Chapter 20
Digital Audio

In the past several decades, digital electronics, as represented by computers, have made great changes in the way we do things. Audio is no exception. The purpose of this chapter is to describe important principles involved in digital electronics and its application to audio. It especially contrasts digital audio with analog audio.

20.1 Digital vs Analog

Even in a digital world, audio begins and ends with an analog signal. The output of a microphone or other acoustical transducer is an electrical signal that is analogous to the acoustical waveform that produced it. The signal remains in analog form until converted to digital form by an analog to digital converter (ADC). Once in the digital domain the sound can be transmitted, processed, or stored as digital data, i.e., as a series of "ones" and "zeros." To hear the sound again requires that it first be transformed back to analog form by a digital to analog converter (DAC). The analog signal can then be amplified to produce an analogous current that will drive loudspeakers to make an analogous acoustical signal. What is different about digital audio is the digital means of transmission, processing, and storage.

Transmission: The sound that comes over the Internet, or in digital broadcasting, or DMX digital cable is digital transmission. It is transmitted as a series of high and low voltages, representing ones and zeros. In this way it is different from analog transmission, as in AM or FM radio, where voltages follow the original acoustical material.

Processing: Digital processing exploits two splendid advantages of the digital domain, the availability of delay and noise immunity.

Digital methods can delay a signal by any desired amount of time because the signal is readily available in digital memory. There it can be processed in detail, like any other digital file, by a computer or a computer-like device. An obvious

W.M. Hartmann, *Principles of Musical Acoustics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-1-4614-6786-1_20,

application is an ambiance synthesizer, where a signal (music or speech) is delayed and added back many times to the original to simulate the reflections in a room. The delayed signal can be delayed by a small amount to simulate early reflections and it can be given many different long delays and added back to simulate reverberation. The delayed signals can be attenuated and can be filtered to represent the absorption of high frequencies on reflecting surfaces. Less obvious, a delay operation is at the heart of digital filtering, which can change the spectral profile of a signal, just like an analog filter, but with more precision and with more flexibility.

An analog signal processor—an analog filter for example—is a *causal* device. Its output at any given time depends on the input signal at that time, or at times in the past. Such a processor cannot make any use of the signal as it will be in the future because it has no knowledge of the future. This seems to us to be a reasonable limitation. We ourselves live with this limitation every day. It is a constraint called "causality." Digital processing devices do not have to live with this constraint. An input signal can be stored in memory, processed using all the information in the memory, and then reproduced—with a time lag that depends on the memory length, of course, but normally only a few milliseconds. Thus, digital devices can be *acausal*, meaning that what happens now can be affected by what will happen in the future. Freedom from causality opens up an entire new world of signal processing opportunities.

Noise immunity is an important advantage of processing in the digital domain, because the analog domain is not immune to noise. As analog processing increases, noise and distortion are added—combining signals in a mixer, filtering them to improve frequency response, filtering again to eliminate some problem, running the signals across the room in cables, etc. Each added stage or process degrades the signal somewhat. This limits the number of operations that can be chained. It particularly limits the number of stages (hence the precision) of filtering that can be done in the analog domain. By contrast, digital processing is immune to noise. Adding another digital process to a signal path, or adding ten more digital processes, does not increase the amount of noise or distortion. That is because digital operations are numerical. Exercise 1 illustrates this point.

Among the most important processing operations is the simple operation of copying. With analog recordings there is some signal degradation with each successive recording. With digital recording and storage there is no degradation from one generation to the next. (See Exercise 13.)

Storage: Compact discs, DVDs, digital audio tape (DAT), diskette, and hard drives are examples of digital storage media. So are computer RAM and flash drives. In their physical principles, some of these storage devices are not very different from the devices that have traditionally stored sound in analog form. The compact disc, or DVD, uses optical storage, resembling the optical storage of sound tracks on film. The DAT, the hard drive, and the diskette are magnetic media, storing digital data as small regions of magnetism on the medium, just like a tape recorder stores analog data. What is different about the digital storage is not the physical principle of storage; it is the matter of *what* is stored.

Digital data are stored as numbers. The numbers are represented in binary form. Each binary digit (bit) is either a "1" or a "0." The compact disc is a format that uses a 16-bit word. Therefore, an example of a meaningful sample of a waveform stored on a compact disc might be

$$0110\ 0111\ 1101\ 1001.$$

If this number is recorded on a magnetic disk, then successive places on the disk are magnetized so:

$$\downarrow\uparrow\uparrow\downarrow\ \downarrow\uparrow\uparrow\uparrow\ \uparrow\uparrow\downarrow\uparrow\ \uparrow\downarrow\downarrow\uparrow\,,$$

where the arrow points in from the north pole to the south pole of the magnet.

This number can be interpreted from right to left: There is 1 in the "1"s column, zero in the "2"s column, zero in the "4"s column, 1 in the "8"s column, and so on. The four bits on the right add up to $1 \times 1 + 0 \times 2 + 0 \times 4 + 1 \times 8 = 9$. With 16 bits it is possible to encode $2^{16} = 65536$ different numbers. (See Exercise 2.) A problem immediately arises because all the numbers are positive and we wish to store waveforms that are both positive and negative. This problem is solved by a rule, applied in audio, that says that if the first bit on the left is a 1 then the number shall be interpreted as negative. Hence the number above is positive because the first bit is "0". (Digital storage conventions are full of arbitrary rules like that. Digital methods work only because everybody agrees on the rules.)

It now becomes possible to see why digital storage is immune to noise. When a bit is stored on a hard drive the magnetism will point up or down. Of course, there will be random variations in magnetism on the disk. At some places the magnetism will be slightly larger than other places because of inevitable noise on the magnetic medium. So long as this variation in magnetism is small compared to the difference between upward and downward magnetism, the noise will not cause the direction of the magnetism to be misinterpreted. A "1" will remain a "1," and a "0" will remain a "0." The digital information is not corrupted at all by this noise. The information is preserved perfectly as shown by Fig. 20.1. This immunity can be contrasted with analog storage where any added noise is reproduced along with the signal.

20.2 Digital Noise

There is, however, a form of noise that occurs particularly in digital storage, namely quantization noise. A 16-bit word allows only 65536 different possible values of the waveform at an instant. It cannot store values of the waveform that happen to lie between allowed values and therefore it makes an error. Since allowed values are 1 unit apart, the largest possible error is 1 unit. This idea leads to a calculation of signal to noise ratio (SNR). The calculation goes as follows: First, the quantization

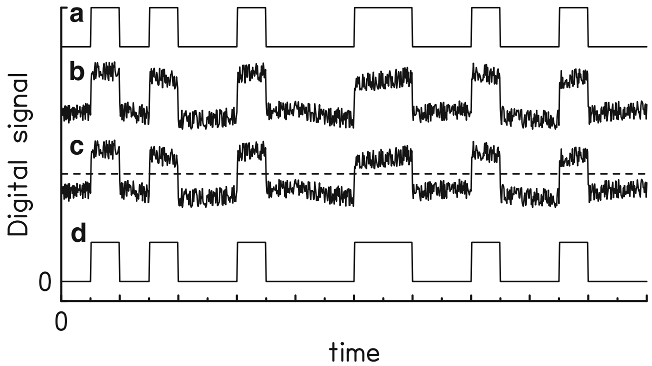

Fig. 20.1 (**a**) A 20-bit digital word, [0101 0010 0011 0010 0100], is put out to a storage device. (**b**) The storage device records the word as a signal, adding hum and noise. (**c**) The recorded signal is interpreted with a threshold set at the *dashed line*. Whenever the signal is greater than the threshold, the signal is interpreted as a 1; else it's a zero. (**d**) The interpreted signal is recovered from the storage device. It represents the word perfectly

noise is 1 unit. Second, the largest possible signal is 65536 units. Therefore, the SNR is 65536/1, and in decibels this becomes $20 \log(65536) = 96$ dB. This is an excellent ratio of signal to noise. The best vinyl records achieve a ratio of only 60 or 70 dB. In general the SNR in dB can be written as

$$SNR = 20 \log(2^N) = 20\,N \log(2) = 20 \cdot 0.301N \approx 6N. \tag{20.1}$$

This formula, as developed from left to right, says that for every extra bit in the digital word one gains 6 dB of SNR. To understand the factor of 20 that multiplies the logarithm, please refer to Exercise 8 in Chap. 10 that shows how to deal with amplitudes on the decibel scale.

As usual, it is hard to illustrate realistic noise with a drawing. The ear is so much more sensitive than the eye in this matter. To illustrate quantization noise one can imagine what would happen if the word length were 3 bits instead of 16 bits. Figure 20.2 shows a sine signal and the sampled sine signal using a 3-bit *rounding* DAC. The rounding DAC replaces a number by the nearest integer value. The quantization noise would be even worse with a *truncating* DAC. A truncating DAC replaces each number by the integer part of that number. For instance, at the time of the second sample in Fig. 20.2 the signal has a value of 2.7. The rounding DAC replaces this value by 3. The truncating DAC replaces it by 2.

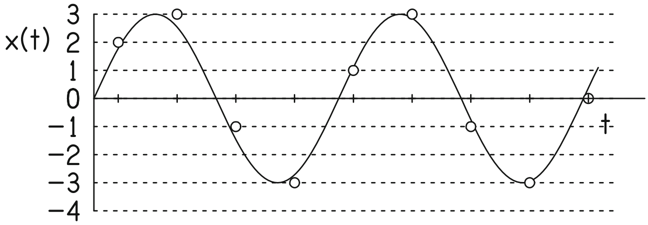

Fig. 20.2 A signal with amplitude of 3 units is shown by a *continuous line*. It is sampled regularly at times shown by *tic marks* on the *horizontal axis*. The sampled version of the signal is shown by *circles*. The analog-to-digital conversion is only 3 bits and so there are only 8 possible different values −4, −3, −2, −1, 0, +1, +2, +3. The resulting quantization error is evident in the figure. The *circles* do not agree perfectly with the *solid line*

20.3 Sampling

When sound is recorded on analog magnetic tape or vinyl records it is recorded continuously. A digital recording cannot be continuous. Instead, a digital recording consists of a series of samples of the sound. These samples are taken by the ADC at a precisely regular rate. Figure 20.3 shows such sampling.

Obviously, if the samples are taken too infrequently important details of the waveform will be missed. In order to create a faithful sampled image of the waveform, the samples must be taken frequently. But how frequently do the samples have to be? If the waveform has only low frequencies so that it changes only slowly, the sample rate does not need to be high. But high frequencies will cause the waveform to change rapidly and a high sample rate will be needed to capture those rapid changes. In the end, the correct sample rate to use depends on how high the highest frequencies are. The rule is called the Nyquist rule.

The Nyquist Rule:
The sample rate must be at least twice as high as the highest frequency in the signal to be sampled.

The sample rate for compact disc is 44,100 samples per second (sps). Therefore, the sound to be recorded digitally cannot have a frequency greater than $44,100/2 = 22,050$ Hz. That's OK. The audible range extends only to 20,000 Hz. The sample rate for DAT is slightly higher than for CDs. It is 48,000 sps.

20.4 Contemporary Digital Audio

Digital techniques are so flexible technically that they lead in many different directions. We note some of them here:

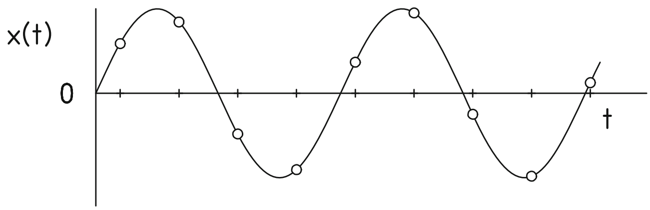

Fig. 20.3 The signal from Fig. 20.2 is sampled regularly at times shown by *tic marks* on the *horizontal axis*. The analog-to-digital conversion is 16 bits and this sampling is so accurate that it is not possible to see the difference between the true signal, shown by the *solid line*, and the samples, shown by the *circles*. Some people think that they can hear the difference though

Improved quality: The compact disc standard—16 bits and 44.1 ksps—leads to high quality sound, but there are many who think that it is not good enough. Already 24-bit recording is standard practice in digital studios. The DVD medium supports both longer word lengths—such as 24 bits—and higher sample rate. The most important reason for the flexibility of the DVD is that its storage capacity is nearly seven times greater than the CD.

Data compression: The compact disc standard requires a great deal of data (see Exercise 3). If there are 700 MB on a compact disc, then a 32-GB computer flash drive can comprehend the contents of only 45 compact discs. The problems of great amounts of digital data become really serious when transferring audio over the Internet. The slow speed of many links lead to long download times.

Reducing the amount of data that is required to transmit and store audio, while minimizing the degradation in audio quality is the object of data compression techniques. These techniques work by taking time slices of the original material, making a spectral analysis of each slice, and processing slices with a data reduction algorithm based on what listeners are likely to hear. Then the slices are joined back together to make a digital audio stream with fewer bits per second. Data reductions of one bit *out* for every ten bits *in* are possible with techniques such as MP3 encoding. The operation of compressive encoding schemes demands that the playback process be compatible with the encoding process. Tight standards must be in force.

Convergence: In the days of vinyl microgroove records there were just a few formats. Long-playing records were made at 33 RPM (revolutions per minute) and the disks were 10 in. or 12 in. in diameter. Singles were recorded at 45 RPM, and the disks were 7 in. in diameter. There was a clear reason to restrict the number of formats because introducing a new format required new hardware on the part of the user. Users resisted frequent changes of hardware.

The matter of formats is entirely different with digital media. A new format may require only a new program for interpretation, and new programs are easily and cheaply distributed—for instance by the Internet. There is no compelling reason to limit audio to two channels (stereophonic), or four, or six. Audio can be combined with video and other data. Programs in different formats can be accommodated on the same distribution medium so long as the instructions for decoding are included. The combination of all forms of communication in a comprehensive and flexible data space has been called “convergence.” Convergence requires that competing manufacturers, and competing distributors of program material, and governments agree on standards to be followed by the industry as a whole. With all interested parties jockeying for position, it is sometimes a slow-moving process.

Exercises

Exercise 1, The joy of integer numbers

You know that $1 + 1 = 2$. That is an accurate statement. How much less accurate is the statement: $1 + 1 + 1 - 1 + 1 - 0 - 1 = 2$? The answer to this question explains why digital signal processing can be chained indefinitely with no degradation in signal quality.

Exercise 2, Test the formula yourself.

A formula in the text says that with an N-bit word, it is possible to encode 2^N different numbers. This statement clearly works for $N = 2$. The formula gives 2^2 and this is equal to 4. It is easy to write down all possible combinations of two bits:

$$00 \quad 01 \quad 10 \quad 11,$$

and we discover that there are four possibilities as predicted by the formula. Show that the formula works for $N = 3$ and $N = 4$. This exercise is intended to persuade you that the formula is correct for any value of N. It is an example of *inductive* logic.

Exercise 3, Compact discs

The sample rate for compact discs is 44,100 samples per second per channel. Each sample is stored as a 16-bit word. Show that it requires at least 1.4 million bits to store a single second of stereophonic sound. There are 8 bits in a byte. How many bytes does it take to store a single second of sound?

Exercise 4, Nyquist theorem

Explain why the Nyquist theorem says that there must be at least two samples per cycle of a waveform.

Exercise 5, Digital telephones

Signals transmitted over the telephone are normally restricted to a maximum frequency of 5,000 Hz. What is the smallest sample rate that can be used to digitize a telephone conversation?

Exercise 6, No apparent audio source

The first section of this chapter explained that digital audio required an ADC to convert analog signals into digital form before the signals could be processed and stored using digital methods. Can you think of an exception to this statement?

Exercise 7, Communicating with extraterrestrials

We are going to send a capsule into deep space with artifacts from our civilization with the expectation that someone out there will find it and learn about our human culture. It is the space-age equivalent of the message in a bottle tossed into the sea. We want to include an audio recording of some human speech and music. What recording format do you recommend, a digital compact disc or an analog vinyl record?

Exercise 8, Digital ready?

In the early days of digital audio, some loudspeakers and headphones were marketed as "digital ready." What might that mean?

Exercise 9, Quantization noise?

Why is it called "quantization noise?" Why isn't it called "quantization distortion" instead?

Exercise 10, An 8-bit system

Early digital audio systems used 8-bits. Quantization by 8 bits is still used in some digital applications such as digital oscilloscopes, but 8-bits is not good enough for audio. There is too much quantization noise.

(a) How many different possible values can be represented by an 8-bit system? (b) Calculate the SNR of an 8-bit system. Do you agree that this is not good enough for audio?

Exercise 11, An 8-bit word

A byte is an eight-bit word. Show that the byte below represents the number 109.

0 1 1 0 1 1 0 1

Exercise 12, Have another byte

Write the 1s and 0s in a byte to represent the number 85.

Exercise 13, Sharing commercial music

Music in digital form can be copied, recopied, and copied again many times with no loss in quality. What are the implications for the industry that produces and distributes music?

Exercise 14, Causality

Making a decision now that is influenced by events that will happen in the future violates the principle of causality. (a) Recall an instance in which you wished that you could violate causality in this way.

Violating causality risks a time paradox where making a decision based on the future changes the future in such a way that the decision is no longer based on the future. (b) Does your recalled instance create a time paradox?

(c) The first section of this chapter says that digital signal processing can violate causality. Really? Explain why it does not involve a time paradox.

♠