forget that the duration of the audio also affects the file size. Audio is a time-based medium. Audio files (such as speech and music) often require long, continuous playback. Therefore, the file size increases very rapidly with a higher sampling rate and bit depth. Let's look at an example of a 1-minute CD-quality stereo audio. A stereo audio has two ***channels***—a left channel and a right channel. The bit depth of CD-quality audio is 16 bits per channel—16 bits per sample for each channel. Its sampling rate is 44.1 kHz—that is, 44,100 samples per second. The number of bits rquired to store a 1-minute CD-quality stereo audio is calculated as follows:

$$1 \text{ minute} \times 60 \text{ seconds/minute} = 60 \text{ seconds}$$

$$60 \text{ seconds} \times 44{,}100 \text{ samples/second} = 2{,}646{,}000 \text{ samples}$$

$$2{,}646{,}000 \text{ samples} \times 16 \text{ bits/sample} = 42{,}336{,}000 \text{ bits}$$

Because stereo audio has two channels, the total bit size $= 42{,}336{,}000$ bits $\times\, 2 =$ 84,672,000 bits.

To convert the bit size into bytes,

$$84{,}672{,}000 \text{ bits/(8 bits/byte)} = 10{,}584{,}000 \text{ bytes} \cong 10 \text{ MB}$$

As you see, it requires about 10 MB for each minute of CD-quality audio. An hour of such audio requires about 600 MB. At the time of writing, the average download speed of the 4G wireless connection is in the range of 2–6 mbps (megabits per second) for laptop modems and 1–2 mbps for smartphones.[1] The time it takes to download this 1-minute audio from the Web at 2 mbps would be

$$84{,}672{,}000 \text{ bits}/(2{,}000{,}000 \text{ bits/second}) \approx 42 \text{ seconds}$$

With 1 mbps on some smartphones, the time would be about 84 seconds, or almost 1.5 minutes. To reduce the download time, you may want to reduce the file size of the audio. As you may recall from Chapter 1, the three general strategies of reducing a digital media file size are reducing the sampling rate, reducing bit depth, and applying file compression. As you see in the example, the equation for the audio file size calculation can be expressed as follows:

$$\text{File size} = \text{duration} \times \text{sampling rate} \times \text{bit depth} \times \text{number of channels}$$

The equation for calculating audio file size includes an additional factor—the number of channels. Therefore, by examining the equation of audio file size calculation, you see four general strategies of reducing a digital audio file's size:

- Reduce the sampling rate
- Reduce the bit depth
- Apply compression
- Reduce the number of channels

Note that duration of the audio is also a factor in the equation for file size calculation. We do not include reduction of audio duration in the listed strategies because in most situations you want to keep the duration of the audio. However, keep in mind that the duration of the audio has a direct effect on the file size. You may want to remove the unnecessary silent segments in an audio file where possible. For example, most voice-over or speech contains unnecessary silence or pauses that may be removed to reduce the audio duration and thus the file size.

- **Reducing the number of channels.**
  Stereo audio has two channels. If you reduce a stereo audio file to mono—which becomes one channel—then you reduce the file size in half. This may suit speech and short sound effects for online games. Your decision to reduce the audio from stereo to mono is very much dependent on the nature of your project. Reducing a channel causes a noticeable difference unless your final product is expected to be listened to with a mono-speaker.
- **Reducing the sampling rate.**
  Reducing the sampling rate and bit depth sacrifices the fidelity of the digitized audio, which means it will sound less like the original. However, when working with digital media files, you often have to weigh the quality against the file size. When you do so, you often need to take into consideration both human perception of the medium and how you're going to use the audio.

  First, let's consider that the human ear can hear sound ranging from approximately 20 Hz to 20,000 Hz. The range varies with individuals and their ages. Not all people can hear the two ends of the average range. The human ear is most sensitive in the range of about 2,000 to 5,000 Hz, not the two ends of the range.

  Second, according to a rule called ***Nyquist's theorem***, we must sample at least two points in each sound wave cycle to be able to reconstruct the sound wave satisfactorily.

[1] Mark Sullivan. "4G Wireless Speed Tests: Which Is Really the Fastest? AT&T, Sprint, T-Mobile, and Verizon: PCWorld's exclusive performance tests reveal which 4G network delivers the fastest data speeds." *PCWorld*. March 13, 2011. URL: http://www.pcworld.com/printable/article/id,221931/printable.html

In other words, the sampling rate of the audio must be at least twice that of the audio frequency—called a ***Nyquist rate***. Therefore, a higher-pitch sound requires a higher sampling rate than a lower-pitch sound. In reality, the sound we hear, such as music and speech, is made up of multiple-frequency components. Then the sampling rate can be chosen as twice the highest-frequency component in the sound in order to reproduce the audio satisfactorily. Reducing the sampling rate sacrifices the sound quality of the higher frequencies.

The most common sampling rates you may encounter in a digital audio editing program are:

11,025 Hz: AM radio quality/speech
22,050 Hz: Near FM radio quality (common for multimedia projects)
44,100 Hz: CD quality
48,000 Hz: DAT (digital audio tape) quality
96,000 Hz: DVD audio quality
192,000 Hz: DVD audio quality

Based on the human hearing range and the Nyquist theorem, the sampling rate for CD-quality audio of 44.1 kHz is reasonable. But can we reduce the sampling rate, and if so, to what extent? Because the human ear is most sensitive in the range of about 2 to 5 kHz, then 11,025 Hz and 22,050 Hz seem to be reasonable sampling rates. 11,025 Hz causes more noticeable degradation to music than to speech because music has more higher-frequency components than speech. Because the human voice in speech normally does not exceed 5 kHz, the sampling rate setting of 11,025 Hz is reasonable for speech. Depending on the nature of your final product, a sampling rate setting of 11,025 Hz may also suit short sound effects (such as breaking glass and explosions), which may not require high fidelity.

So, how much can we lower the file size by reducing the sampling rate? From the file size equation, the file size can be reduced in the same proportion as the reduction of the sampling rate. In the example of 1 minute of CD-quality stereo audio, lowering the sampling rate from 44.1 kHz to 22.05 kHz will reduce the file size from about 10 MB to about 5 MB.

- **Reducing the bit depth.**
  The most common bit-depth settings you may encounter in a digital audio editing program are 8 bit and 16 bit. According to the file size equation, lowering the bit depth from 16 to 8 reduces the file size in half. In the example of a 1-minute CD-quality stereo audio, lowering the bit depth from 16 to 8 will reduce the file size from about 10 MB to about 5 MB. Suppose the sampling rate of the audio has been lowered from 44.1 kHz to 22.05 kHz, creating a 5-MB file. Lowering the bit depth of this file from 16 to 8 will reduce the file size further, from 5 MB to 2.5 MB.

  Eight-bit resolution is usually sufficient for speech. However, for music, 8-bit resolution is too low to accurately reproduce the sound satisfactorily. Our ears usually can notice the degradation in playback. Typically, 16-bit resolution is used for music.
- **Applying file compression.**
  An audio file can be compressed to reduce the audio file size. Compression can be lossless or lossy. Lossy compression gets rid of some data, but human perception is taken into consideration so that the data removed causes the least noticeable distortion. For example, the popular audio file format MP3 uses a lossy compression that gives a good compression rate while preserving the *perceptible* quality of the audio by selectively removing the least perceptible frequency components of the audio. Keep in mind that a file compressed with a lossy compression method should not be used as a source file for further editing. To achieve the best result in editing an audio file, you should always start with a source file that is uncompressed or compressed with lossless compression.

No matter which file size reduction strategies you want to apply to your audio file, you should always evaluate the acceptability of the audio quality of the reduced file based on the nature of your audio project and weigh the quality against the file size limitation. The intended use of your final audio dictates the acceptability of trade-offs.

Even working within a limitation of file size, you should record or digitize with at least CD or DAT quality and then apply file size reduction strategies later, instead of digitizing at a lower sampling rate and bit depth in the first place. One reason is that compression algorithms are often designed to selectively remove data that will cause minimal impact on the audio quality perceivable by the human ear. Another reason is that by keeping a higher sampling rate and bit depth, you will have a choice of optimizing the file size by a combination of the file optimization strategies. You can weigh the file size reduction against the quality at that point. However, if you do not digitize at a high enough sampling rate or bit depth, you will get stuck with an unsatisfactory audio.

Many digital audio editing programs let you choose the format in which you want to save your audio file. The common file types are listed in Table 4.1. Some file types already dictate the compression option to be used. Others allow you to choose whether you want to compress the file, and possibly the compression options.

Generally, the intended use of your audio file determines the file format. You should take into consideration the following factors:

The information in Section 6.10 on streaming video and progressive download also apply to audio.

HTML5 video and audio are covered in Chapter 15.

- **The file size limits.**
  If your audio is intended to be used on the Web, you may want to consider a file format that offers high compression or even a streaming audio file format to minimize the wait time to play your audio. Streaming audio means the audio will be played while it is being downloaded. The audio file does not have to be downloaded in its entirety before it can be played. However, it requires Internet connection during its playback.
- **The intended audience of your audio file.**
  How is your target audience going to listen to your audio? What equipment do they have? If they are listening on a computer, what is the operating system? If your audio will be played on multiple platforms, then the file should be a cross-platform format. If you want the audio to be played on a Web page without having to use external plug-ins, you may want to use HTML5 audio.
- **Keeping the file as a source file.**
  If you are keeping the file for future editing, then you should choose a file format that is uncompressed or uses lossless compression.

CLOUD COMPUTING FOR VIDEO AND AUDIO DELIVERY

The term "cloud" refers to the Internet and cloud computing refers to the technologies that provide computing services (such as storage, software, database access) to users via the Internet. For online storage, you can download your files onto your devices whenever needed. The downloaded video and audio files can be played back *from your devices*. For video and audio files, cloud-based service providers often also support streaming the media—the media is played back *over the Internet*. For example, Amazon Cloud Drive offers online storage and supports downloading and streaming of music files that are saved on your Cloud Drive. Apple iCloud lets you stream the music files stored on your iCloud and download them on any of your devices.

TABLE 4.1 Common Audio File Types

File Type	Acronym For	Originally Created By	File Information and Type of Compression	Platforms and Additional Information
.aiff	Audio Interchange File Format	Apple, adopted later by Silicon Graphics	Usually not compressed, but has a compressed version	Primarily for Mac OS; also supported on Windows
.wav		IBM and Microsoft	• Supports uncompressed and a number of different compression formats • One of the HTML5 audio formats	• Primarily for Windows, but can be used on other systems • Plays in Web browsers that support the .wav format of HTML5 audio; at the time of writing, the supported browsers are Firefox, Safari, Chrome, and Opera
.au and .snd	Also called μ-law or Sun μ-law format	Sun and NeXT	μ-law encoding compresses the file at a ratio of 2:1; slow decompression	Sun, Unix, or Linux operating system
.mp3	MPEG audio layer 3	Moving Pictures Experts Group	• Good compression rate with perceivably high-quality sound • One of the HTML5 audio formats	• Cross-platform; many digital audio players can play it • Plays in Web browsers that support the MP3 format of HTML5 audio; at the time of writing, it is supported by Safari, Internet Explorer (IE), and Chrome, but Chrome is dropping future support of MP3
.m4a	MPEG-4 format without the video data	Moving Pictures Experts Group	• AAC compression; same compression as the MPEG-4 H.264 without the video data • One of the HTML5 audio formats	Plays in Web browsers that support the AAC format of HTML5 audio; at the time of writing, it is supported by Safari, IE, and Chrome
.ogg or .oga		Xiph.Org Foundation	• Usually referred to as Ogg Vorbis format • Ogg is a container format • The audio codec is Vorbis	Plays in Web browsers that support the Ogg Vorbis format of HTML5 audio; at the time of writing, it is supported by Firefox, Chrome, and Opera
.mov	QuickTime movie	Apple	• Not just for video • Supports audio track and a MIDI track • Supports a variety of sound compressors • Files can be streamed with QuickTime Streaming Server • "Fast Start" technology also allows users to listen to media as it is being downloaded	Cross-platform; requires QuickTime player
.wma	Windows Media Audio	Microsoft		
.asf	Advanced streaming format	Microsoft	Proprietary compression algorithm	Primarily used with Windows Media Player

4.7 MIDI

So far in this chapter, we have been describing the digital audio of captured analog audio that is produced by the vibration of objects in air. The continuous fluctuation of pressure in a propagating pressure wave is captured by a device or microphone, then digitized by sampling and quantization. The audio can be speech, music, noise, or a combination of these.

There is another method of storing music information—in ***MIDI*** format. MIDI (Musical Instrument Digital Interface) is a communications protocol, not a physical object. It defines the common interface for electronic digital musical instruments to communicate with computers or other instruments or devices containing microprocessors. It specifies the configurations of cables and cable plugs and the format of the data.

Many electronic keyboards have built-in synthesizers. A MIDI keyboard looks like a small piano, but upon receiving a signal such as a key being hit, its electronic device synthesizes sound using its own internal microprocessor (i.e., computer). Computers also can be attached directly to a MIDI keyboard to capture the musical notes being played. There are also software programs that let you enter the notes directly via the computer's mouse and keyboard. The composed music also can be played through a MIDI keyboard that has a synthesizer.

MIDI signals are not digitized audio sample points but contain note *information* played with a virtual instrument. Such information includes the instrument being played, the note being played, the duration of the note, and how loud to play the note. Unlike the digital audio we have been describing in the previous sections of this chapter, MIDI music creation does not involve capturing and digitizing analog sound waves or any analog information. Therefore, it does not involve sampling and quantization; that is, there is no sampling rate or bit-depth option in creating and editing a MIDI file.

Compared to the sampled audio files, the MIDI format has both advantages and disadvantages. Its file size can be much more compact. For example, the file size of a 1-minute MIDI file can be about 2 KB, but, as you have seen in the file size calculations, a 1-minute, stereo, 16-bit audio with a sampling rate of 44.1 kHz is about 10 MB. If you convert this MIDI file into a 44.1 kHz, 16-bit, stereo digital audio file, it will become about 10 MB. MIDI music can be easily edited like sheet music by changing the notation, timing, and instruments.

The composed music of a MIDI file requires a synthesizer to play. The quality of the sound depends on the quality of the synthesizer that plays back the composed music. This can be a disadvantage because the actual sound produced by different synthesizers may differ even for the same note of the same instrument. The composed music you hear from your MIDI keyboard may not sound the same on your friend's MIDI keyboard or computer synthesizer. For example, QuickTime has its own selection of high-quality MIDI instruments and thus can play MIDI without an external MIDI synthesizer. But it sounds different from a full-fledged synthesizer.

The analogy for MIDI music versus digitized audio may be a cake recipe versus the baked cake. It is much easier and lighter to send a cake recipe to your friends. But how exactly the cake tastes and looks may vary depending on the exact ingredients your friends use—even the same temperature setting varies from oven to oven. On the other hand, a baked cake is more bulky to mail to your friends. But it ensures the cake's taste and look that you intend the recipient to experience.

4.8 SUMMARY

Sound is a wave that is generated by vibrating objects in a medium such as air. The vibrating objects can be the vocal cords of a person, a guitar string, or a tuning fork. An object vibrating in air creates a disturbance of the surrounding air molecules, causing periodic changes of air pressure—forming a sound wave.

No matter what the vibrating object is, the object is vibrating or moving back and forth at a certain frequency. This causes the surrounding air molecules to vibrate at this same frequency. The common unit for frequency is Hertz (Hz); 1 Hz refers to 1 cycle per second. The sound frequency is related to the pitch of the sound. Higher frequencies correspond to higher pitches.

Sound intensity is related to, but not exactly the same as, the perceived loudness of a sound. The loudness of a sound is a subjective perception, but sound intensity is an objective measurement. Sound intensity is often measured in decibels (dB). A decibel is based on a ratio of a louder sound to a softer one; it is not an absolute measurement.

When two or more sound waves meet, their amplitudes add up, resulting in a more complex waveform. The waveforms of the sound we perceive every day (such as speech, music, and noise) are complex waveforms that result when multiple waveforms of different frequencies are added together.

Like the process of digitizing any analog information, the process of digitizing a sound wave involves sampling and quantizing an analog sound wave. In the sampling step, the sound wave is sampled at a specific rate into discrete samples of amplitude values. The higher the sampling rate, the higher the accuracy in capturing the sound. But a high sampling rate will generate more sample points, which will require more storage space and processing time. In the quantizing step, each of the discrete samples of amplitude values obtained from the sampling step will be mapped to the nearest value on a scale of discrete levels. Therefore, the more levels available in the scale, the higher the accuracy in reproducing the sound. For digital audio, having more levels means higher resolution. But higher resolution requires more storage space for the same reason that higher bit depth for any digital media will increase the file size.

Audio files, such as speech and music, usually require long, continuous playback. Therefore, the file size increases very rapidly with higher sampling rate and bit depth. To reduce the file size, there are four general file optimization approaches: reduce the sampling rate, reduce the bit depth, apply compression, and reduce the number of channels. No matter which file size reduction strategies you want to apply to your audio file, you should always evaluate the acceptability of the audio quality of the reduced file based on the nature of your audio project, and weigh the quality against the file size limitation. The intended use of your final audio dictates the consideration of the acceptable trade-offs.

Even working within a limitation of file size, you will get better results by recording or digitizing at a sampling rate and bit depth that produce good audio quality and then apply file size reduction strategies later.

Another method of storing music information is in MIDI format. MIDI stands for Musical Instrument Digital Interface. MIDI files do not contain digitized audio sample points. Instead, they contain information about musical notes to be played with a virtual instrument. Such information includes the instrument being played, the note being played, the duration of the note, and how loud to play the note. Unlike digitized audio, MIDI music creation does not involve digitizing analog sound waves or any analog information. Therefore, there are no sampling rate and bit depth settings. MIDI music has the advantage of very small file size. Another advantage is that the music content can be edited easily. The disadvantage is that the sound depends on the synthesizer that plays back the composed music. Thus, you do not know how your MIDI file sounds on your target audience's devices.

TERMS

bit depth 116
channels 119
decibels (dB) 111
dynamic range 117
frequency 110
Hertz (Hz) 110
loudness 111
MIDI 124
Nyquist rate 121
Nyquist's theorem 120
pitch 110
quantizing 116
resolution 116
sampling 115
sound intensity 111
waveform 108

LEARNING AIDS

The following learning aids can be found at the book's companion Web site.

Sound as a Pressure Wave
An illustration of a sound wave as a longitudinal air-pressure wave.

Sampling and Quantizing in Digital Audio
An interactive tutorial illustrating the concepts of sampling and quantizing in digitizing a sound wave.

Audio Examples (a Short Musical Sound and a Sound Effect) with Various Combinations of Sampling Rate and Bit Depth
Audio examples of a short musical sound and an explosion sound effect with various sampling rates (44,100 Hz, 22,050 Hz, and 11,025 Hz) and bit depths (8 bit and 16 bit) are used to demonstrate the impact of sampling rate and bit depth on (i) the audio quality and (ii) the file size. Review questions on the impact of sampling rate and bit depth, the audio's waveforms, and frequency spectral views are also available.

REVIEW QUESTIONS

When applicable, please select all correct answers.

1. A sound with higher ______ is perceived to have a higher pitch.
   A. volume
   B. frequency
   C. fidelity
   D. sampling rate
   E. bit depth

2. The unit used for measuring ______ is Hertz (Hz).
   A. amplitude
   B. frequency
   C. sampling rate
   D. bit depth
   E. dynamic range

3. A waveform i
   A. pressure–t
   B. space–tim
   C. pressure–s

4. The horizonta
   A. pressure
   B. distance
   C. time

5. The vertical a
   A. pressure
   B. distance
   C. time

6. **True/False:**

7. The ______ c
   A. amplitude
   B. frequency
   C. sampling r
   D. bit depth
   E. dynamic ra

8. The ______ o
   stored.
   A. amplitude
   B. frequency
   C. sampling r
   D. bit depth
   E. dynamic ra

9. In digital audi
   A. amplitude
   B. frequency
   C. sampling r
   D. bit depth
   E. dynamic ra

10. In digital audi
    A. amplitude
    B. frequency
    C. sampling r
    D. bit depth
    E. dynamic ra

11. How many lev

12. How many lev

13. Generally, the