
1

Digital Audio Concepts

with John Strawn

Background: History of Digital Audio Recording

- Experimental Digital Recording**
- Digital Sound for the Public**
- Digital Sound for Musicians**
- Digital Multitrack Recording**

Basics of Sound Signals

- Frequency and Amplitude**
- Time-domain Representation*
- Frequency-domain Representation*
- Phase**
- Importance of Phase*

Analog Representations of Sound

Digital Representations of Sound

- Analog-to-digital Conversion**
- Binary Numbers**
- Digital-to-analog Conversion**
- Digital Audio Recording versus MIDI Recording**
- Sampling**
- Reconstruction of the Analog Signal*
- Aliasing (Foldover)**
- The Sampling Theorem**
- Ideal Sampling Frequency*

Antialiasing and Anti-imaging Filters
Phase Correction
Quantization
Quantization Noise
Low-level Quantization Noise and Dither
Converter Linearity

Dynamic Range of Digital Audio Systems

Decibels
Dynamic Range of a Digital System

Oversampling

Multiple-bit Oversampling Converters
1-bit Oversampling Converters

Digital Audio Media

Synthesis and Signal Processing

Conclusion

The merger of digital audio recording with computer music technology creates a supple and powerful artistic medium. This chapter introduces the history and technology of digital audio recording and playback. After studying this introduction, you should be familiar with the basic vocabulary and concepts of digital audio. In the interest of brevity we condense topics that are large specialities unto themselves; for more literature sources see D. Davis (1988, 1992).

Background: History of Digital Audio Recording

Sound recording has a rich history, beginning with Thomas Edison and Emile Berliner's experiments in the 1870s, and marked by V. Poulsen's Telegraphone magnetic wire recorder of 1898 (Read and Welch 1976). Early audio recording was a mechanical process (figure 1.1).

Although the invention of the triode vacuum tube in 1906 launched the era of electronics, electronically produced records did not become practical until 1924 (Keller 1981). Figure 1.2 depicts one of the horn-loaded loudspeakers typical in the 1920s.

Optical sound recording on film was first demonstrated in 1922 (Ristow 1993). Sound recording on tape coated with powdered magnetized material was developed in the 1930s in Germany (figure 1.3), but did not reach the rest of the world until after World War 2. The German Magnetophon tape

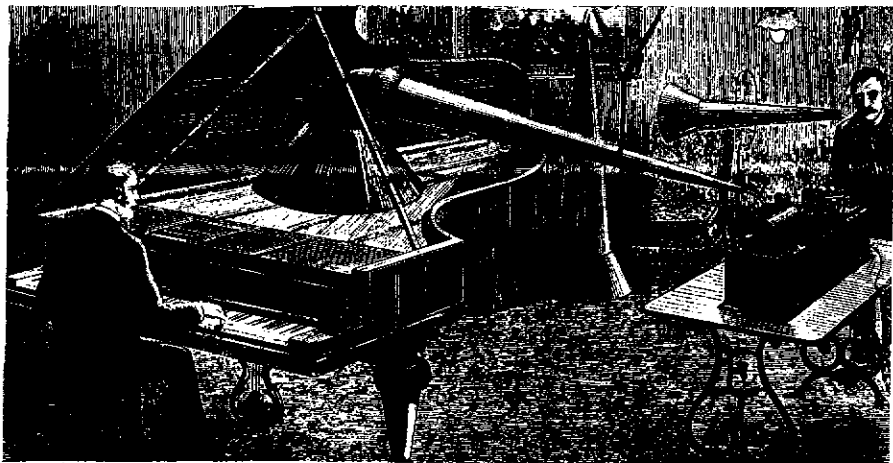
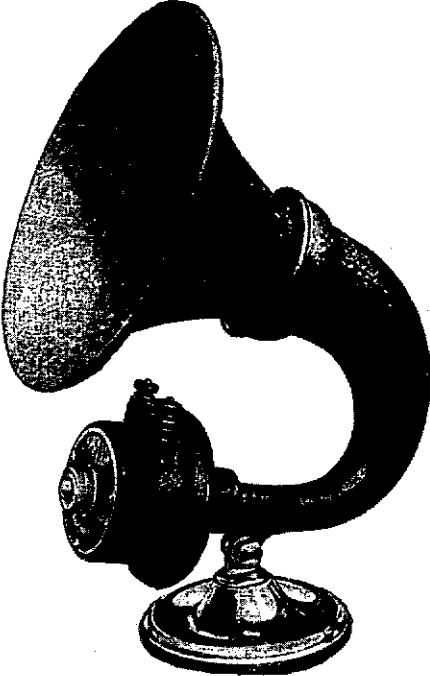


Figure 1.1 Mechanical recording session before 1900. Sound vibrations picked up by the large cone over the piano were transduced into vibrations of a cutting stylus piercing a rotating wax cylinder.

Haut-Parleurs
AMPLION
Brevets E.-A. GRAHAM



Amplion Libellule, Prix **135** francs
Auditions à l'Exposition Internationale de T. S. F., Arts Décoratifs, quai d'Orsay

Compagnie Française AMPLION
131, rue de Vaugirard, 131, PARIS (15^e)
R. C. Seine 216.437 B

Figure 1.2 Amplion loudspeaker, as advertised in 1925.

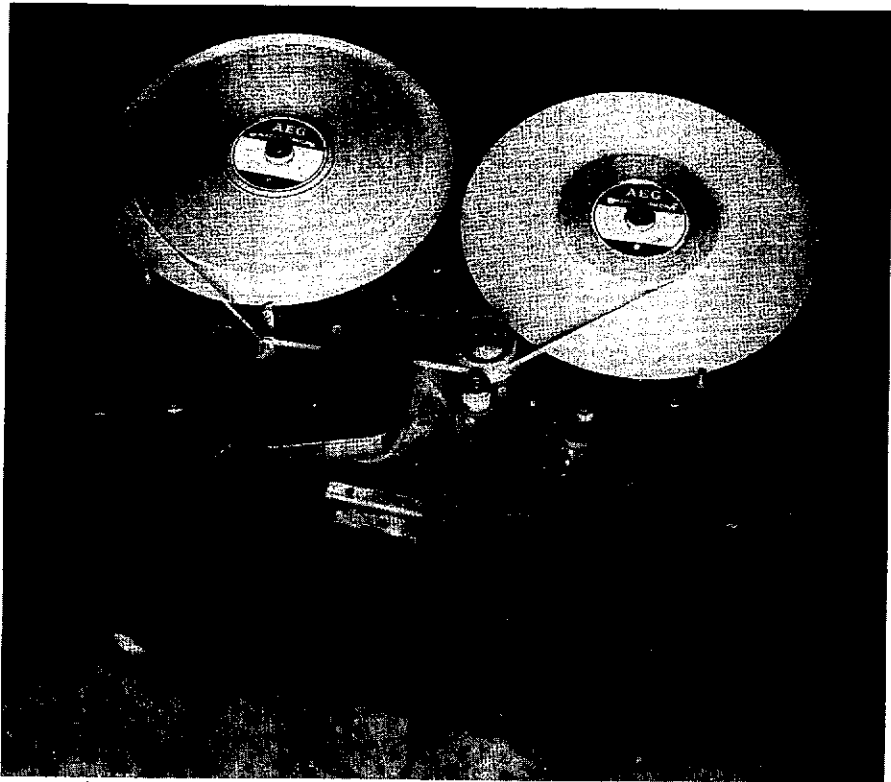


Figure 1.3 Prototype of a portable *Magnetophon* tape recorder from 1935, made by AEG. (Photograph courtesy of BASF Aktiengesellschaft.)

recorders were a great advance over previous wire and steel band recorders, which required soldering or welding to make a splice. The Magnetophons and their descendants were *analog* recorders. The term “analog” refers to the fact that the waveform encoded on tape is a close analogy to the original sound waveform picked up by a microphone. Analog recording continues to be refined, but faces fundamental physical limits. These limits are most apparent when making copies from one analog medium to another—additional noise is inescapable.

For more on the history of analog recording, with particular reference to multitrack machines, see chapter 9.

Experimental Digital Recording

The core concept in digital audio recording is *sampling*, that is, converting continuous analog signals (such as those coming from a microphone) into discrete *time-sampled* signals. The theoretical underpinning of sampling is

the *sampling theorem*, which specifies the relation between the sampling rate and the audio bandwidth (see the section on the sampling theorem later in this chapter). This theorem is also called the *Nyquist theorem* after the work of Harold Nyquist of Bell Telephone Laboratories (Nyquist 1928), but another form of this theorem was first stated in 1841 by the French mathematician Augustin Louis Cauchy (1789–1857). The British researcher A. Reeves developed the first patented *pulse-code-modulation* (PCM) system for transmission of messages in “amplitude-dichotomized, time-quantized” (digital) form (Reeves 1938; Licklider 1950; Black 1953). Even today, digital recording is sometimes called “PCM recording.” The development of *information theory* contributed to the understanding of digital audio transmission (Shannon 1948). Solving the difficult problems of converting between analog signals and digital signals took decades, and is still being improved. (We describe the conversion processes later.)

In the late 1950s, Max Mathews and his group at Bell Telephone Laboratories generated the first synthetic sounds from a digital computer. The samples were written by the computer to expensive and bulky reel-to-reel computer tape storage drives. The production of sound from the numbers was a separate process of playing back the tape through a custom-built 12-bit vacuum tube “digital-to-sound converter” developed by the Epsco Corporation (Roads 1980; see also chapter 3).

Hamming, Huffman, and Gilbert originated the theory of *digital error correction* in the 1950s and 1960s. Later, Sato, Blesser, Stockham, and Doi made contributions to error correction that resulted in the first practical systems for digital audio recording. The first dedicated one-channel digital audio recorder (based on a videotape mechanism), was demonstrated by the NHK, the Japan broadcasting company (Nakajima et al. 1983). Soon thereafter, Denon developed an improved version (figure 1.4), and the race began to bring digital audio recorders to market (Iwamura et al. 1973).

By 1977 the first commercial recording system came to market, the Sony PCM-1 processor, designed to encode 13-bit digital audio signals onto Sony Beta format videocassette recorders. Within a year this was displaced by 16-bit PCM encoders such as the Sony PCM-1600 (Nakajima et al. 1978). At this point product development split along two lines: professional and “consumer” units, although a real mass market for this type of digital recording never materialized. The professional Sony PCM-1610 and 1630 became the standards for compact disc (CD) mastering, while Sony PCM-F1-compatible systems (also called EIAJ systems, for Electronics Industry Association of Japan) became a de facto standard for low-cost digital audio recording on videocassette. These standards continued throughout the 1980s.

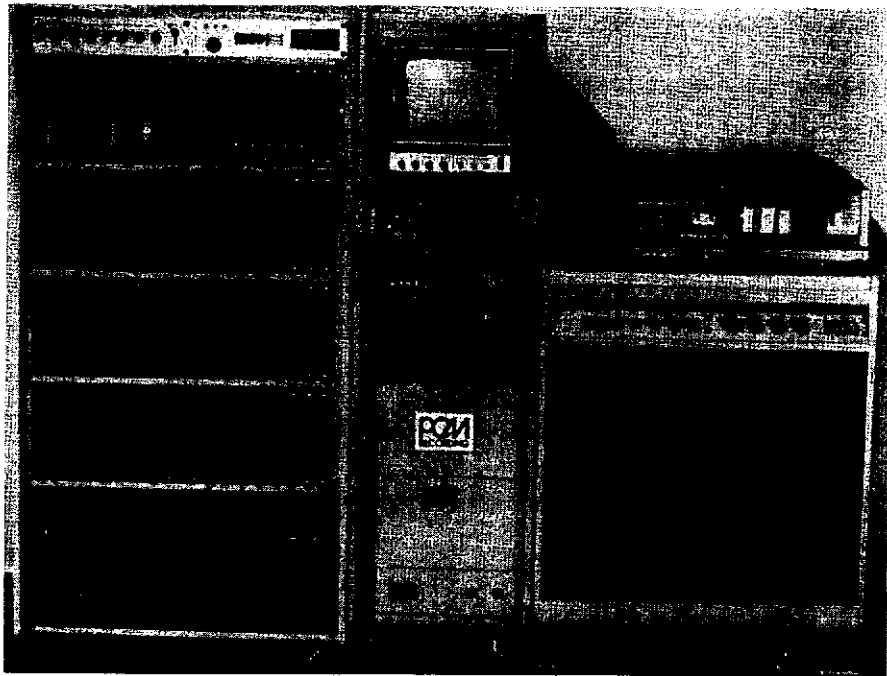


Figure 1.4 Nippon Columbia (Denon) digital audio recorder made in 1973 based on a 1-inch videotape recorder (on the right).

The Audio Engineering Society established two standard sampling frequencies in 1985: 44.1 and 48 KHz. They revised their specification in 1992 (Audio Engineering Society 1992a, 1992b). (A 32 KHz sampling frequency for broadcast purposes also exists.) Meanwhile, a few companies developed higher-resolution digital recorders capable of encoding more than sixteen bits at higher sampling rates. For example, a version of Mitsubishi's X-86 reel-to-reel digital tape recorder encoded 20 bits at a 96 KHz sampling frequency (Mitsubishi 1986). A variety of high-resolution recorders are now available.

Digital Sound for the Public

Digital sound first reached the general public in 1982 by means of the compact disc (CD) format, a 12-cm optical disc read by a laser (figure 1.5). The CD format was developed jointly by the Philips and Sony corporations after years of development. It was a tremendous commercial success, selling over 1.35 million players and tens of millions of discs within two years (Pohlman 1989). Since then a variety of products have been derived from



Figure 1.5 The Sony-Philips compact disc.

CD technology, including CD-ROM (Read Only Memory), CD-I (Interactive), and other formats that mix audio data, texts, and images.

By the early 1990s, manufacturers targeted the need for recordable digital media. Various stereo media appeared, including Digital Audio Tape (DAT), Digital Compact Cassettes (DCC), the Mini-Disc (MD), and recordable CDs (CD-R). (See the section on digital audio media below.)

Digital Sound for Musicians

Although CD players had inexpensive 16-bit DACs, good-quality converters attached to computers were not common before 1988. Prior to this time,

a few institutional computer music centers developed custom-made ADCs and DACs, but owners of the new personal computer systems had to wait. They could buy digital synthesizers and control them from their computer using the MIDI protocol (see chapter 21), but they could not directly synthesize or record sound with the computer.

Only in the late 1980s did low-cost, good-quality converters become available for personal computers. This development heralded a new era for computer music. In a short period, sound synthesis, recording, and processing by computer became widespread. Dozens of different *audio workstations* reached the musical marketplace. These systems let musicians record music onto the hard disk connected to a personal computer. This music could be precisely edited on the screen of the computer, with playback from the hard disk.

Digital Multitrack Recording

In contrast to stereo recorders that record both left and right channels at the same time, *multitrack* recorders have several discrete channels or *tracks* that can be recorded at different times. Each track can record a separate instrument, for example, allowing flexibility when the tracks are later mixed. Another advantage of multitrack machines is that they let musicians build recordings in several layers; each new layer is an accompaniment to previously recorded layers.

The British Broadcasting Company (BBC) developed an experimental ten-channel digital tape recorder in 1976. Two years later, the 3M company, working with the BBC, introduced the first commercial 32-track digital recorder (figure 1.6) as well as a rudimentary digital tape editor (Duffy 1982). The first computer disk-based random-access sound editor and mixer was developed by the Soundstream company in Salt Lake City, Utah (see figure 16.38). Their system allowed mixing of up to eight tracks or *sound files* stored on computer disk at a time (Ingebretsen and Stockham 1984).

By the mid-1980s, both 3M and Soundstream had withdrawn from the digital multitrack tape recorder market, which was then dominated by the Sony and Mitsubishi conglomerates, later joined by the Studer company. For a number of years, digital multitrack recording was a very expensive enterprise (figure 1.7). The situation entered a new phase in the early 1990s with the introduction of low-cost multitrack tape recorders by Alesis and Tascam, and inexpensive multitrack disk recorders by a variety of concerns. (Chapter 9 recounts the history of analog multitrack recording.)

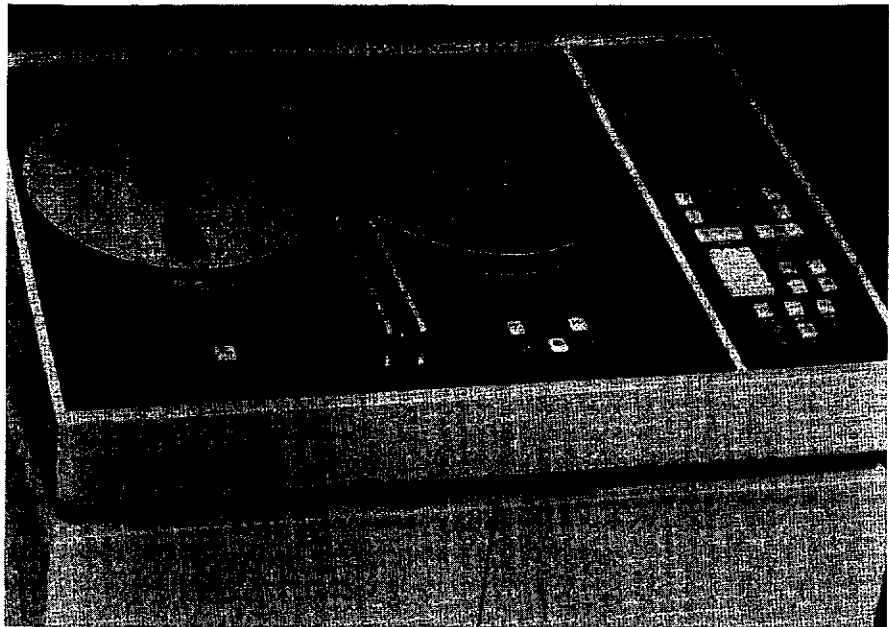


Figure 1.6 3M 32-track digital tape recorder, introduced in 1978.

Basics of Sound Signals

This section introduces the basic concepts and terminology for describing sound signals, including frequency, amplitude, and phase.

Frequency and Amplitude

Sound reaches listeners' ears after being transmitted through air from a source. Listeners hear sound because the air pressure is changing slightly in their ears. If the pressure varies according to a repeating pattern we say the sound has a *periodic waveform*. If there is no discernible pattern it is called *noise*. In between these two extremes is a vast domain of quasi-periodic and quasi-noisy sounds.

One repetition of a periodic waveform is called a *cycle*, and the *fundamental frequency* of the waveform is the number of cycles that occur per second. As the length of the cycle—called the *wavelength* or *period*—increases, the frequency in cycles per second decreases, and vice versa. In the rest of this book we substitute Hz for “cycles per second” in accordance with standard acoustical terminology. (Hz is an abbreviation for Hertz, named after the German acoustician Heinrich Hertz.)

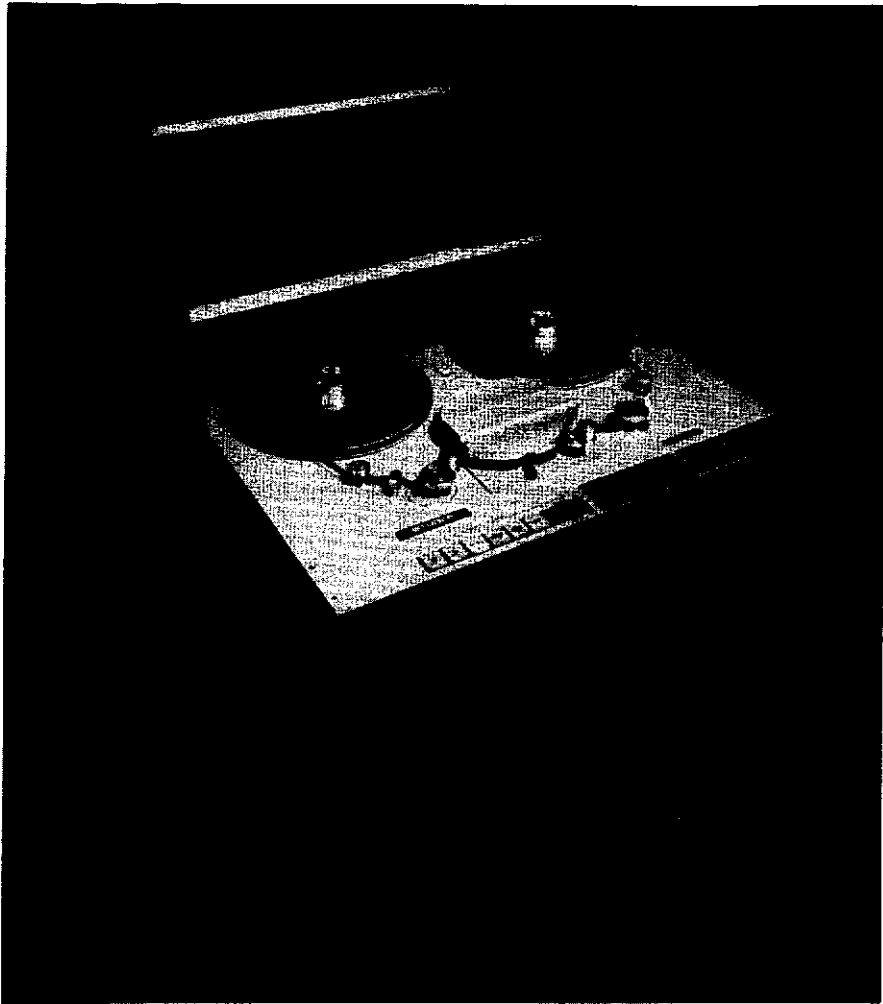


Figure 1.7 Studer D820-48 DASH digital multitrack recorder introduced in 1991 with a retail price of about \$270,000.

Time-domain Representation

A simple method of depicting sound waveforms is to draw them in the form of a graph of air pressure versus time (figure 1.8). This is called a *time-domain* representation. When the curved line is near the bottom of the graph, then the air pressure is lower, and when the curve is near the top of the graph, the air pressure has increased. The *amplitude* of the waveform is the amount of air pressure change; we can measure amplitude as the vertical distance from the zero pressure point to the highest (or lowest) points of a given waveform segment.

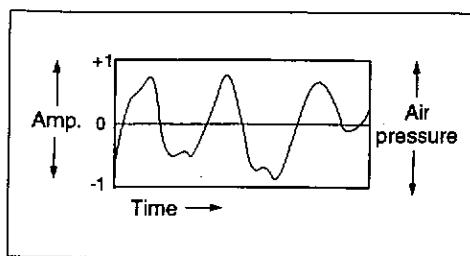


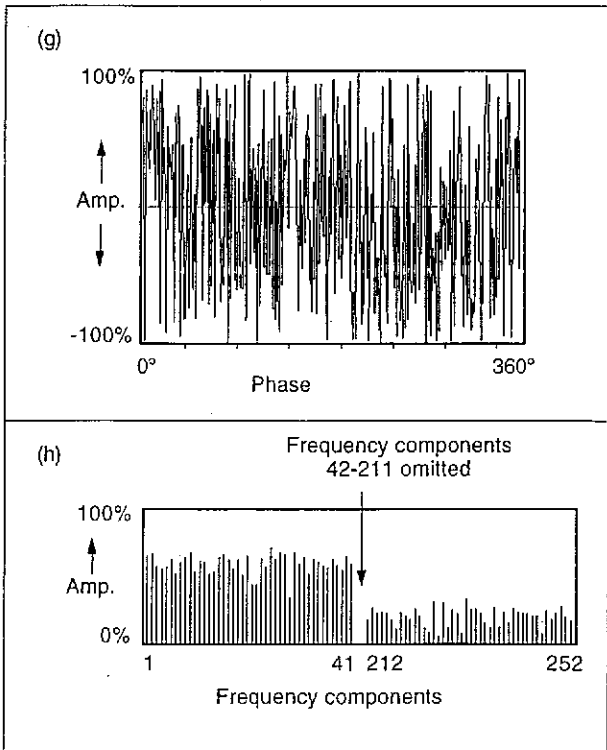
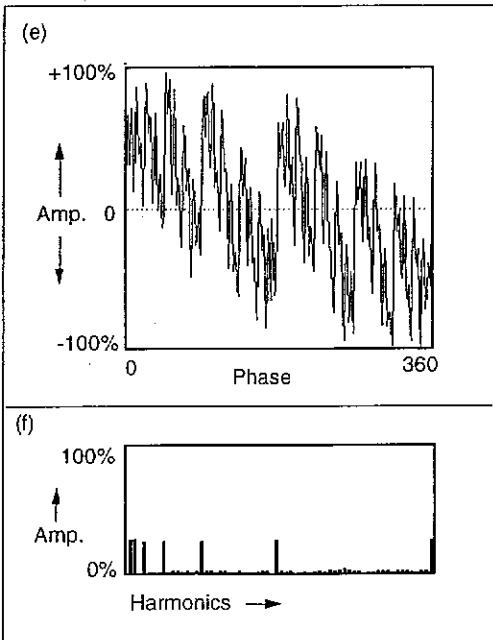
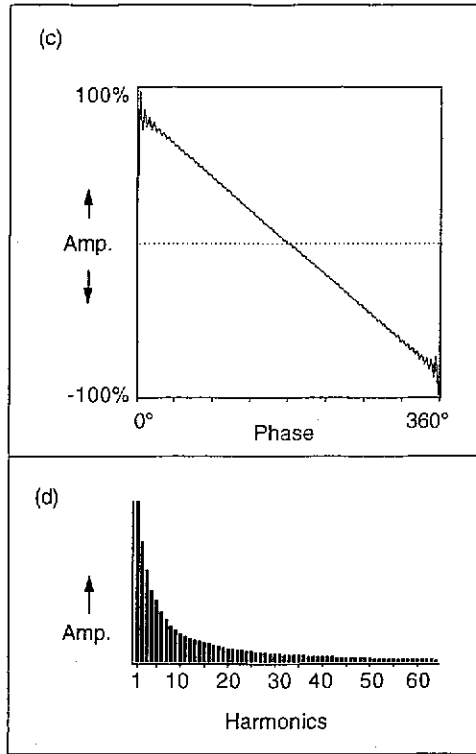
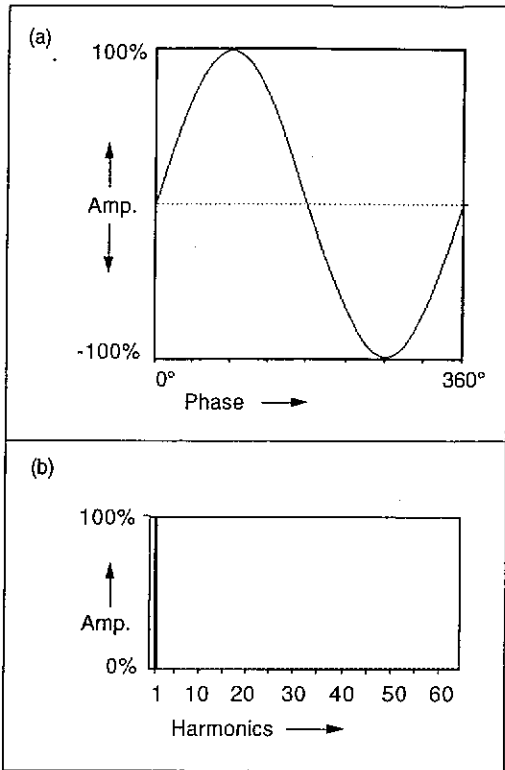
Figure 1.8 Time-domain representation of a signal. The vertical dimension shows the air pressure. When the curved line is near the top of the graph, the air pressure is greater. Below the solid horizontal line, the air pressure is reduced. Atmospheric pressure variations heard as sound can occur quickly; for musical sounds, this entire graph might last no more than one-thousandth of a second (1 ms).

An acoustic instrument creates sound by emitting vibrations that change the air pressure around the instrument. A loudspeaker creates sound by moving back and forth according to voltage changes in an electronic signal. When the loudspeaker moves “in” from its position at rest, then the air pressure decreases. As the loudspeaker moves “out,” the air pressure near the loudspeaker is raised. To create an audible sound these in/out vibrations must occur at a frequency in the range of about 20 to 20,000 Hz.

Frequency-domain Representation

Besides the fundamental frequency, there can be many frequencies present in a waveform. A *frequency-domain* or *spectrum* representation shows the frequency content of a sound. The individual frequency components of the spectrum can be referred to as *harmonics* or *partials*. Harmonic frequencies are simple integer multiples of the fundamental frequency. Assuming a

Figure 1.9 Time-domain and frequency-representations of four signals. (a) Time-domain view of one cycle of a sine wave. (b) Spectrum of the one frequency component in a sine wave. (c) Time-domain view of one cycle of a sawtooth waveform. (d) Spectrum showing the exponentially decreasing frequency content of a sawtooth wave. (e) Time-domain view of one cycle of a complex waveform. Although the waveform looks complex, when it is repeated over and over its sound is actually simple—like a thin reed organ sound. (f) The spectrum of waveform (e) shows that it is dominated by a few frequencies. (g) A random noise waveform. (h) If the waveform is constantly changing (each cycle is different from the last cycle) then we hear noise. The frequency content of noise is very complex. In this case the analysis extracted 252 frequencies. This snapshot does not reveal how their amplitudes are constantly changing over time.



fundamental or *first harmonic* of 440 Hz, its second harmonic is 880 Hz, its third harmonic is 1760 Hz, and so on. More generally, any frequency component can be called a partial, whether or not it is an integer multiple of a fundamental. Indeed, many sounds have no particular fundamental frequency.

The frequency content of a waveform can be displayed in many ways. A standard way is to plot each partial as a line along an x -axis. The height of each line indicates the strength (or amplitude) of each frequency component. The purest signal is a *sine* waveform, so named because it can be calculated using trigonometric formulae for the sine of an angle. (Appendix A explains this derivation.) A pure sine wave represents just one frequency component, or one line in a spectrum. Figure 1.9 depicts the time-domain and frequency-domain representations of several waveforms. Notice that the spectrum plots are labeled “Harmonics” on their horizontal axis, since the analysis algorithm assumes that its input is exactly one period of the fundamental of a periodic waveform. In the case of the noise signal in figure 1.9g, this assumption is not valid, so we relabel the partials as “frequency components.”

Phase

The starting point of a periodic waveform on the y or amplitude axis is its *initial phase*. For example, a typical sine wave starts at the amplitude point 0 and completes its cycle at 0. If we displace the starting point by 2π on the horizontal axis (or 90 degrees) then the sinusoidal wave starts and ends at 1 on the amplitude axis. By convention this is called a cosine wave. In effect, a cosine is equivalent to a sine wave that is *phase shifted* by 90 degrees (figure 1.10).

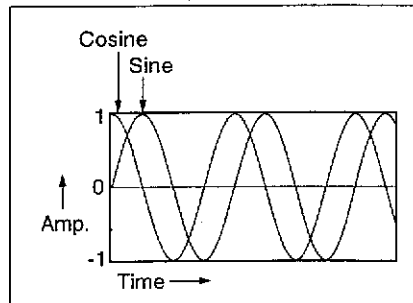


Figure 1.10 A sine waveform is equivalent to a cosine waveform that has been delayed or phase shifted slightly.

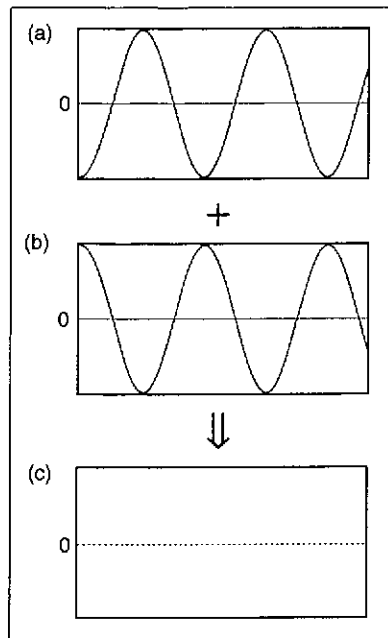


Figure 1.11 The effects of phase inversion. (b) is a phase-inverted copy of (a). If the two waveforms are added together, they sum to zero (c).

When two signals start at the same point they are said to be *in phase* or *phase aligned*. This contrasts to a signal that is slightly delayed with respect to another signal, in which the two signals are *out of phase*. When a signal *A* is the exact opposite phase of another signal *B* (i.e., it is 180 degrees out of phase, so that for every positive value in signal *A* there is a corresponding negative value for signal *B*), we say that *B* has *reversed polarity* with respect to *A*. We could also say that *B* is a *phase-inverted copy* of *A*. Figure 1.11 portrays the effect when two signals in inverse phase relationship sum.

Importance of Phase

It is sometimes said that phase is insignificant to the human ear, because two signals that are exactly the same except for their initial phase are difficult to distinguish. Actually, research indicates that 180-degree differences in absolute phase or *polarity* can be distinguished by some people under laboratory conditions (Greiner and Melton 1991). But even apart from this special case, phase is an important concept for several reasons. Every filter uses phase shifts to alter signals. A filter phase shifts a signal (by delaying its input for a short time) and then combines the phase-shifted version with the original signal to create *frequency-dependent phase cancellation* effects that

alter the spectrum of the original. By “frequency-dependent” we mean that not all frequency components are affected equally. When the phase shifting is time-varying, the affected frequency bands also vary, creating the sweeping sound effect called *phasing* or *flanging* (see chapter 10).

Phase is also important in systems that resynthesize sound on the basis of an analysis of an existing sound. In particular, these systems need to know the starting phase of each frequency component in order to put together the different components in the right order (see chapter 13 and Appendix A.) Phase data are particularly critical in reproducing short, rapidly changing *transient* sounds, such as the onset of an instrumental tone.

Finally, much attention has been invested in recent years to audio components that phase shift their input signals as little as possible, because frequency-dependent phase shifts distort musical signals audibly and interfere with loudspeaker *imaging*. (Imaging is the ability of a set of loudspeakers to create a stable “audio picture” where each audio source is localized to a specific place within the picture.) Unwanted phase shifting is called *phase distortion*. To make a visual analogy, a phase-distorted signal is “out of focus.”

Now that we have introduced the basic properties of audio signals, we take a comparative look at two representations for them: analog and digital.

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

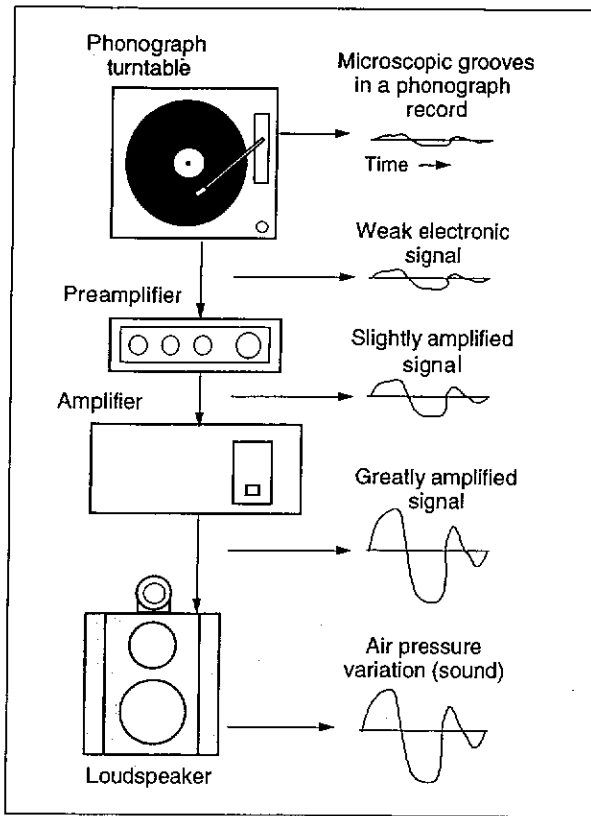


Figure 1.12 The analog audio chain, starting from an analog waveform transduced from the grooves of a phonograph record to a voltage sent to a preamplifier, amplifier, loudspeaker, and projected into the air.

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

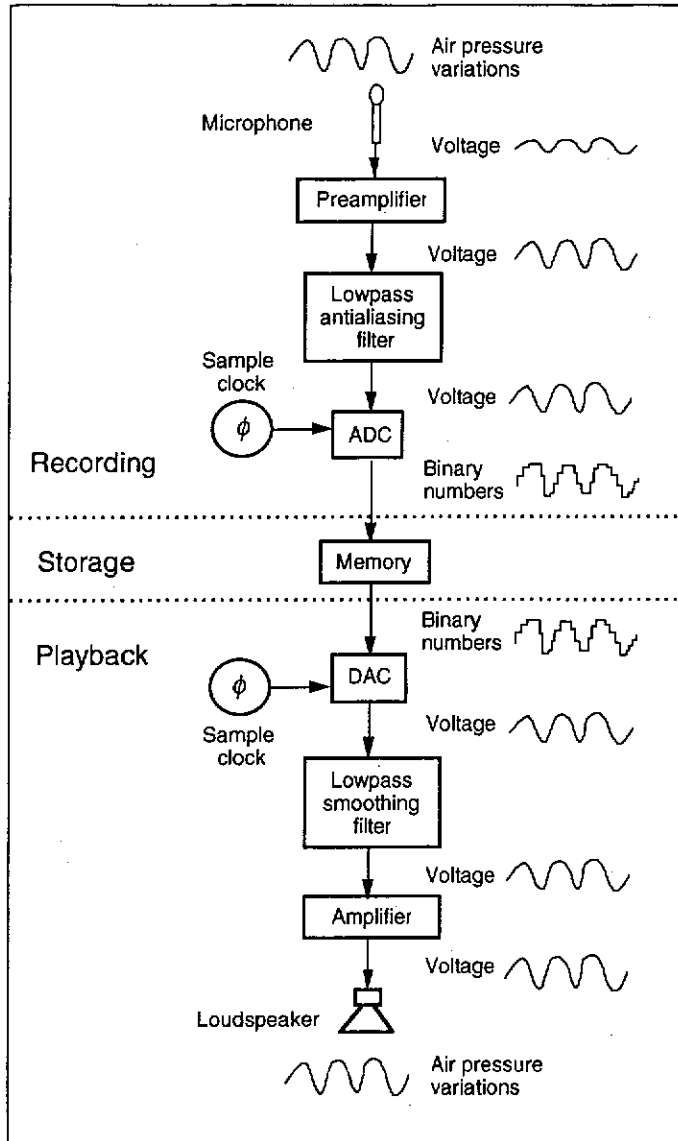


Figure 1.13 Overview of digital recording and playback.

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.

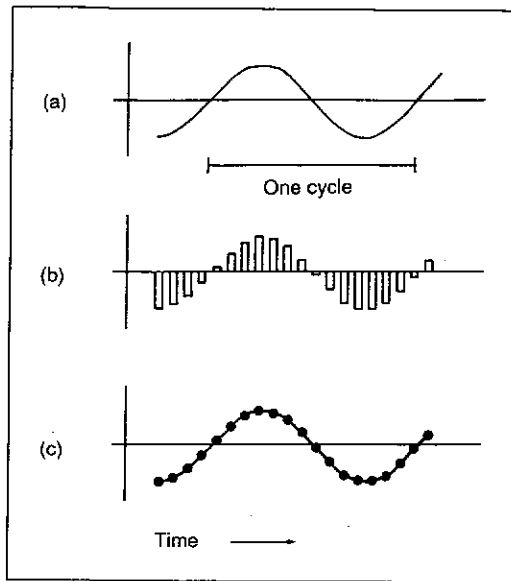


Figure 1.14 Analog and digital representations of a signal. (a) Analog sine waveform. The horizontal bar below the wave indicates one period or cycle. (b) Sampled version of the sine waveform in (a), as it might appear at the output of an ADC. Each vertical bar represents one sample. Each sample is stored 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 ($44,100 \times 2 \text{ channels} \times 4 \text{ seconds}$). 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.

Reconstruction of the Analog Signal

Sampling frequencies around 50 KHz are common in digital audio systems, although both lower and higher frequencies can also be found. In any case,

50,000 numbers per second is a rapid stream of numbers; it means there are 6,000,000 samples for one minute of stereo sound.

The digital signal in figure 1.13b does not show the value between the bars. The duration of a bar is extremely narrow, perhaps lasting only 0.00002 second (two hundred-thousandths of a second). This means that if the original signal changes “between” bars, the change is not reflected in the height of a bar, at least until the next sample is taken. In technical terms, we say that the signal in figure 1.13b is defined at *discrete* times, each such time represented by one sample (vertical bar).

Part of the magic of digitized sound is that if the signal is bandlimited, the DAC and associated hardware can exactly reconstruct the original signal from these samples! This means that, given certain conditions, the missing part of the signal “between the samples” can be restored. This happens when the numbers are passed through the DAC and smoothing filter. The smoothing filter “connects the dots” between the discrete samples (see the dotted line in figure 1.13c). Thus, a signal sent to the loudspeaker looks and sounds like the original signal.

Aliasing (Foldover)

The process of sampling is not quite as straightforward as it might seem. Just as an audio amplifier or a loudspeaker can introduce distortion, sampling can play tricks with sound. Figure 1.15 gives an example. Using the input waveform shown in figure 1.15a, suppose that a sample of this waveform is taken at each point in time shown by the vertical bars in figure 1.15b (each vertical bar creates one sample). As before, the resulting samples of figure 1.15c are stored as numbers in digital memory. But when we attempt to reconstruct the original waveform, as shown in figure 1.15d, the result is something completely different.

In order to understand better the problems that can occur with sampling, we look at what happens when we change the *wavelength* (the length of one cycle) of the original signal without changing the length of time between samples. Figure 1.16a shows a signal with a cycle eight samples long, figure 1.16d shows a cycle two samples long, and figure 1.16g shows a waveform with eleven cycles per ten samples. This means that one cycle takes longer than the interval between samples. This relationship could also be expressed as 11/10 cycles per sample.

Again, as each of the sets of samples is passed through the DAC and associated hardware, a signal is reconstructed (figures 1.16c, f, and i) and sent to the loudspeaker. The signal shown by the dotted line in figure 1.16c

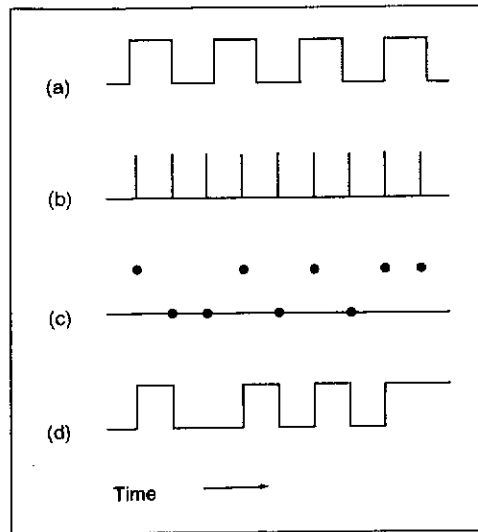


Figure 1.15 Problems in sampling. (a) Waveform to be recorded. (b) The sampling pulses; whenever a sampling pulse occurs, one sample is taken. (c) The waveform as sampled and stored in memory. (d) When the waveform from (c) is sent to the DAC, the output might appear as shown here (after Mathews 1969).

is reconstructed more or less accurately. The results of the sampling in figure 1.16f are potentially a little less satisfactory; one possible reconstruction is shown there. But in figure 1.16i, the resynthesized waveform is completely different from the original in one important respect. Namely, the wavelength (length of the cycle) of the resynthesized waveform is different from that of the original. In the real world, this means that the reconstructed signal sounds at a pitch different from that of the original signal. This kind of distortion is called *aliasing* or *foldover*.

The frequencies at which this aliasing occurs can be predicted. Suppose, just to keep the numbers simple, that we take 1000 samples per second. Then the signal in figure 1.16a has a frequency of 125 cycles per second (since there are eight samples per cycle, and $1000/8 = 125$). In figure 1.16d, the signal has a frequency of 500 cycles per second (because $1000/2 = 500$).

The frequency of the input signal in figure 1.16g is 1100 cycles per second. But the frequency of the output signal is different. In figure 1.16i you can count ten samples per cycle of the output waveform. In actuality, the output waveform occurs at a frequency of $1000/10 = 100$ cycles per second. Thus the frequency of the original signal in figure 1.16g has been changed by the *sample rate conversion* process. This represents an unacceptable change to a musical signal, which must be avoided if possible.

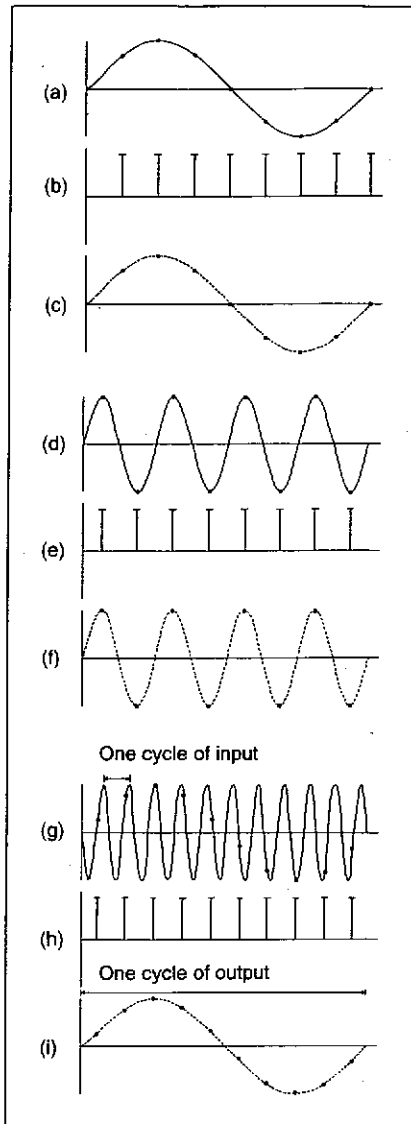


Figure 1.16 Foldover effects. At the bottom of each set of three graphs, the thick black dots represent samples, and the dotted line shows the signal as reconstructed by the DAC. Every cycle of the sine waveform (a) is sampled eight times in (b). Using the same sampling frequency, each cycle of (d) is sampled only twice in (e). If the sampling pulses in (e) were moved to the right, the output waveform in (f) might be phase-shifted, although the frequency of the output would still be the same. In (h), there are ten samples for the eleven cycles in (g). When the DAC tries to reconstruct a signal, as shown by the dashed lines in (i), a sine waveform results, but the frequency has been completely changed due to the foldover effect. Notice the horizontal double arrow above (g), indicating one cycle of the input waveform, and the arrow above (i), indicating one cycle of the output waveform.

The Sampling Theorem

We can generalize from figure 1.16 to say that as long as there are at least two samples per period of the original waveform, we can assume that the resynthesized waveform will have the same frequency. But when there are fewer than two samples per period, the frequency (and perhaps the timbre) of the original signal is lost. In this case, the new frequency can be found by the following formula. If the original frequency is higher than half the sampling frequency, then:

$$\text{new frequency} = \text{sampling frequency} - \text{original frequency}$$

This formula is not mathematically complete, but it is sufficient for our discussion here. It means the following. Suppose we have chosen a fixed sampling frequency. We start with a signal at a low frequency, sample it, and resynthesize the signal after sampling. As we raise the pitch of the input signal (but still keep the sampling frequency constant), the pitch of the resynthesized signal is the same as the pitch of the input signal until we reach a pitch that corresponds to one-half of the sampling frequency. As we raise the pitch of the input signal even higher, the pitch of the output signal goes down to the lowest frequencies! When the input signal reaches the sampling frequency, the entire process repeats itself.

To give a concrete example, suppose we introduce an analog signal at 26 KHz into an analog-to-digital converter operating at 50 KHz. The converter reads it as a tone at 24 KHz, since $50 - 26 = 24$ KHz.

The *sampling theorem* describes the relationship between the sampling rate and the bandwidth of the signal being transmitted. It was expressed by Harold Nyquist (1928) as follows:

For any given deformation of the received signal, the transmitted frequency range must be increased in direct proportion to the signalling speed. . . . The conclusion is that the frequency band is directly proportional to the speed.

The essential point of the sampling theorem can be stated precisely as follows:

In order to be able to reconstruct a signal, the sampling frequency must be at least twice the frequency of the signal being sampled.

In honor of his contributions to sampling theory, the highest frequency that can be produced in a digital audio system (i.e., half the sampling rate) is called the *Nyquist frequency*. In musical applications, the Nyquist frequency is usually in the upper range of human hearing, above 20 KHz. Then the

sampling frequency can be specified as being at least twice as much, or above 40 KHz.

In some systems the sampling frequency is set somewhat greater than twice this highest frequency, because the converters and associated hardware cannot perfectly reconstruct a signal near half the sampling frequency (an idealized reconstruction of such a case is shown in figure 1.16f).

Ideal Sampling Frequency

The question of what sampling frequency is ideal for high-quality music recording and reproduction is an ongoing debate. Part of the reason is that mathematical theory and engineering practice often conflict: converter clocks are not stable, converter voltages are not linear, filters introduce phase distortion, and so on. (See the sections on phase correction and over-sampling later.)

Another reason is that many people hear information (referred to as "air") in the region around the 20 KHz "limit" on human hearing (Neve 1992). Indeed, Rudolf Koenig, whose precise measurements set international standards for acoustics, observed at age 41 that his own hearing extended to 23 KHz (Koenig 1899). It seems strange that a new digital compact disc should have less bandwidth than a phonograph record made in the 1960s, or a new digital audio recorder should have less bandwidth than a twenty-year old analog tape recorder. Many analog systems can reproduce frequencies beyond 25 KHz. Scientific experiments confirm the effects of sounds above 22 KHz from both physiological and subjective viewpoints (Oohashi et al. 1991; Oohashi et al. 1993).

In sound synthesis applications, the lack of "frequency headroom" in standard sampling rates of 44.1 and 48 KHz causes serious problems. It requires that synthesis algorithms generate nothing other than sine waves above 11 KHz (44.1 KHz sampling rate) or 12 KHz (48 KHz sampling rate), or foldover will occur. This is because any high-frequency component with partials beyond the fundamental has a frequency that exceeds the Nyquist rate. The third harmonic of a tone at 12.5 KHz, for example, is 37.5 KHz, which in a system running at 44.1 KHz sampling rate will reflect down to an audible 6600 Hz tone. In sampling and pitch-shifting applications, the lack of frequency headroom requires that samples be lowpass filtered before they are pitch-shifted upward. The trouble these limits impose is inconvenient.

It is clear that high-sampling rate recordings are preferable from an artistic standpoint, although they pose practical problems of additional storage

and the need for high-quality audio playback systems to make the effort worthwhile.

Antialiasing and Anti-imaging Filters

In order to make sure that a digital sound system works properly, two important filters are included. One filter is placed before the ADC, to make sure that nothing (or as little as possible) in the input signal occurs at a frequency higher than half of the sampling frequency. As long as this filter does the proper work, aliasing should not occur during the recording process. Logically enough, such a filter is called an *antialiasing filter*.

The other filter is placed after the DAC. Its main function is to change the samples stored digitally into a smooth, continuous representation of the signal. In effect, this lowpass *anti-imaging* or *smoothing filter* creates the dotted line in figure 1.14c by connecting the solid black dots in the figure.

Phase Correction

The issue of *phase correction* came rushing to the fore following the introduction of the first generation of digital audio recorders and players. Many complained about the harsh sound of digital recordings, a problem that could be traced to the the *brickwall* antialiasing filters before the ADCs (Woszczyk and Toole 1983; Preis and Bloom 1983). They are called brick-wall filters because of their steep frequency rejection curve (over 90 dB/octave at the Nyquist frequency, typically). These steep filters can cause significant time-delays (phase distortion) in midrange and high audio frequencies (figure 1.17). A smaller frequency-dependent delay is contributed by the smoothing filter at the output of a DAC.

No analog filter can be both extremely steep and *phase linear* around the cutoff point. (Phase linear means that there is little or no frequency-dependent delay introduced by the filter.) Hence, the effect of a steep filter “spills over” into the audio range. For compact disc recordings at a 44.1 KHz sampling rate, the Nyquist frequency is 22.05 KHz, and a steep antialiasing filter can introduce phase distortion that extends well below 10 KHz (Meyer 1984). This type of phase distortion lends an unnaturally harsh sound to high frequencies.

There are various ways to tackle this problem. The simplest is to trade off the antialiasing properties of the filter in favor of less phase distortion. A less steep antialiasing filter (40–60 dB/octave, for example) introduces less phase distortion, but at the risk of foldover for very high frequency sounds. Another solution is to apply a *time correction filter* before the ADC to

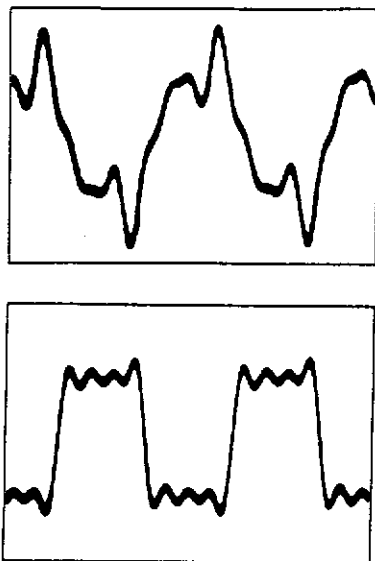


Figure 1.17 Phase distortion caused by an antialiasing filter. (a) 2.5 KHz square wave distorted by a brickwall antialiasing filter. (b) Phase-corrected square wave.

skew the phase relationships in the incoming signal so as to preserve the original phase relationships in the recording (Blessner 1984; Greenspun 1984; Meyer 1984). At present, however, the high-technology solution to phase correct conversion is to use *oversampling* techniques at both the input and output stages of the system. We discuss oversampling later.

Quantization

Sampling at discrete time intervals, discussed in the previous sections, constitutes one of the major differences between digital and analog signals. Another difference is *quantization*, or discrete amplitude resolution. The values of the sampled signal cannot take on any conceivable value. This is because digital numbers can only be represented within a certain range and with a certain accuracy, which varies with the hardware being used. The implications of this are an important factor in digital audio quality.

Quantization Noise

Samples are usually represented as integers. If the input signal has a voltage corresponding to a value between 53 and 54, for example, then the converter might round it off and assign a value of 53. In general, for each sample taken, the value of the sample usually differs slightly from the value

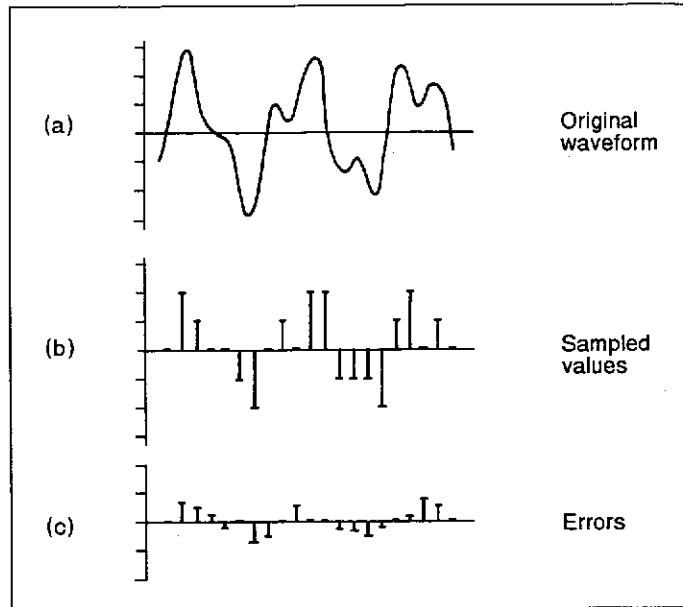


Figure 1.18 Effects of quantization. (a) Analog waveform. (b) Sampled version of the waveform in (a). Each sample can be assigned only certain values, which are indicated by the short horizontal dashes at the left. The difference between each sample and the original signal is shown in (c), where the height of each bar represents the quantization error.

of the original signal. This problem in digital signals is known as *quantization error* or *quantization noise* (Blessner 1978; Maher 1992; Lipshitz et al. 1992; Pohlmann 1989a).

Figure 1.18 shows the kinds of quantization errors that can occur. When the input signal is something complicated like a symphony, and we listen to just the errors, shown at the bottom of figure 1.18, it sounds like noise. If the errors are large, then one might notice something similar to analog tape hiss at the output of a system.

The quantization noise is dependent on two factors: the signal itself, and the accuracy with which the signal is represented in digital form. We can explain the sensitivity to the signal by noting that on an analog tape recorder, the tape imposes a soft halo of noise that continues even through periods of silence on the tape. But in a digital system there can be no quantization noise when nothing (or silence) is recorded. In other words, if the input signal is silence, then the signal is represented by a series of samples, each of which is exactly zero. The small differences shown in figure 1.18c disappear for such a signal, which means that the quantization noise disappears. If, on the other hand, the input signal is a pure sinusoid, then

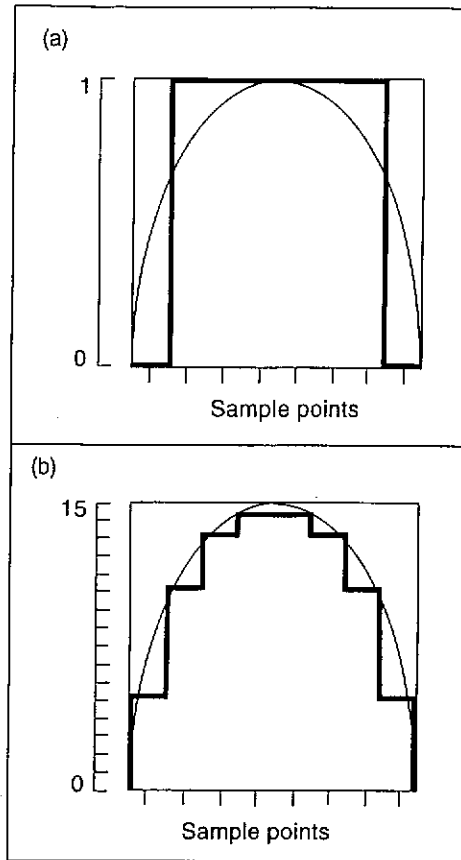


Figure 1.19 Comparing the accuracy of 4-bit quantization with that of 1-bit quantization. The thin rounded curve is the input waveform. (a) 1-bit quantization provides two levels of amplitude resolution, while (b) 4-bit quantization provides sixteen different levels of amplitude resolution.

the quantization error is not a random function but a deterministic truncation effect (Maher 1992). This gritty sound, called *granulation noise*, can be heard when very low level sinusoids decay to silence. When the input signal is complicated, the granulation becomes randomized into white noise.

The second factor in quantization noise is the accuracy of the digital representation. In a PCM system that represents each sample value by an integer (a *linear PCM* system), quantization noise is directly tied to the number of bits that are used to represent a sample. This specification is the *sample width* or *quantization level* of a system. Figure 1.19 illustrates the effects of different quantization levels, comparing the resolution of 1-bit versus 4-bit quantization. In a linear PCM system generally, the more bits used to represent a sample, the less the quantization noise. Figure 1.20

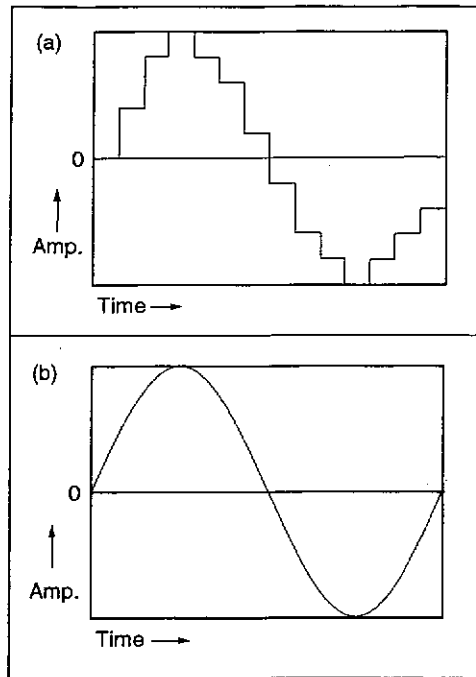


Figure 1.20 Effect of quantization on sine wave smoothness. (a) “Sine” wave with ten levels of quantization, corresponding to a moderately loud tone emitted by a 4-bit system. (b) Smoother sinusoid emitted by an 8-bit system.

shows the dramatic improvement in sine wave accuracy achieved by adding more bits of resolution.

The quantization measure is confused by *oversampling* systems, which use a high-speed “1-bit” converter. The quantization of a system that uses a “1-bit” converter is actually much greater than 1 bit. See the section on oversampling later.

Low-level Quantization Noise and Dither

Although a digital system exhibits no noise when there is no input signal, at very low (but nonzero) signal levels, quantization noise takes a pernicious form. A very low level signal triggers variations only in the lowest bit. These 1-bit variations look like a square wave, which is rich in odd harmonics. Consider the decay of a piano tone, which smoothly attenuates with high partials rolling off—right until the lowest level when it changes character and becomes a harsh-sounding square wave. The harmonics of the square wave may even extend beyond the Nyquist frequency, causing aliasing and introducing new frequency components that were not in the

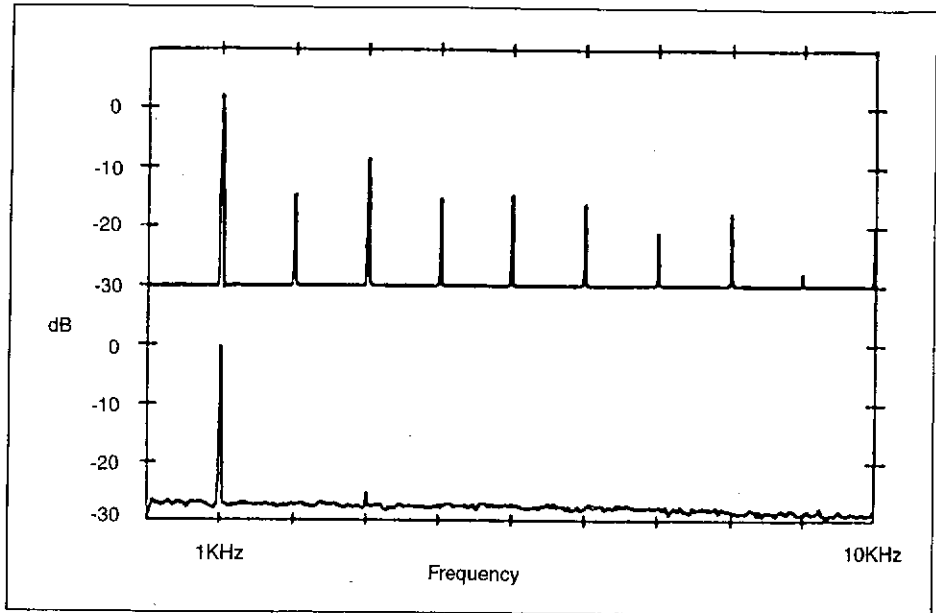


Figure 1.21 Dither reduces harmonic distortion in a digital system. The top part of the figure shows the spectrum of 1 KHz sine wave with an amplitude of 1/2 bit. Note the harmonics produced by the action of the ADC. The lower part shows the spectrum of the same signal after dithering of about 1 bit in amplitude is applied before conversion. Only a small amount of third harmonic noise remains, along with wideband noise. The ear can resolve the sine wave below the noise floor.

original signal. These artifacts may be possible to ignore if the signal is kept at a low monitoring level, but if the signal is heard at a high level or if it is digitally remixed to a higher level, it becomes more obvious. Hence it is important that the signal be quantized as accurately as possible at the input stage.

To confront low-level quantization problems, some digital recording systems take what seems at first to be a strange action. They introduce a small amount of analog noise—called *dither*—to the signal prior to analog-to-digital conversion (Vanderkooy and Lipshitz 1984; Lipshitz et al. 1992). This causes the ADC to make random variations around the low-level signal, which smooths out the pernicious effects of square wave harmonics (figure 1.21). With dither, the quantization error, which is usually signal-dependent, is turned into a wide-band noise that is uncorrelated with the signal. For decrescendos like the piano tone mentioned previously, the effect is that of a “soft landing” as the tone fades smoothly into a bed of low-level random noise. The amount of added noise is usually on the order of 3 dB, but the ear can reconstruct musical tones whose amplitudes fall

below that of the dither signal. See Blesser (1978, 1983), Rabiner and Gold (1975), Pohlmann (1989a), and Maher (1992) for more details on quantization noise and methods for minimizing it. Lipshitz, Wannamaker, and Vanderkooy (1992) present a mathematical analysis of quantization and dither. See Hauser (1991) for a discussion of dither in oversampling converters.

Dither may not be necessary with an accurate 20-bit converter, since the low bit represents an extremely soft signal in excess of 108 dB below the loudest signal. But when converting signals from a 20-bit to a 16-bit format, for example, dithering is necessary to preserve signal fidelity.

Converter Linearity

Converters can cause a variety of distortions (Blesser 1978; McGill 1985; Talambiras 1985). One that is pertinent here is that an n -bit converter is not necessarily accurate to the full dynamic range implied by its n -bit input or output. While the *resolution* of an n -bit converter is one part in 2^n , a converter's *linearity* is the degree to which the analog and digital input and output signals match in terms of their magnitudes. That is, some converters use 2^n steps, but these steps are not linear, which causes distortion. Hence it is possible to see an "18-bit converter," for example, that is "16-bit linear." Such a converter may be better than a plain 16-bit converter, which may not be 16-bit linear. (See Polhmann 1989a for a discussion of these problems.)

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

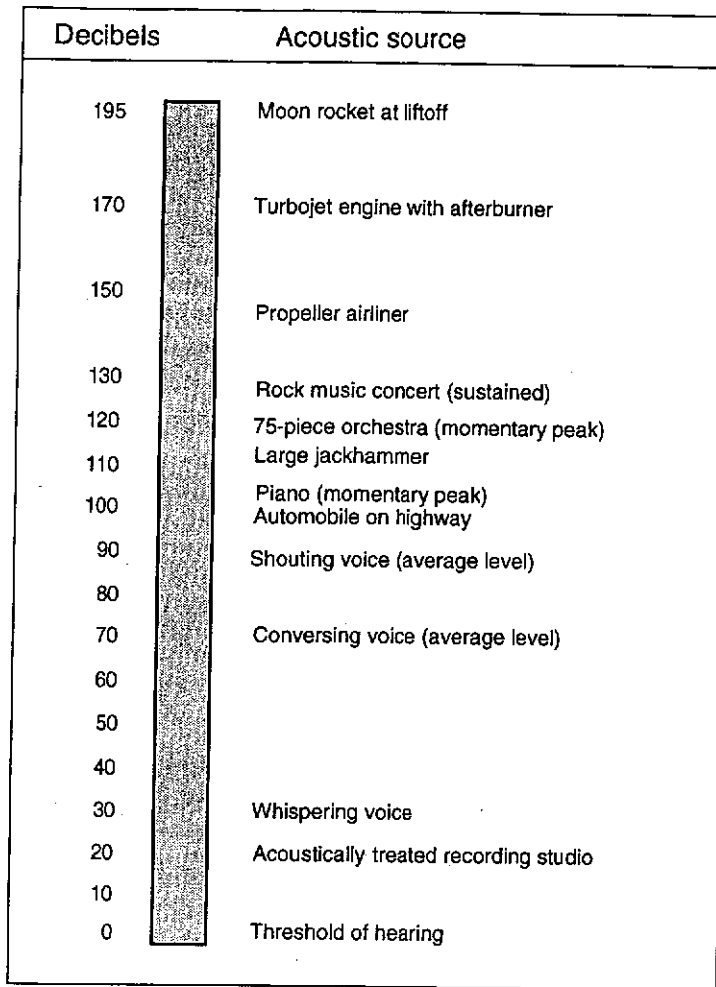


Figure 1.22 Typical acoustic power levels for various acoustic sources. All figures are relative to 0 dB = 10^{-12} watts per square meter.

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

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:

$$\text{maximum dynamic range in decibels} = \text{number of bits} \times 6.11.$$

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.

Oversampling

So far we have mainly discussed linear PCM converters. A linear PCM DAC transforms a sample into an analog voltage in essentially one straightforward step. In contrast to linear PCM converters, oversampling converters use more samples in the conversion stage than are actually stored in the recording medium. The theory of oversampling is an advanced topic, however, and for our purposes here it is sufficient to present the basic ideas, leaving ample references for those who wish to investigate the topic further.

Oversampling is not one technique but a family of methods for increasing the accuracy of converters. We distinguish between two different types of oversampling:

1. Multiple-bit oversampling DACs developed for compact disc players in the early 1980s by engineers at the Philips company (van de Plassche 1983; van de Plassche and Dijkmans 1984)
2. 1-bit oversampling with *sigma-delta modulation* or a related method as used in more recent ADCs and DACs (Adams 1990; Hauser 1991)

The first method converts a number of bits (e.g., 16) at each tick of the sampling clock, while the second method converts just one bit at a time, but at a very high sampling frequency. The distinction between multibit and 1-bit systems is not always clear, since some converters use a combination of these two approaches. That is, they perform multibit oversampling first, and then turn this into a 1-bit stream that is again oversampled.

Multiple-bit Oversampling Converters

In the mid-1980s many CD manufacturers used a DAC chip set designed by Philips that introduced the benefits of oversampling technology to home listeners. These converters take advantage of the fact that digital filters can provide a much more linear phase response than the steep brickwall analog filters used in regular DACs. (ADCs based on this concept have also been made, but we restrict the discussion here to the DAC side.) In a CD player, 44,100 16-bit samples are stored for each second per channel, but on playback they may be *upsampled* four times (to 176.4 KHz) or eight times (to 352.8 KHz), depending on the system. This is accomplished by interpolating three (or seven) new 16-bit samples in between every two original samples. At the same time all of the samples are filtered by a linear phase digital filter, instead of a phase-distorting brickwall analog filter. (This digital filter is a *finite-impulse-response* filter; see chapter 10.)

Besides phase linearity, a main benefit of oversampling is a reduction in quantization noise—and an increase in signal-to-noise ratio—over the audio bandwidth. This derives from a basic principle of converters stating that the total quantization noise power corresponds to the resolution of the converter, independent of its sampling rate. This noise is, in theory, spread evenly across the entire bandwidth of the system. A higher sampling rate spreads a constant amount of quantization noise over a wider range of frequencies. Subsequent lowpass filtering eliminates the quantization noise power above the audio frequency band. As a result, a four-times oversampled recording has 6 dB less quantization noise (equivalent to adding another bit of resolution), and an eight-times oversampled recording has 12 dB less noise. The final stage of the systems is a gently sloping analog lowpass filter that removes all components above, say, 30 KHz, with insignificant audio band phase shift.

1-bit Oversampling Converters

Although the theory of 1-bit oversampling converters goes back to the 1950s (Cutler 1960), it took many years for this technology to become incorporated into digital audio systems. The 1-bit oversampling converters constitute a family of different techniques that are variously called *sigma-delta*, *delta-sigma*, *noise-shaping*, *bitstream*, or *MASH* converters, depending on the manufacturer. They have the common thread that they sample one bit at a time, but at high sampling frequencies. Rather than trying to represent the entire waveform in a single sample, these converters measure the differences between successive samples.

1-bit converters take advantage of a fundamental law of information theory (Shannon and Weaver 1949), which says that one can trade off sample width for sample rate and still convert at the same resolution. That is, a 1-bit converter that “oversamples” at 16 times the stored sample rate is equivalent to a 16-bit converter with no oversampling. They both process the same number of bits. The benefits of oversampling accrue when the number of bits being processed is greater than the number of input bits.

From the standpoint of a user, the rate of oversampling in a 1-bit converter can be a confusing specification, since it does not necessarily indicate how many bits are being processed or stored. One way to try to decipher oversampling specifications is to determine the total number of bits being processed, according to the relation:

oversampling factor × *width of converter*.

For example, a “128-times oversampling” system that uses a 1-bit converter is processing 128×1 bits each sample period. This compares to a traditional 16-bit linear converter that handles 1×16 bits, or 8 times less data. In theory, the 1-bit converter should be much cleaner sounding. In practice, however, making this kind of determination is sometimes confounded by converters that use several stages of oversampling and varying internal bit widths.

In any case, all the benefits of oversampling accrue to 1-bit converters, including increased resolution and phase linearity due to digital filtering. High sampling rates that are difficult to achieve with the technology of multibit converters are much easier to implement with 1-bit converters. Oversampling rates in the MHz range permit 20-bit quantization per sample.

Another technique used in 1-bit oversampling converters is *noise shaping*, which can take many forms (Hauser 1991). The basic idea is that the “re-quantization” error that occurs in the oversampling process is shifted into a high-frequency range—out of the audio bandwidth—by a highpass filter in a feedback loop with the input signal. This *noise-shaping loop* sends only the requantization error through the highpass filter, not the audio signal.

The final stage of any oversampling converter is a decimator/filter that reduces the sampling rate of the signal to that required for storage (for an ADC) or playback (for a DAC) and also lowpass filters the signal. In the case of a noise shaping converter this decimator/filter also removes the requantization noise, resulting in dramatic improvements in signal-to-noise ratio. With *second-order noise shaping* (so called because of the second-order highpass filter used in the feedback loop), the maximum signal-to-noise level of a 1-bit converter is approximately equivalent to 15 dB (2.5

bits) per octave of oversampling, minus a fixed 12.9 dB penalty (Hauser 1991). Thus an oversampling factor of 29 increases the signal-to-noise ratio of a 16-bit converter by the equivalent of 10 bits or 60 dB.

For more details on the internals of oversampling noise-shaping converters, see Adams (1986, 1990), Adams et al. (1991), and Fourré, Schwarzenbach, and Powers (1990). Hauser (1991) has written a survey paper that explains the history, theory, and practice of oversampling techniques in tutorial form and contains many additional references.

Digital Audio Media

Audio samples can be stored on any digital medium: tape, disk, or integrated circuit, using any digital recording technology, for example, electromagnetic, magneto-optical, or optical. Using a given medium, data can be written in a variety of *formats*. A format is a kind of *data structure* (see chapter 2). For example, some manufacturers of digital audio workstations implement a proprietary format for storing samples on a hard disk. For both technological and marketing reasons, new media and formats appear regularly. Table 1.2 lists some media and their distinguishing features.

Some media are capable of handling more bits per second and so have the potential for higher-quality recording. For example, certain digital tape recorders can encode 20-bits per sample with appropriate converters (Angus and Faulkner 1990). A hard disk can handle 20-bit samples at rates in excess of 100 KHz (for a certain number of channels at a time), while for semiconductor media (memory chips) the potential sample width and sampling rate are much greater.

Another characteristic of media is lifespan. Archival-quality optical disks made of etched tempered glass and plated with gold will last decades and can be played many thousands of times (Digipress 1991). Magnetic media like DAT and floppy disks are inexpensive and portable, but not nearly as robust.

An outstanding advantage of digital storage media is that one can transfer the bits from one medium to another with no loss. (This assumes compatibility between machines and absence of copy-protection circuits, of course.) One can clone a recording any number of times—from the original or from any of the copies. It also means that one can transfer a recording from an inexpensive serial medium (such as DAT) to a random-access medium (such as disk) that is more suited to editing and processing. After one is done editing, one can transfer the samples back to DAT. These transfers

Table 1.2 Digital audio media

Medium	Serial or random access	Notes
Stationary head (magnetic tape)	Serial	Typically used for professional multitrack (24, 32, 48 track) recording; several formats coexist; limited editing.
Rotary head videotape (magnetic tape)	Serial	Professional and consumer formats; consumer videocassettes are inexpensive; two machines needed for assembly editing (see Chapter 16); several tape formats (U-matic, Beta, VHS, 8 mm, etc.) and three incompatible international video encoding formats (NTSC, PAL, SECAM)
Rotary head audiotape (magnetic tape)	Serial	Professional Nagra-D format for four-channel location recording.
Digital Audio Tape (DAT) (magnetic tape)	Serial	Small portable cassettes and recorders; compatible worldwide; some machines handle SMPTE timecode (see Chapter 21)
Digital Compact Cassette (DCC) (magnetic tape)	Serial	A digital format that can also be used in traditional analog cassette recorders. Uses data compression. Inferior sound quality as compared to CD format.
Hard disks (magnetic and optical)	Random	Nonremovable hard disks are faster (several milliseconds access time); removable hard disks are convenient for backup and transporting of sound samples. Note: a removable optical hard disk attached to a computer is usually not the same format as an audio CD, though they may look similar.
Floppy diskettes (magnetic)	Random	Floppy disks are small, inexpensive and convenient, but they are slow and can store only short sound files. Not reliable for long-term storage.
Sony Mini Disc (MD) (magnetic)	Random	A floppy disk format for sound that employs data compression. Inferior sound quality with respect to CD format.
Compact disc (optical)	Random	Small thin disc storing maximum of 782 Mbytes for a 74-minute disc; archive-quality disks last decades; can playback images as well as audio. Various levels of audio quality, depending on the application, from speech grade (CD-ROM) to very high

Table 1.2 (cont.)

Medium	Serial or random access	Notes
		<p>Medium</p> <p>fideliy (20-bit format). Slow access and transfer rate compared to other random-access media (Pohlmann 1989b, d)</p>
Semiconductor memory (electronic)	Random	<p>Semiconductor memory (electronic)</p> <p>Very fast access time (less than 80 nanoseconds typically); excellent for temporary storage (for editing) but too expensive for large databases.</p>

are accomplished through *digital input/output connectors* (hardware jacks on the playback and recording systems) and *standard digital audio transmission formats* (software protocols for sending audio data between devices; see chapter 22).

Synthesis and Signal Processing

As we have seen, sampling transforms acoustical signals into binary numbers, making possible digital audio recording. For musical purposes, the applications of sampling go beyond recording, to *synthesis* and *signal processing*. Synthesis is the process of generating streams of samples by algorithmic means. The six chapters in part II enumerate the many possible paths to synthesis.

Signal processing transforms streams of samples. In music we use signal processing tools to sculpt sound waves into aesthetic forms. Typical audio applications of signal processing include the following:

- Dynamic range (amplitude) manipulations—reshaping the amplitude profile of a sound
- Mixing—combining multiple tracks of audio, including crossfading
- Filters and equalizers—changing the frequency spectrum of the sound
- Time-delay effects—echoes, chorus effect, flanging, phasing
- Convolution—simultaneous time-domain and frequency-domain transformations
- Spatial projection, including reverberation
- Noise reduction—cleaning up bad recordings

- Sample rate conversion—with or without pitch shift
- Sound analysis, transformation, and resynthesis
- Time compression/expansion—changing duration without affecting pitch, or vice versa

Although it is a relatively new field, *digital signal processing* (DSP) has blossomed into a vast theoretical science and applied art. Parts III and IV explain essential concepts of DSP as they pertain to music.

Conclusion

This chapter has introduced fundamental concepts of digital audio recording and playback. This technology continues to evolve. In the realms of AD and DA conversion, signal processing, and storage technology—where there is always room for improvement—we can look forward to new developments for many years to come.

While recording technology marches on, the aesthetics of recording take this technology in two opposing directions. The first is the “naturalist” or “purist” school of recording, which attempts to recreate the ideal concert hall experience with as little artifice as possible. Listening to these recordings, it is as if we are suspended in air (where the microphones are) in the ideal listening location, eavesdropping on a virtuoso performance. The opposite approach, no less valid, is often employed in pop, electronic, and computer music: the creation of an artificial sound stage in which sources can move and we are presented with illusions such as sounds emanating from different spaces simultaneously. These illusions are created by the signal processing operations described in part III.