Cambridge University Press
978-0-521-17042-0 - Cambridge Introductions to Music: Music Technology
Julio D'Escriván
Excerpt
More information

Chapter 1

Representing and storing sound

In this chapter we will describe the basic physics of sound and what an audio signal is. We will look at how we work out sound levels and, in simple terms, some of the maths behind that. We also discuss digital recording and therefore sampling. We describe different ways of representing sound and the way we perceive different frequencies. Finally, we discuss the capture and storage of sound as computer data and the most common standards for carrying digital audio.

Physics of sound: the audio signal

What is sound?

First, a problem: to study sound properly, it would be really useful to be able to represent it visually. But sound is, well . . . sound, a sensation for the ears. When we are able to see sound, what we see is how sound affects other things: the mythical soprano who breaks a wine-glass with a high-pitched warble, or the rippling wave patterns on the surface of a metal plate on top of a loudspeaker. In the same way, to represent sound graphically we need to show how one or more of its characteristics can be visualised. It may be how its amplitude changes over time, or how the different frequencies present in the sound change over time, or, to complicate things slightly, it can be to show how loud those different frequencies are in comparison to one another.

Although this is not a book on acoustics, let's for a minute look at what sound contains that may be useful to represent visually for our purposes. Sound is a phenomenon. This means it is something we perceive, in this case, through our sense of hearing and, in fact, through our whole body (which you will know if you have been to any concerts of drone electronica!). Sound is, in fact, what we call the sensation produced by changes in air pressure as they are perceived through our whole body and more specifically through our ears. These changes are in turn generated by the interaction between a number of objects that excite each other into motion. For instance, if we strike a large metallic sheet with a stone, we will get a loud **noise** with some accompanying resonant sound. The loud noise is produced

Cambridge University Press
978-0-521-17042-0 - Cambridge Introductions to Music: Music Technology
Julio D'Escriván
Excerpt
More information

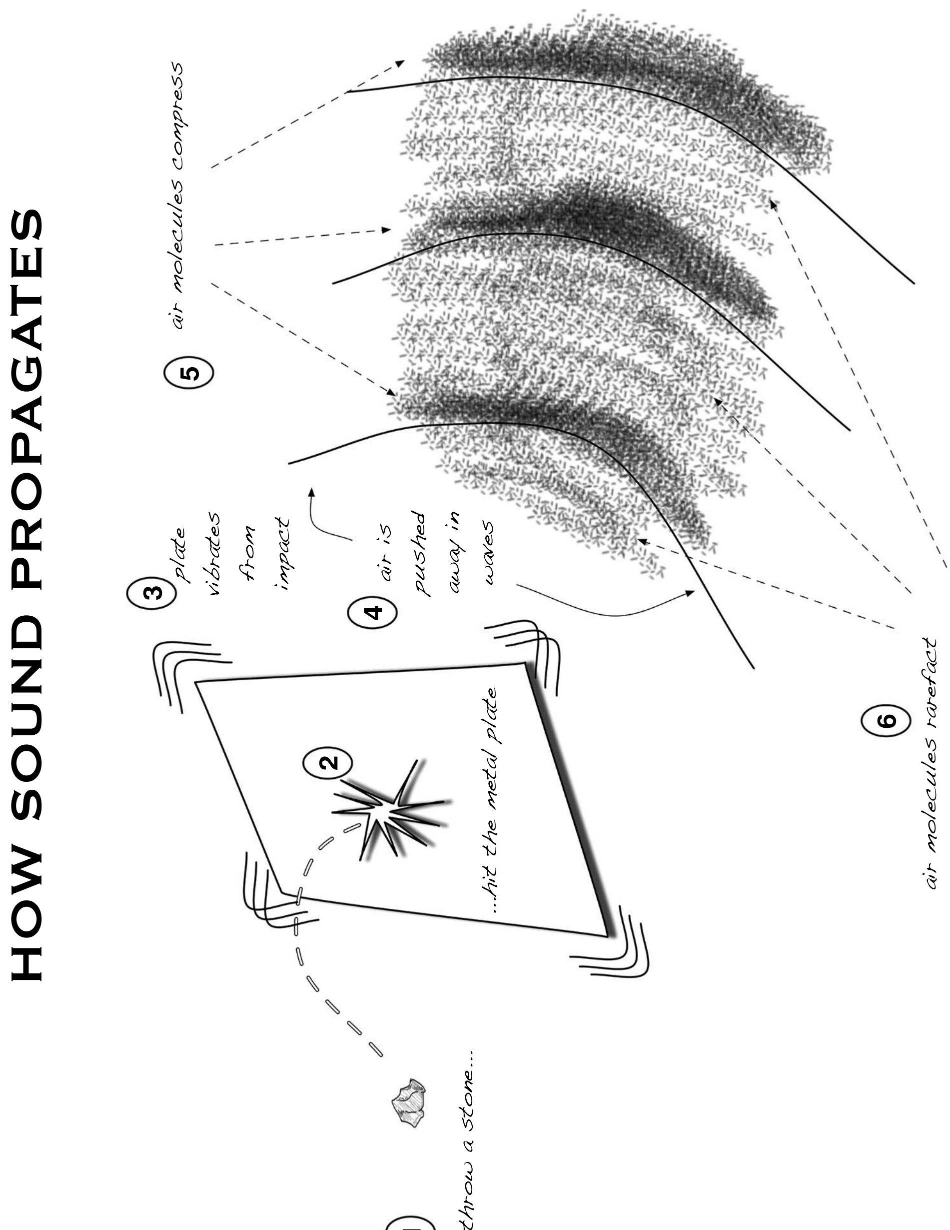

Figure 1.1 How sound propagates

by the sudden impact between stone and metal, and as the metal sheet is made to vibrate by the energy of the impact, it creates strong and sudden, but gradually decreasing, changes in air pressure as it finally comes to rest again. The molecules of air are compressed as the metal sheet bends out (called compression because the air pressure is greater as it is being pushed) and then as it bends in they are rarefied (**rarefaction** is a decrease in air pressure); see Figure 1.1.

The strength of those air-pressure changes at every instant you measure them is called *amplitude*. Our perception, with our ears, of that magnitude is called *loudness* (this is measured in *phons*, but more of that later). The speed with which those changes occur is called *frequency*. The number of changes in pressure per second, or rather *periods* of the pressure waves – the time it takes for the wave to complete its cycle – is measured in cycles per second, a unit known as the Hertz.[1] The kind or quality of the sound is called *timbre*. When the changes in air pressure are steady, regular, predictable, they are perceived as pitches we can identify: E, A, D, G, B, etc. When the changes are not regular, as in the example given in Figure 1.1, we perceive *noise* (which is a subject unto itself![2] But for now, let's just be practical and call it noise).

To model the way the air compresses and rarefacts we use a sine wave. It is the simplest sound wave possible: it represents a simple and smooth repetitive oscillation. We describe its amplitude as positive to represent compression and as negative to represent rarefaction. Figure 1.2 illustrates the different aspects we can measure in a sound wave (using a sine for simplicity).

How do we measure sound?

To represent sound, we first need to measure its characteristics in some way. Before we start we need a scale that can represent pretty large numbers within a few measurements; this is due to the way we perceive frequency and amplitude of sound. What we perceive as 'equal changes' when we measure frequencies are not in fact separated by equal magnitudes. For instance, if we play the note *middle C* (*MIDI* note 60, or C3)[3] on the piano and go up in *octave intervals* to the following Cs for a few octaves, we believe we perceive a regular progression where the distance between consecutive Cs is simply one octave, each time, to our ears. But in fact, where the ear perceives the following linear increments – 'one octave above C', 'two octaves above C', 'three octaves above C', 'four octaves above C', 'five octaves above C', etc. – the actual physical measurements are 261.62 Hz, 523.25 Hz, 1046.50 Hz, 2093 Hz, 4186 Hz, 8372 Hz, etc. . . . which means that where our brain tells us we have simply jumped one similar interval higher (an octave), the physical measurement tells us that we have doubled the previous number of frequencies each time. In our

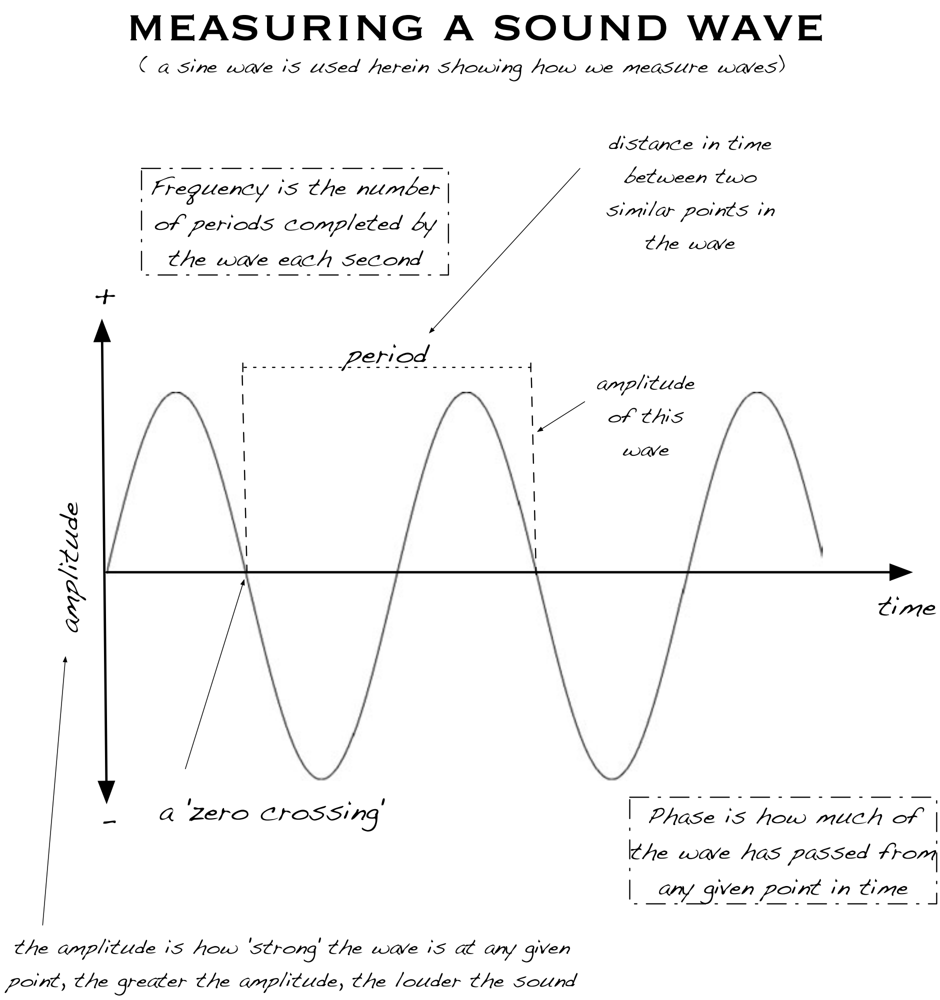

Figure 1.2 Measuring a sound wave

example, jumping up three octaves is actually multiplying the original frequency almost by 10! To cut a long story short, we will need to use *exponential* scales to measure sound. This doesn't only apply to pitch. It applies to amplitude as well, so that something which is twice as loud as something else to our ears is, in fact, ten units of amplitude larger; *loudness* is a physiological impression of how intense a sound is to our ears. Just imagine: if you have two guitarists playing, for you to perceive their amplitude as being twice as loud you need twenty guitarists! (there's a thought . . .)

So how can we represent a series of numbers that grow exponentially to be able to measure frequency and amplitude? We are going to need a mathematical concept:

the logarithm.[4] At this point you may want to make yourself a coffee because the following is a little involved. OK, here it goes . . .

If you take a series of numbers like:

1, 10, 100, 1000, 10000, 100000

you could find a different way to represent them by observing that they are powers of 10. So, as shorthand, you could represent them by writing down what power of 10 is needed to make each number. The sequence above then becomes:

0, 1, 2, 3, 4, 5

which is quite compact! And to get the original number you simply elevate 10 to the number found in that simpler and compact series:

$$10^0 = 1,\ 10^1 = 10,\ 10^2 = 100,\ 10^3 = 1000,\ 10^4 = 10000, \text{ etc.}$$

These exponents are logarithms in base 10:

$$\log(1) = 0,\ \log(10) = 1,\ \log(100) = 2,\ \log(1000) = 3, \text{ etc.}$$

In summary, we have a series of numbers which looks linear:

0, 1, 2, 3, 4, 5

actually representing an exponential one:

1, 10, 100, 1000, 10000, 100000

Very useful, don't you agree? Getting back to our measurements of sound, remember that amplitude grows exponentially, as we said earlier. And, our measurements of loudness start with very small numbers. For instance, we can hear a change in air pressure of 0.00002 Pascals, but as we consider other things we normally hear, we will see that the numbers grow quite rapidly. (By the way, a Pascal, or Pa, is a measure of force exerted over an area, so it can be used to measure the pressure of the atmosphere upon the earth, and sound results from changes in air pressure, so you get the idea!)

Your bedroom,[5] if it is reasonably quiet at night, will have average changes in air pressure of about 0.00063 Pa. If somebody speaks to you in a normal voice from one metre away, that exerts upon your ears around 0.02 Pa. If you stand about ten metres away from a diesel truck, the measurement would be about 0.63 Pa. One metre away from a chainsaw, 6.3 Pa. Sounds will be painfully loud at 63.2 Pa, and a jet plane, fifty metres away, will produce 200 Pa.[6] Let's put these numbers in a list:

0.00002, 0.00063, 0.02, 0.63, 6.3, 63.2, 200

They don't look very memorable, do they? So, if we apply some maths we mentioned earlier (and some we have not mentioned yet!) we can represent our loudness differences above like this:

0, 30, 60, 90, 110, 130, 140

which immediately looks simpler and more memorable.[7] So how did we arrive at the latter numbers? Sound-pressure levels and loudness are measured in a logarithmic scale with units called decibels. A decibel, loosely defined, is the logarithm of a ratio between a certain quantity and a reference level. If this confuses you, just think that we are measuring ratios, that is, relationships between two numbers, to see how much bigger or smaller one is than the other. Then think that the resulting ratios will grow very rapidly, so we need our trusty logarithms to represent those numbers more compactly. The first ratio should be 1, meaning that the numbers being compared are equal. Given our measurements of loudness, then we would have:

$$0.00002/0.00002 = 1 \quad \text{and} \quad \log(1) = 0$$

So, if two sounds have the same level, the difference between them is 0 decibels, or 0 dB. This is why in the last series of numbers we started with 0. All other changes will be expressed in reference to that one. 0 dB is, then, the reference level meaning that there is no difference between the two loudnesses being compared. To show that we are measuring sound-pressure levels we notate this as:

0 dB SPL

At this point, know that we can use the decibel for measuring any ratios, so if our decibels are used to measure power, say for amplification of the sound, then the reference would be to milliwatts and the unit would be dBm.

The formula for calculating decibels of sound-pressure level is:

$$LeveldB = 20\log\left(\frac{measured\ amplitude}{reference\ amplitude}\right)$$

So, now we can find where the other numbers in the simpler list above came from.

Your quiet bedroom (0.00063 Pa) in relation to the smallest change in sound-pressure level you can detect (0.00002 Pa) would be 30 dB, because:

$$LeveldB = 20\log\left(\frac{0.00063}{0.00002}\right) = 29.9662$$

which is near enough to 30 dB. The voice, one metre away (0.02 Pa), in relation to the smallest change in sound-pressure level you can detect (0.00002 Pa) would be 60 dB, because:

$$LeveldB = 20\log\left(\frac{0.02}{0.00002}\right) = 60$$

And now if you keep plugging the pressure levels of the diesel truck, the chainsaw, the painful sound and the jet plane, you will get the other values we arrived at earlier.

What does this mean for people working in a studio? For a start, mixer-channel strip faders should make a little more sense to you now. Notice that they have a marking for 0 dB right at the top of the fader. This means that you can set your recording reference level (0 dB) as the loudest that you can get without unwanted distortion of the audio signal. Meaning: nothing should go above the reference. Once you have established this through trial and error by getting the musicians to play as loud as they will play in the song (not as loud as they *can* play!), you then measure all other levels in relation to that as a reference, making sure they are below 0 dB and that all of them combined do not go beyond 0 dB.

The audio signal

At this point, let's briefly review the process of how you capture an audio signal. Earlier we said that sound is made of changes in air pressure that are perceived through our ears. To model that and be able to store sound, engineers have replicated that process by designing very sensitive membranes that simulate the way our eardrums work. Simply put, the membranes vibrate mechanically and are attached to *transducers* (devices to convert one form of energy into another) that create electric current at intensities that are higher or weaker depending on how strongly or weakly the membrane is affected by the changes in air pressure (see Figure 1.1).

Once sound has been converted to electrical impulses, we refer to it as an *audio signal*. A signal being simply that: something which represents something else, in this case audio. Note that the behaviour of the electrical current mimics the sound-pressure levels picked up by the microphone membrane (also known as a *diaphragm*). This is to say that there will be fluctuations in current for each fluctuation in air pressure; actually, this is what is known as an *analogue* process: one thing affects another so efficiently that the second one can be used to represent the first (chew on that for a moment, it is a simple but key statement). By the same logic, if we take an electrical current and make it stronger (i.e. amplify it), then feed it into loudspeakers, we will hear the sound again, but potentially louder. Before the digital era, the two main (though not exclusive) recording methods of the twentieth century consisted

Cambridge University Press
978-0-521-17042-0 - Cambridge Introductions to Music: Music Technology
Julio D'Escriván
Excerpt
More information

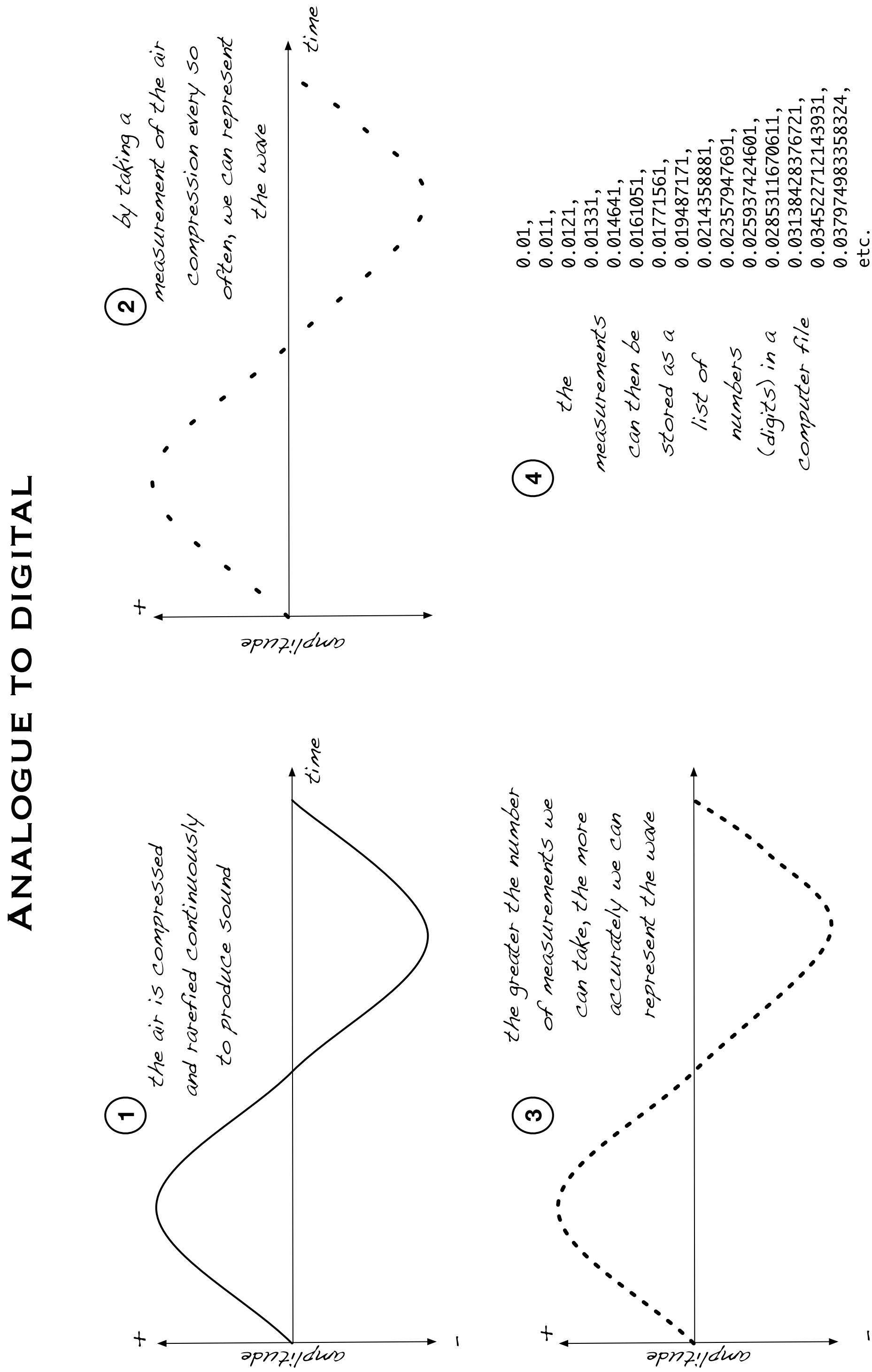

Figure 1.3 Analogue to digital

of preserving this electrical current as a groove on vinyl or as magnetised areas on tape with a magnetic covering. In the late twentieth century, digital recording became commonplace. The simplest explanation is that the electric current is now measured and the resulting numbers stored as a computer file.

Digital recording

Since it is the most common current method of recording, let's dwell on the digital recording process for a moment. I will just outline the key concepts.[8]

An audio signal needs to be *sampled* so we can get a list of values, where each one corresponds to one electrical fluctuation (which, as we saw, represents a fluctuation in air pressure). Think for a moment about this. You will see that the success of sampling will depend on two things: the number of measurements we can make per second (*sampling rate*) and how many numbers we can represent (more numbers → more accuracy → greater *sampling resolution*).

Computer technology allows us to do this and then use the data to make useful representations of the captured sound. Special software programs and protocols are needed for encoding and, later, decoding the data into audio formats; these are known as *codecs* and we will describe them later in this chapter. For now, just note that this method of capturing sound by measuring amplitudes at a particular constant speed (*quantisation*) and then representing those magnitudes in binary numbers is called *PCM*, or *Pulse Code Modulation*. This term comes from telephony as the technique was first developed for sending voice signals over the telephone. The word 'pulse' refers to each amplitude measurement at a constant speed which is made on the original sound; each quantised pulse is represented by a number belonging to a set of signals known as a *codeword* hence the use of the word 'code'; and, finally, the resulting transformation into binary numbers is a *modulation*: Pulse Code Modulation (for more precise information, you really need to learn about telecommunications!).

If you study Figure 1.3, you will better understand the key concepts. Notice that the fluctuation in current, represented by a simple waveform, has an unbroken curved shape. When sound occurs, our ears naturally perceive all the subtlety of this fluctuation. But if we want a computer to capture this phenomenon, we need to take as many measurements as possible per unit of time so we can approximate how the ear works. We will also need the computer to keep track of time, by use of a clock (the *sample clock*). This is called the 'analogue to digital conversion', and the electronic device used to achieve this is thus called the *ADC*. The list of measurements is kept as an *audio file*, which can be played back by performing a 'digital to analogue conversion' with the (similarly named) *DAC*. All we would need to do is, while keeping time (with the sample clock), output the number values so they can be

Table 1.1 *Commonly used audio sampling frequencies*

Sampling frequency in kHz	Commonly used for:
22.05	Internet podcasts, online streaming videos
44.1	CD recordings, music
48	Audio for video on DVD and digital video production
96 and 192	High Definition audio for digital audio workstations

converted into electrical current values and made strong enough (*amplified*) so they can be connected to a loudspeaker and drive it. The loudspeaker, roughly speaking, functions in an inverse way to the microphone. It takes an electromagnetically produced oscillation (fluctuation of values) and moves a membrane (speaker cone) in sync with it. The movement of the membrane creates differences in air pressure and the audio signal becomes audible again.

Sampling has evolved as computers have become more powerful, allowing us to capture sound with ever-increasing fidelity. There is a threshold level of sampling rate and resolution (or bit-depth) which today we consider a starting point for audio quality and which is used by commercial recordings on CD: this is the 44.1 kHz sampling rate at 16 bits. It is what people actually mean when they say 'CD quality' (see Table 1.1). Now, this is not the best quality possible, but it has become a standard, although it should be superseded thoroughly in the next few years. So, how did it come about?

First, let's make clear that the term 'sampling', refers not only to digital recording, but also to the artistic process of 'lifting' sound fragments from publicly available music tracks. However, both activities are accomplished with the same technology. The sampling rate that is considered acceptable nowadays is a result of research carried out by Dr Harry Nyquist, early in the twentieth century, whose concern with the optimum way to transmit an audio signal arose out of research into improving telegraphic communications.[9] Nyquist, together with Dr Claude Shannon, developed the theory of sampling at Bell Labs in the United States and their theorem is named after them (the Nyquist–Shannon sampling theorem). They discovered that, to reproduce a sampled signal without loss or distortion, it had to have been sampled at least twice as fast as the highest frequency it contained.[10] Since human hearing can perceive up to 20,000 Hz, then it follows that we need a sampling rate of at least 40,000 Hz to make sure we can later reproduce the signal without distortion or loss.

As somebody working in music technology, you will always be trying to get the best quality of audio that you can store in your chosen media. So, the next important item to consider is *bit-depth*, which is to say, how many values can you represent? I think it is safe to assume that everybody knows, at least informally, that computers