Chapter 1

What Is Signal Processing? A Look At the CD Player

Signal processing is an immense and diverse field. There are perhaps 50,000 engineers who regard signal processing as their specialty and hundreds of thousands more whose work involves signal processing.[1] It is also a field that did not exist 50 years ago[2] and one that remains mysterious, or quite unknown, to most people, though many of its tasks, such as analog-to-digital conversion, error-correction coding, speech synthesis, and image compression, have become familiar to laymen of the present communications- and computer-dominated world.

Before attempting a definition of signal processing in general, let us take a look at an everyday device, the compact-disk player, whose performance depends upon several types of signal processing.

To better understand the CD player, consider first the acoustic phonograph, that is, one without electronic amplification. In the studio a record is made as follows: pressure waves in the air move a diaphragm, the center of which is connected to a stylus, and the stylus cuts a groove whose undulations record the sound. In playback, motion of the groove under the

---

[1]Craig Marven and Gillian Ewers, *A Simple Approach to Digital Signal Processing* (New York: John Wiley & Sons, 1996), p. 3.

[2]The term 'signal processing' did not exist 50 years ago, nor was there much of what would today be called digital signal processing; there were, however, a great many engineers concerned with analog signals (as for telephone or radio), and many of their activities, such as designing filters, would today be regarded as signal processing.

needle causes vibrations in the tone-arm; the vibrations travel to a diaphragm at the base of a speaking horn, which produces the pressure waves in the air. Electric recording, introduced in the 1920s, made this operation more complicated by converting the acoustic signal to an electrical signal both in making the recording and in playing it back. (See Figure 1.) The vital advantage was that electrical signals, unlike acoustic signals, could easily be amplified.

The basic operation of a CD player is almost as simple. The electrical signal is converted from a continuously varying waveform to a sequence of binary digits, 0s and 1s. These are then recorded on a spiral path around the disk. (See Figure 2.) There are pits or "dimples" in the track, and the falling and rising edges of a dimple correspond to the falling and rising edges of the pulse train. In playback a laser beam tracks the spiral path. Where there are height changes, the beam is not reflected sharply, but attenuated. A light sensor detects the reflections, producing again the sequence of 0s and 1s,

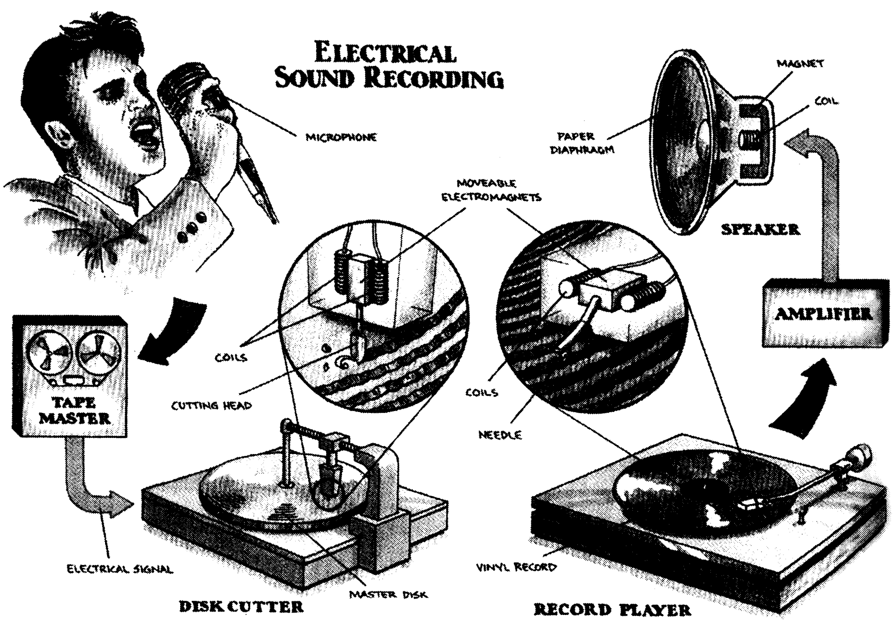

FIGURE 1. The drawing shows how electrical sound-recording works. A microphone converts the sound waves to an electrical signal, which, after recording on a magnetic tape, drives electromagnets to cut a track in the master disk. When the record is played, the motion of the needle generates a tiny current, which is amplified and sent to the speakers, which convert the signal back to sound waves. (Reproduced by permission of the artist.)

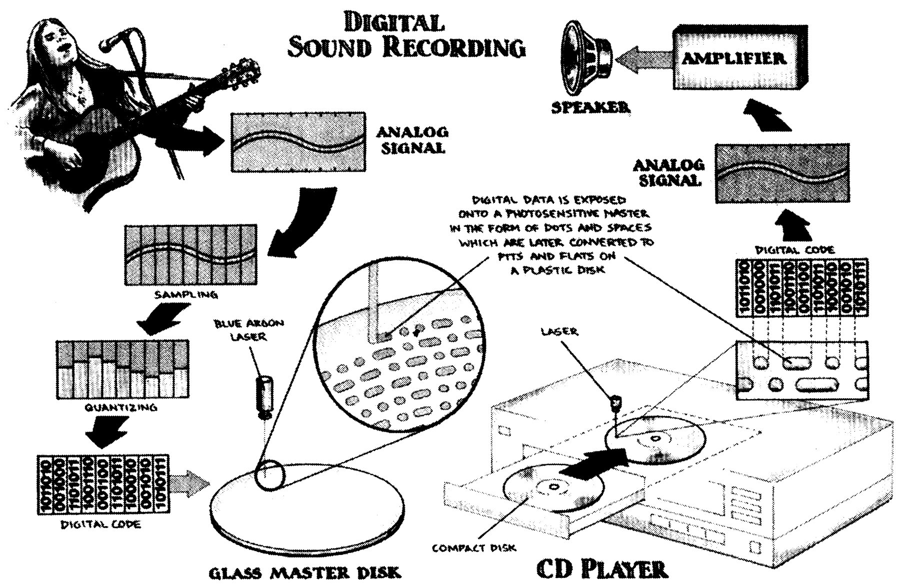

FIGURE 2. The drawing shows how digital sound-recording works. The electrical signal coming from the microphone is digitized, that is, converted from a continuous waveform to a sequence of binary digits by sampling and quantizing. The sequence is recorded on a compact disk as a series of closely spaced pits. In the CD player a laser pickup traces the spiral track, generating the original sequence of binary digits. The digital code is then converted to an analog signal, which, after amplification, drives the speakers. (Reproduced by permission of the artist.)

which is then converted back to an analog sound signal.[3] (The sequence of 0s and 1s is a long one: the digital sound consists of 44,100 samples per second and each sample is represented by 16 bits, so one can calculate that the 74 minutes of music that a CD can hold require some 3.1 billion bits. Actually, error correction and other coding requirements mean that three times this number is required.)[4]

CD players provide better sound than conventional record players: increased bandwidth (a lower bound of 20 hertz rather than 30 hertz), flatter

[3]Fred Guterl, "Compact disc" (*IEEE Spectrum*, vol. 25 (1988), no. 11, pp. 102–108); Senri Miyaoka, "Digital audio is compact and rugged" (*IEEE Spectrum*, vol. 21 (1984), no. 7, pp. 35–39); and John G. Truxal, *The Age of Electronic Messages* (Cambridge, MA: MIT Press, 1990), p. 245.

[4]Guterl *op. cit.* Since there are some 31 million seconds in a year, if a person were to read off the sequence of 0s and 1s without stop, one per second, it would take 300 years, longer than even a teenager would want to spend with a CD.

frequency-response curve (plus or minus 0.5 decibels rather than 3 decibels), greater dynamic range (90 dB rather than 70 dB), better signal-to-noise ratio (90 dB rather than 60 dB), much smaller harmonic distortion (0.01 percent rather than 1 to 2 percent), and much better separation between the two stereo channels (90 dB rather than 30 dB). Other advantages are smaller size, longer playing-time, and greater durability (there being no mechanical contact in playback to cause wear of the disks). In addition, one has immediate access to different tracks, and programmed play is possible. Finally, CDs are almost immune to the dust, scratches, and fingerprints that disturb the sound of LPs.[5] It is not, then, surprising that in less than a decade after their introduction in the early 1980s, CD players drove record players from the market.[6]

While the vital advantage of electric recording was that it permitted amplification, the vital advantage of CD recording is that it permitted many types of signal processing, which confer the advantages just named. The electron tube made amplification possible; microelectronics made complex, real-time signal processing possible.

One of the most important of the signal-processing tasks is the conversion of the signal from analog to digital form, which is done in such a way that the original waveform can be recreated with a high degree of accuracy. The digital encoding itself confers an important advantage: slight imperfections in the disk or low levels of electrical noise in the digital circuits—ones small enough not to change a 0 into a 1 or vice versa—cause no degradation of the signal at all.

Errors inevitably occur, but the signal-processing capability of the CD player can usually correct them. First, there is redundancy in the digital coding so that if a small number of bits are misread, the original sequence is automatically restored. Second, successive binary digits are not recorded together, but physically separated. This means that if a defect in the disk af-

---

[5]Peter J. Bloom, "High-quality digital audio in the entertainment industry: an overview of achievements and challenges" (*IEEE ASSP Magazine*, vol. 2 (1985), no. 4, pp. 2–25), and Miyaoka *op. cit.*

[6]Steven Lubar, *Infoculture: The Smithsonian Book of Information Age Inventions* (Boston: Houghton Mifflin, 1993), p. 195. In the late 1990s there is something of an LP-record cult that believes in the superiority of vinyl records to CDs. *Time* magazine reported [14 July 1997, p. 20] that the number of LPs sold had increased from 625,000 in 1994 to 1,100,000 in 1996, and that at the end of June 1997 2,200,000 had already been sold that year. And turntables, as part of home-entertainment centers, continue to be sold; *Atlantic Monthly* reported [December 1997, p. 106] that there were more high-end turntables to choose from in 1997 than ever before.

INTERVIEWER: Once people started talking about "Digital Signal Processing," did that retroactively define "signal processing?" Was "signal processing" a coherent field before the mid-'60s?

JAMES KAISER: We never really thought about it as signal processing. In the electronics it was maybe signal generation or signal shaping or that kind of thing. I can never remember us saying, "We'll do it by signal processing." But then once you've gone digital, now also you realize, "Gee, stock market prices are digital signals. They are sampled signals. And a sampled signal from a musical instrument, a music signal. These all have something in common." It's not just this little communications thing that you were working on, or this radar thing. It's very broad. The operations you are doing are, essentially, the same kind of things in each case. And the development of computer processor power is what prompted that realization. It was the tool that allowed you, with the software, to do things on signals no matter where they came from. . . .[1]

INTERVIEWER: In looking at your 1965 book I was very impressed by how much you yourself used computers in those early years, both to do calculations and to simulate phenomena. You must have been very early in those applications.

JAMES FLANAGAN: We found that you could get useful solutions for just about any signal process that you could turn into difference equations. That's the way we did most of the speech synthesis, with vocal tract simulation by differential equations, then solving them simultaneously. Not in real time, obviously. This early speech-processing work required sampled data, and the understanding of sampled data signals. But it also needed all kinds of filtering and spectral analysis.

I think the whole area of digital signal processing, particularly digital filter design, was driven by the speech processing community. I made a mark here. Roger Golden and I did something called the phase vocoder in 1966. This required simulation of electrical filters. We had some infinite impulse response filters that approximated Bessel characteristics. We hadn't thought about finite impulse response filters very much then—they were developed a bit later—but we used them to good effect. The whole business of having to do filtration of signals, spectral analysis, and algorithmic operations on sampled data, of recognizing what happens when you square a signal or take a square root, or watching what happens to the bandwidth, this all drove the development of digital signal processing at that time. There might have been a parallel in image processing that I do not know about, but speech processing was a research activity that galvanized digital signal techniques.[2]

[1]James Kaiser oral-history interview 11 February 1997, p. 36.

[2]James Flanagan oral-history interview 8 April 1997, pp. 16–17.

fects a large number of bits, these will be drawn from different sampling areas of the original signal.[7] Thus, a dust particle, a scratch, a fingerprint, or even a small hole through the disk will not affect the sound. Finally, larger errors, which the program cannot fully correct, are detected and minimized through digital filtering.[8]

The CD player exhibits sophisticated signal processing, and many of the tasks could not have been performed 25 or even 10 years earlier. In this monograph we will review the evolution, over the past 50 years, of signal processing, describing the development of new and improved techniques and pointing out many of their applications. We will review also the evolution of the professional organization today known as the IEEE Signal Processing Society, which played a vital role in the technological development and in the emergence of a recognized profession of signal processing.[9] (See Figure 3.)

A signal may be defined as a time-variant quantity that conveys information.[10] In the signal-processing world, that quantity almost always is, or is converted to, an electrical quantity such as current or voltage.[11] And, increasingly in recent decades, continuously varying signals are converted to sequences of bits, which is to say analog signals are converted to digital signals. A typical signal-processing system is depicted in Figure 4, where the upper row shows the signal-processing tasks and the lower row shows how the signal changes.[12]

An analog signal, such as might come from the pickup cartridge of a phonograph, passes through a low-pass filter, also called an anti-aliasing filter, which removes all the frequency components of the signal above a certain frequency. The low-pass filter is called an anti-aliasing filter, because if it were not present the very-high-frequency components might produce, in the subsequent processing of the signal, lower-frequency "aliases". The next

---

[7]Truxal *op. cit.*, p. 250.

[8]Miyaoka *op. cit.*

[9]As noted in the Preface, a much more complete history of the Signal Processing Society is contained in the monograph by Frederik Nebeker, *The IEEE Signal Processing Society: Fifty Years of Service, 1948–1998* (New Brunswick, NJ: IEEE History Center, 1998).

[10]James A. Cadzow, *Foundations of Digital Signal Processing and Data Analysis* (New York: Macmillan, 1987), p. 1.

[11]This is true also in the biological world for organisms with central nervous systems: sense organs convert visual, aural, tactile, olfactory, and other signals into electrical signals to convey them to the brain.

[12]This description of a typical signal-processing system and the drawing in Figure 4 come from Chapter 3 (pp. 31–72) of Marven and Ewers, *op. cit.*

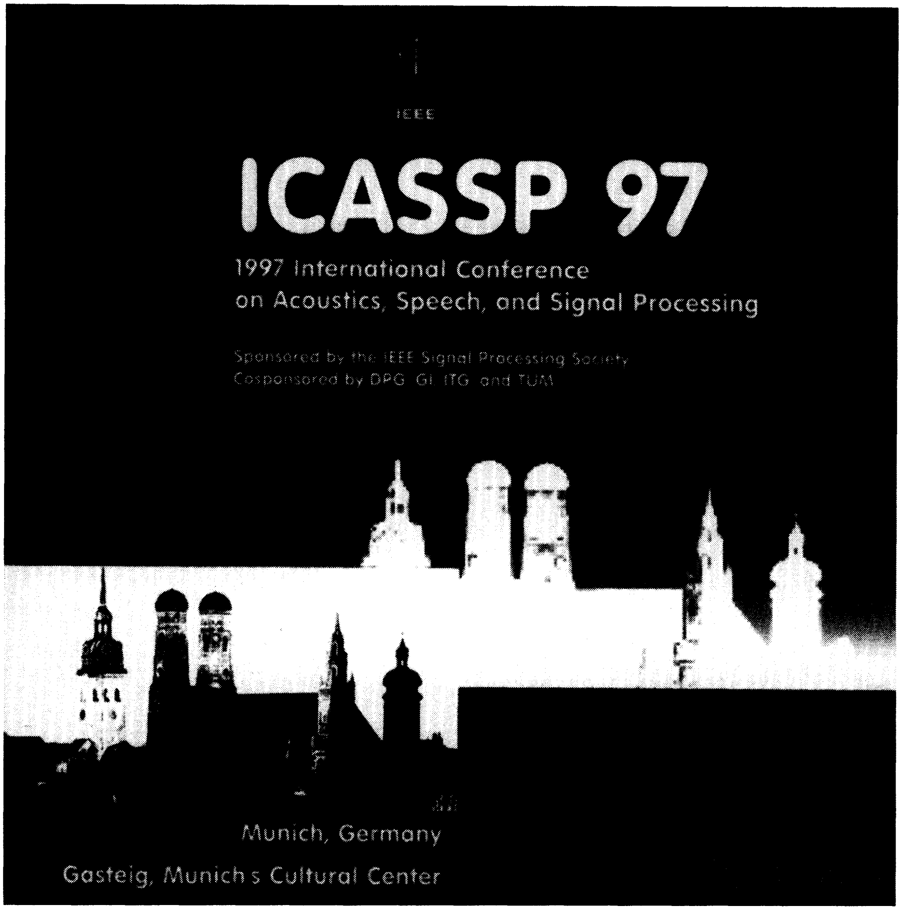

FIGURE 3. The largest of the annual conferences sponsored by the IEEE Signal Processing Society is the International Conference on Acoustics, Speech, and Signal Processing (ICASSP). At the 1997 ICASSP in Munich, attendees received a CD-ROM containing the entire proceedings of the 4-day conference. (Reproduced by permission of IEEE.)

step, which produces the staircase waveform, is called sample-and-hold: the input signal is sampled at particular points in time, and the sampled value is held constant until the next sampling. Then the heights are quantized, that is, an analog-to-digital converter assigns one of a discrete set of values to each of the heights. Here the sampled values are sorted into classes according to amplitude. Each height may then be represented as a binary number. If the number contained, for example, 6 bits, this would allow each height to be coded as one of 2^6 or 64 different levels. The entire signal has thus been turned into a sequence of binary numbers. Next comes some digital-

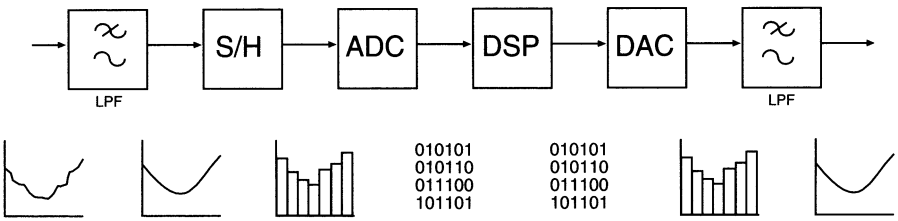

FIGURE 4. The two rows show a typical signal-processing sequence, the upper row showing the signal-processing tasks and the lower row the signal changes. Missing from this sequence is the particular signal-processing task for which the whole process is undertaken. (Redrawn after Figure 3.55 of Craig Marven and Gillian Ewers, *A Simple Approach to Digital Signal Processing* (New York: John Wiley & Sons, 1996).)

signal-processing task. This could be enhancing certain features of the signal; or extracting the desired information from a noisy signal; or analyzing a waveform into component frequencies; or encrypting information; or compressing the information for transmission and de-compressing it after transmission. Then a reverse process recreates an analog waveform: digital-to-analog conversion produces a staircase waveform, which passes through a low-pass filter to remove high-frequency components, thus smoothing the waveform.

Every step of this sequence has been subjected to great scrutiny and improved upon by ingenious design. For example, the quantization need not be uniform (with equal-sized bands), but may be tailored to the type of signal being processed (with, for example, closely spaced bands where variation in the signal is most significant). The quantization scheme may even change from moment to moment to deal better with the changing signal; this technique is called adaptive quantization.

The conversion from an analog signal to a sequence of bits need not entail any significant loss of information. If a signal contains only frequencies below some limit, say f_{max}, then the original signal can be reconstructed exactly from sampled values, provided the sampling rate is at least twice f_{max}.[13] There is necessarily some loss of information when the sampled val-

[13]This is usually called the Nyquist rate, in honor of Bell Labs engineer Harry Nyquist, who in the 1920s studied the maximum signaling rate that could be used over a telegraph channel of given bandwidth [Sidney Millman, ed., *A History of Engineering and Science in the Bell System: Communication Sciences (1925–1980)* (AT&T Bell Telephone Laboratories, 1984), pp. 9–10]. See Harry Nyquist, "Certain topics on telegraph transmission theory" (*Transactions of the AIEE*, vol. 47 (1928), pp. 617–644).

ues are digitized, but this loss may be made as small as desired by increasing the number of bits given to each sample.

There are two great advantages to having information in digital form. First, as we have already mentioned, digital signals frequently may be reconstructed exactly despite noise and error in transmission, storage, or retrieval. Second, digital signals may be manipulated by electronic circuits or computers in a quite unlimited variety of ways. The remainder of this monograph will make abundantly clear the real-world significance of these two advantages.

Signal processing is not the transmission of signals, as through telephone wires or by radio waves, but the changes made to signals so as to improve transmission or use of the signals. Among the processes studied and devised by signal-processing engineers are filtering, coding, estimating, detecting, analyzing, recognizing, synthesizing, recording, and reproducing. Though signal processing concerns both analog and digital techniques, the field is increasingly dominated by digital techniques. Indeed, as we will see in Chapter 4, the emergence of digital techniques in the 1960s and 1970s played a large part in creating a community of engineers concerned with signal processing.[14] Another distinction is between real-time signal processing and non-real-time signal processing; the former is restricted to procedures that can be carried out as fast as the signal arrives. The discipline of signal processing encompasses the development of theories, the creation of algorithms, the implementation of algorithms in hardware, and the application of software and hardware to communications, instrumentation, and other types of tasks; that is to say, it encompasses theory, techniques, hardware, and applications.

At the end of the 20th century, signal processing is a vital technology in many areas: communications, information processing, consumer electronics, control systems, radar and sonar, medical diagnosis, seismology, and scientific instrumentation generally. In addition to the wide range of application areas, there is a wide range of signal-processing tasks. Examples of these tasks are removing echo from telephone lines, scrambling cellular-phone conversations, controlling the suspension of an automobile so that it responds to road conditions, enabling satellite imaging systems to resolve tiny objects on the ground, and making internal organs stand out in CAT scans. That so many issues related to the processing of

[14]Digital signals are not new—witness signal fires, ship-to-ship flag communications, Morse code, and Braille—but their use in communications, computing, instrumentation, and control systems has increased enormously in the past several decades.

signals have arisen in recent decades and have assumed such economic importance is a result of the fact—and further confirms it—that ours is an Information Age.

Today we make use of a great variety of electronic means of communication and entertainment. Computers are involved in conducting most businesses, and computers—in the form of microprocessors—are found all around us in offices, in homes, and in cars. In earlier times, most people worked either in agriculture or industry. Today a large part of the work force are information workers, such as teachers, researchers, designers, bankers, lawyers, insurance agents, advertisers, broadcasters, and journalists.[15] In the next chapter we will, in a quite partisan way, assign a definite date to the beginning of the Information Age.

---

[15]According to one estimate, 55 percent of those employed in the United States in 1990 were information workers [Karen Wright, "The road to the global village" (*Scientific American*, vol. 262 (1990), no. 3, pp. 83–94)]. The importance of information processing in industry generally is indicated by the fact that in the United States in the mid 1990s some 40 percent of industry's capital budget was for computer and telecommunications equipment [W. Wayt Gibbs, "Taking computers to task" (*Scientific American*, vol. 277 (1997), no. 1, pp. 82–89)]. If one looks specifically at the computing, consumer electronics, and telecommunications industries, the picture is still impressive: in 1996 in the U.S. these industries employed nearly 4.3 million people and created 6.2 percent of the national output of goods and services, which was a larger part of the economy than the automobile industry, the construction industry, or the food industry [*IEEE Spectrum*, vol. 35 (1998), no. 1, p. 16T4].