Analog Representations of Sound

Just as air pressure varies according to sound waves, so can the electrical quantity called *voltage* in a wire connecting an amplifier with a loudspeaker. We do not need to define voltage here. For the purposes of this chapter, we can simply assume that it is possible to modify an electrical property associated with the wire in a fashion that closely matches the changes in air pressure.

An important characteristic of the time-varying quantities we have introduced (air pressure and voltage) is that each of them is more or less exactly analogous to the other. A graph of the air pressure variations picked up by a microphone looks very similar to a graph of the variations in the loudspeaker position when that sound is played back. The term "analog" serves as a reminder of how these quantities are related.

Figure 1.12 shows an analog audio chain. The curve of an audio signal can be inscribed along the groove of a traditional phonograph record, as shown in figure 1.12. The walls of the grooves on a phonograph record

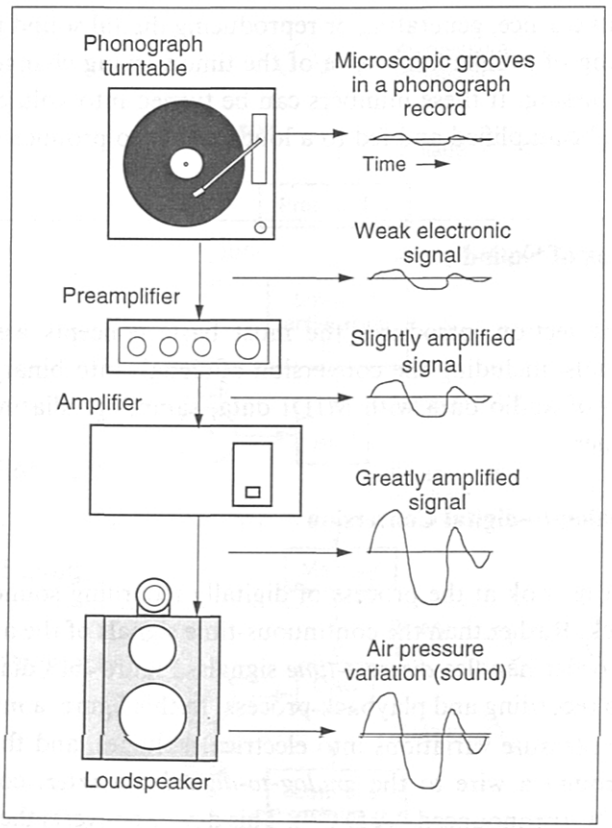

Figure 1.12 The analog audio chain, starting from an analog waveform transduced from the grooves of a phonograph record to a voltage sent to a preamplifier, amplifier, loudspeaker, and projected into the air.

contain a *continuous-time* representation of the sound stored in the record. As the needle glides through the groove, the needle moves back and forth in lateral motion. This lateral motion is then changed into voltage, which is amplified and eventually reaches the loudspeaker.

Analog reproduction of sound has been taken to a high level in recent years, but there are fundamental limitations associated with analog recording. When you copy an analog recording onto another analog recorder, the copy is never as good as the original. This is because the analog recording process always adds noise. For a *first-generation* or original recording, this noise may not be objectionable. But as we continue with three or four generations, making copies of copies, more of the original recording is lost to noise. In contrast, digital technology can create any number of generations of perfect (noise-free) clones of an original recording, as we show later.

In essence, generating or reproducing digital sound involves converting a string of numbers into one of the time-varying changes that we have been discussing. If these numbers can be turned into voltages, then the voltages can be amplified and fed to a loudspeaker to produce the sound.

Digital Representations of Sound

This section introduces the most basic concepts associated with digital signals, including the conversion of signals into binary numbers, comparison of audio data with MIDI data, sampling, aliasing, quantization, and dither.

Analog-to-digital Conversion

Let us look at the process of digitally recording sound and then playing it back. Rather than the continuous-time signals of the analog world, a digital recorder handles *discrete-time* signals. Figure 1.13 diagrams the digital audio recording and playback process. In this figure, a microphone transduces air pressure variations into electrical voltages, and the voltages are passed through a wire to the *analog-to-digital converter*, commonly abbreviated ADC (pronounced "A D C"). This device converts the voltages into a string of *binary numbers* at each period of the sample clock. The binary numbers are stored in a digital recording medium—a type of memory.

Binary Numbers

In contrast to decimal (or *base ten*) numbers, which use the ten digits 0–9, binary (or *base two*) numbers use only two digits, 0 and 1. The term *bit* is an abbreviation of *binary digit*. Table 1.1 lists some binary numbers and their decimal equivalents. There are various ways of indicating negative numbers in binary. In many computers the leftmost bit is interpreted as a sign indicator, with a 1 indicating a positive number, and a 0 indicating a negative number. (Real decimal or *floating-point* numbers can also be represented in binary. See chapter 20 for more on floating-point numbers in digital audio signal processing.)

The way a bit is physically encoded in a recording medium depends on the properties of that medium. On a digital audio tape recorder, for example, a 1 might be represented by a positive magnetic charge, while a 0 is indicated by the absence of such a charge. This is different from an analog

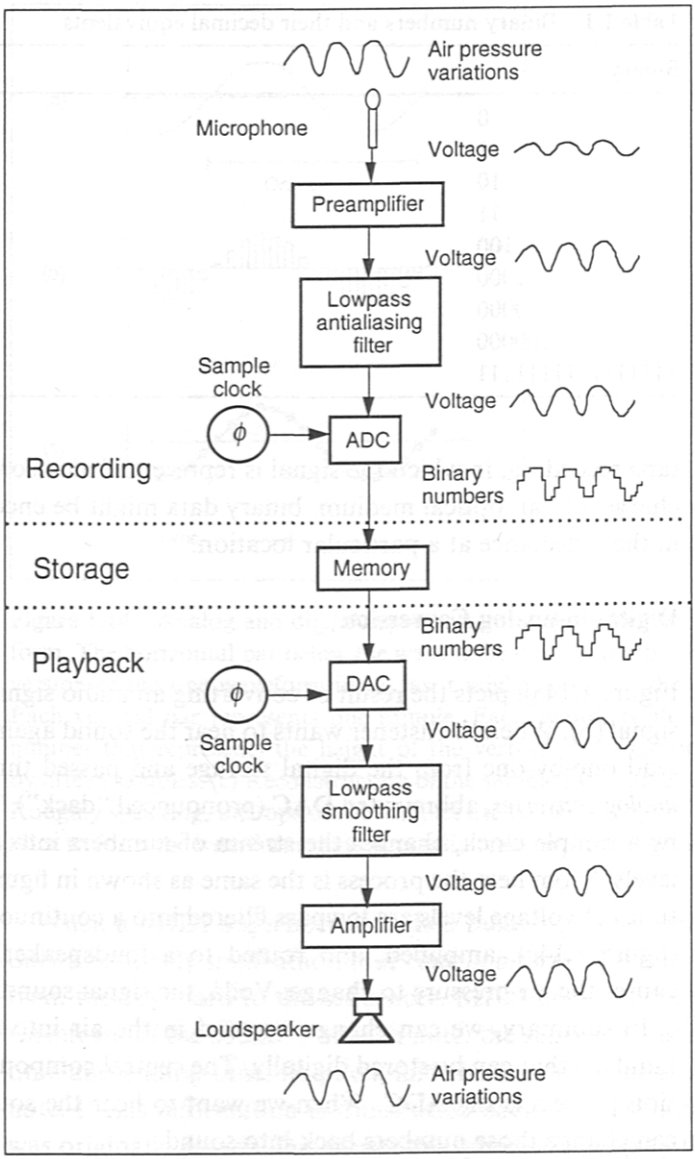

Figure 1.13 Overview of digital recording and playback.

Table 1.1 Binary numbers and their decimal equivalents

Binary	Decimal
0	0
1	1
10	2
11	3
100	4
1000	8
10000	16
100000	32
1111111111111111	65535

tape recording, in which the signal is represented as a continuously varying charge. On an optical medium, binary data might be encoded as variations in the reflectance at a particular location.

Digital-to-analog Conversion

Figure 1.14 depicts the result of converting an audio signal (a) into a digital signal (b). When the listener wants to hear the sound again, the numbers are read one-by-one from the digital storage and passed through a *digital-to-analog converter*, abbreviated DAC (pronounced "dack"). This device, driven by a sample clock, changes the stream of numbers into a series of voltage levels. From here the process is the same as shown in figure 1.13; that is, the series of voltage levels are lowpass filtered into a continuous-time waveform (figure 1.14c), amplified, and routed to a loudspeaker, whose vibration causes the air pressure to change. Voilà, the signal sounds again.

In summary, we can change a sound in the air into a string of binary numbers that can be stored digitally. The central component in this conversion process is the ADC. When we want to hear the sound again, a DAC can change those numbers back into sound.

Digital Audio Recording versus MIDI Recording

This final point may clear up any confusion: the string of numbers generated by the ADC are not related to MIDI data. (MIDI is the Musical Instrument Digital Interface specification—a widely used protocol for control of digital music systems; see chapter 21.) Both digital audio recorders and MIDI sequencers are digital and can record multiple "tracks," but they differ in the amount and type of information that each one handles.

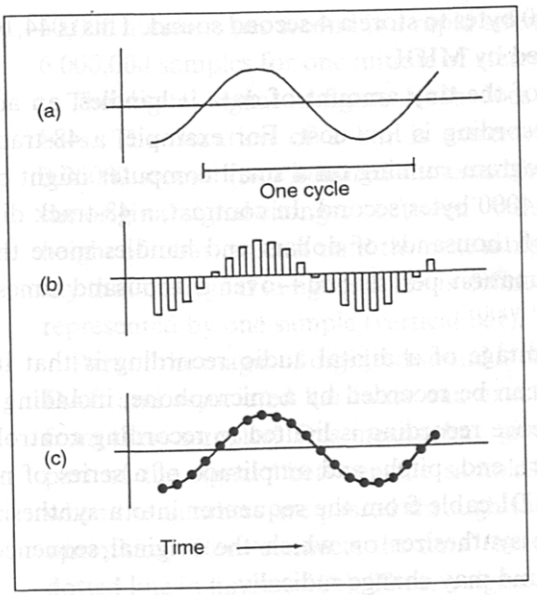

Figure 1.14 Analog and digital representations of a signal. (*a*) Analog sine waveform. The horizontal bar below the wave indicates one period or cycle. (*b*) Sampled version of the sine waveform in (*a*), as it might appear at the output of an ADC. Each vertical bar represents one sample. Each sample is stored in memory as a number that represents the height of the vertical bar. One period is represented by fifteen samples. (*c*) Reconstruction of the sampled version of the waveform in (*b*). Roughly speaking, the tops of the samples are connected by the lowpass smoothing filter to form the waveform that eventually reaches the listener's ear.

When a MIDI sequencer records a human performance on a keyboard, only a relatively small amount of *control information* is actually transmitted from the keyboard to the sequencer. MIDI does not transmit the sampled waveform of the sound. For each note, the sequencer records only the start time and ending time, its pitch, and the amplitude at the beginning of the note. If this information is transmitted back to the synthesizer on which it was originally played, this causes the synthesizer to play the sound as it did before, like a piano roll recording. If the musician plays four quarter notes at a tempo of 60 beats per minute on a MIDI synthesizer, just sixteen pieces of information capture this 4-second sound (four starts, ends, pitches, and amplitudes).

By contrast, if we record the same sound with a microphone connected to a digital audio tape recorder set to a sampling frequency of 44.1 KHz, 352,800 pieces of information (in the form of audio samples) are recorded for the same sound (44,100 × 2 channels × 4 seconds). The storage requirements of digital audio recording are large. Using 16-bit samples, it takes

over 700,000 bytes to store a 4-second sound. This is 44,100 times more data than is stored by MIDI.

Because of the tiny amount of data it handles, an advantage of MIDI sequence recording is low cost. For example, a 48-track MIDI sequence recorder program running on a small computer might cost less than $100 and handle 4000 bytes/second. In contrast, a 48-track digital tape recorder costs tens of thousands of dollars and handles more than 4.6 Mbytes of audio information per second—over a thousand times the data rate of MIDI.

The advantage of a digital audio recording is that it can capture any sound that can be recorded by a microphone, including the human voice. MIDI sequence recording is limited to recording control signals that indicate the start, end, pitch, and amplitude of a series of note events. If you plug the MIDI cable from the sequencer into a synthesizer that is not the same as the synthesizer on which the original sequence was played, the resulting sound may change radically.

Sampling

The digital signal shown in figure 1.14b is significantly different from the original analog signal shown in figure 1.14a. First, the digital signal is defined only at certain points in time. This happens because the signal has been *sampled* at certain times. Each vertical bar in figure 1.14b represents one *sample* of the original signal. The samples are stored as binary numbers; the higher the bar in figure 1.14b, the larger the number.

The number of bits used to represent each sample determines both the noise level and the amplitude range that can be handled by the system. A compact disc uses a 16-bit number to represent a sample, but more or fewer bits can be used. We return to this subject later in the section on "quantization."

The rate at which samples are taken—the *sampling frequency*—is expressed in terms of samples per second. This is an important specification of digital audio systems. It is often called the *sampling rate* and is expressed in terms of Hertz. A thousand Hz is abbreviated 1 KHz, so we say: "The sampling rate of a compact disc recording is 44.1 KHz," where the "K" is derived from the metric term "kilo" meaning thousand.

Dynamic Range of Digital Audio Systems

The specifications for digital sound equipment typically specify the accuracy or *resolution* of the system. This can be expressed as the number of bits that the system uses to store each sample. The number of bits per sample is important in calculating the maximum *dynamic range* of a digital sound system. In general, the dynamic range is the difference between the loudest and softest sounds that the system can produce and is measured in units of *decibels* (dB).

Decibels

The decibel is a unit of measurement for relationships of voltage levels, intensity, or power, particularly in audio systems. In acoustic measurements, the decibel scale indicates the ratio of one level to a *reference level*, according to the relation

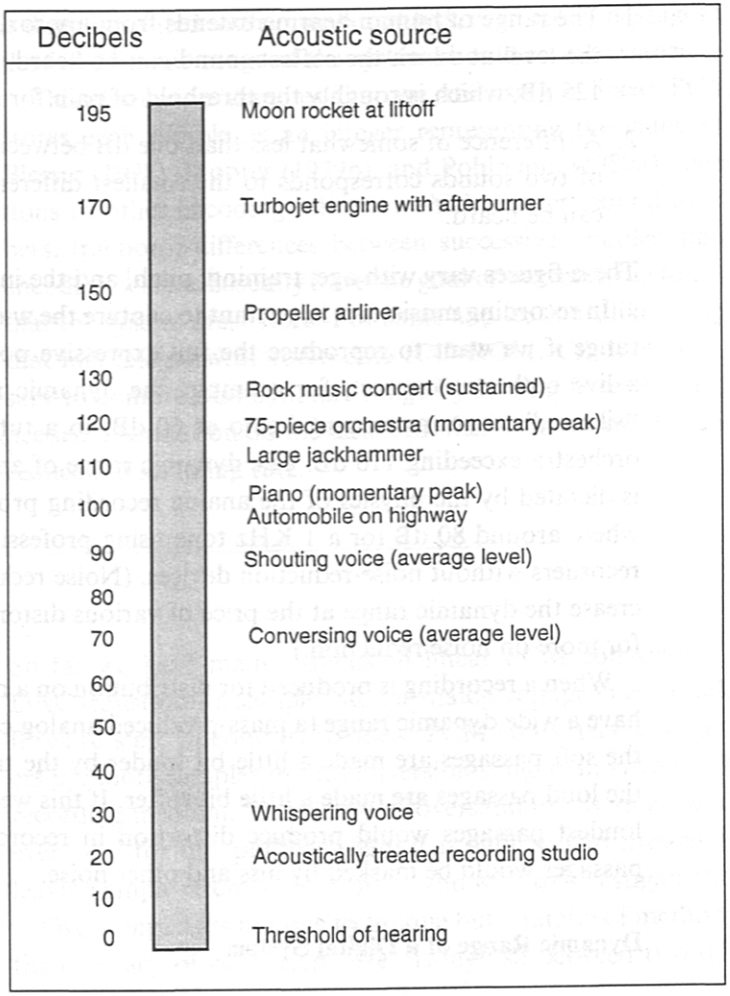

Figure 1.22 Typical acoustic power levels for various acoustic sources. All figures are relative to 0 dB = 10^{-12} watts per square meter.

$$\textit{number of decibels} = 10 \times \log_{10}(\textit{level}/\textit{reference level})$$

where the *reference level* is usually the threshold of hearing (10^{-12} watts per square meter). The logarithmic basis of decibels means that if two notes sound together, and each note is 60 dB, the increase in level is just 3 dB. A millionfold increase in intensity results in a 60 dB boost. (See chapter 23, Backus 1977, or Pohlmann 1989 for more on decibels.)

Figure 1.22 shows the decibel scale and some estimated acoustic power levels relative to 0 dB. Two important facts describe the dynamic range requirements of a digital audio system:

1. The range of human hearing extends from approximately 0 dB, roughly the level at which the softest sound can be heard, to something around 125 dB, which is roughly the threshold of pain for sustained sounds.
2. A difference of somewhat less than one dB between the amplitude levels of two sounds corresponds to the smallest difference in amplitude that can be heard.

These figures vary with age, training, pitch, and the individual.

In recording music, it is important to capture the widest possible dynamic range if we want to reproduce the full expressive power of the music. In a live orchestra concert, for example, the dynamic range can vary from "silence," to an instrumental solo at 60 dB, to a tutti section by the full orchestra exceeding 110 dB. The dynamic range of analog tape equipment is dictated by the physics of the analog recording process. It stands somewhere around 80 dB for a 1 KHz tone using professional reel-to-reel tape recorders without noise-reduction devices. (Noise reduction devices can increase the dynamic range at the price of various distortions. See chapter 10 for more on noise reduction.)

When a recording is produced for distribution on a medium that does not have a wide dynamic range (a mass-produced analog cassette, for example), the soft passages are made a little bit louder by the transfer engineer, and the loud passages are made a little bit softer. If this were not done, then the loudest passages would produce distortion in recording, and the softest passages would be masked by hiss and other noise.

Dynamic Range of a Digital System

To calculate the maximum dynamic range of a digital system, we can use the following simple formula:

maximum dynamic range in decibels = *number of bits* × 6.11.

The number 6.11 here is a close approximation to the theoretical maximum (van de Plaasche 1983; Hauser 1991); in practice, 6 is a more realistic figure. A derivation of this formula is given in Mathews (1969) and Blesser (1978).

Thus, if we record sound with an 8-bit system, then the upper limit on the dynamic range is approximately 48 dB—worse than the dynamic range of analog tape recorders. But if we record 16 bits per sample, the dynamic range increases to a maximum of 96 dB—a significant improvement. A 20-bit converter offers a potential dynamic range of 120 dB, which corresponds roughly to the range of the human ear. And since quantization noise

is directly related to the number of bits, even softer passages that do not use the full dynamic range of the system should sound cleaner.

This discussion assumes that we are using a linear PCM scheme that stores each sample as an integer representing the value of each sample. Blesser (1978), Moorer (1979b), and Pohlmann (1989a) review the implications of other encoding schemes, which convert sound into decimal numbers, fractions, differences between successive samples, and so on. Other encoding schemes usually have the goal of reducing the total number of bits that the system must store. For some applications, like compact disc media that mix images with audio data (CD-ROM, CD-I, etc.), it may be necessary to compromise dynamic range by storing fewer bits in order to fit all needed information on the disk. Another way to save space is, of course, to reduce the sampling rate.