2. Physical sound

2.1 What is sound?

Sound is the human ear's perceived effect of pressure changes in the ambient air. Sound can be modeled as a function of time.

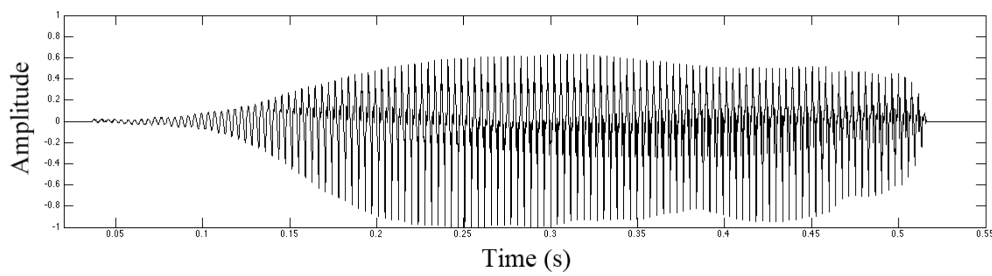
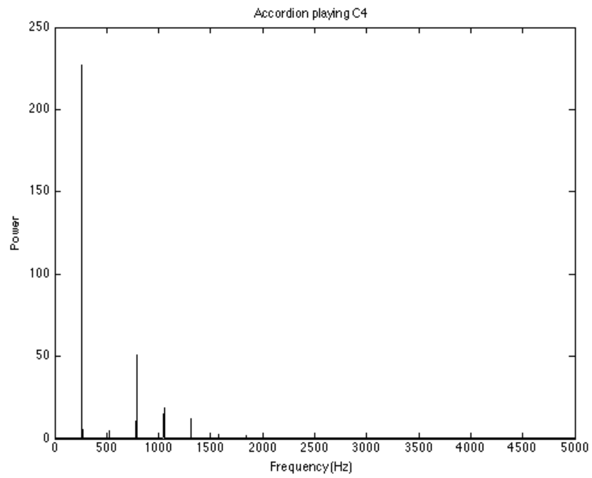


Figure 2.1: A 0.56-second audio clip of an accordion playing C4 (middle C, 261.6 Hz).

When we hear music, we can evaluate its features almost immediately. We can recognize the instrumentation, modality, artist, genre, and perhaps the time and place it was recorded. Graphically, it is difficult to connect this image to what we actually hear: The above graph looks complicated, while the experience of this sound (the *audio signal*) is a single, sustained pitch on an accordion. But when we take a Fourier transform of this clip, we can actually view the frequencies present in a song.

Because pitch and timbre are made up exclusively of the change over time of frequencies and amplitudes, and they tell us so much information about musical features, the Fourier transform is an incredibly useful tool that translate *time domain* signals like music onto an axis of frequencies, i.e., a *frequency domain*. The graph in Figure 2.2 lets our eyes verify what our ears already know: The graph describes the



Freq. (Hz)	Rel. Power	Closest Note
263.2	227.1	C4
528.2	4.461	C5
787.9	50.48	G5
1051	18.76	C6
1313	11.81	E6
1574	2.631	G6
1837	1.683	≈B♭6

Figure 2.2: The spectrum and listed frequencies attained by the discrete Fourier transform of the clip shown in Figure 2.1. Note the locations of the peaks with respect to frequency.

relative strength of the frequencies present in a signal. In this example, we see the frequency characteristics of an accordion playing C.

The frequencies themselves are not as important as the general shape of the spikes and the distance between them; most cannot distinguish between A and C in isolation, but we do have a relatively easy time identifying the difference between a piano and a violin. This is because of the texture of the instrument's sound, called the *timbre* or *tone color*. When the frequencies are more or less equally spaced from one another, we say that the timbre is *harmonic*, or that we have a *harmonic overtone series*. Explicitly, the Fourier transform of the signal in Figure 2.1 and its graphical representation in Figure 2.2 tell us the signal contains the frequencies 263.2 Hz (C), 528.2 Hz (C), 787.9 Hz (G), 1051 Hz (C), and 1313 Hz (E). The frequencies' respective peak in the graph indicates their loudness; hence, they are decreasing in power.

Middle C is 261.6 Hz, so apparently this accordion is slightly out of tune—but furthermore, its timbre is not perfectly harmonic: the difference between its overtones should be 263.2, but $528.2 - 263.2 = 265$, and $787.9 - 528.2 = 259.7$. There are several possible reasons for why these spikes are not exactly equally spaced. Most likely, it is due to the imperfect physical proportion and construction of the instrument's metal reeds, but it could also be error encountered in the recording process or experimental error.

To interpret how exactly this translates to what our ears hear, we must take into account how certain frequencies are perceived by the brain. Young, healthy human ears can detect frequencies within a range of 20-20,000 Hz, where 20 Hz and 20,000 Hz are threshold and limit values, but our ears are not uniformly sensitive to these frequencies [1]. Within the range of 1000 to 5000 Hz, our ears are especially sensitive, meaning that sounds with frequencies within this range do not have to be as loud for our ears to detect them.

Mathematically, the Fourier transform constructs an *orthonormal basis* that takes a complicated sound wave and reduces it to its component waves, which are all simple sine and cosine waves, or *sinusoids*.[1] It shows us every frequency and its amplitude that is present in a complex sound over an interval of time. The connection between the graph of the transform and its mathematical properties is a giant step towards realizing the Fourier transform and its digital applications. Because sight and sound retrieve giant spheres of information, we have to make decisions about what is important and what we can take for granted. Our brains are so excellent at processing information that we can give certain sensations finer resolution (like an important

[1]"Orthonormal" means orthogonal and normal. For a function to be orthogonal to another function, the two functions must be *linearly independent*. The condition of normality is satisfied when each function involved has, in some appropriate sense, energy 1. Finally, a *basis* is a set of functions such that an arbitrary function (within reason) may be written in terms of the basis. See Chapter 7.

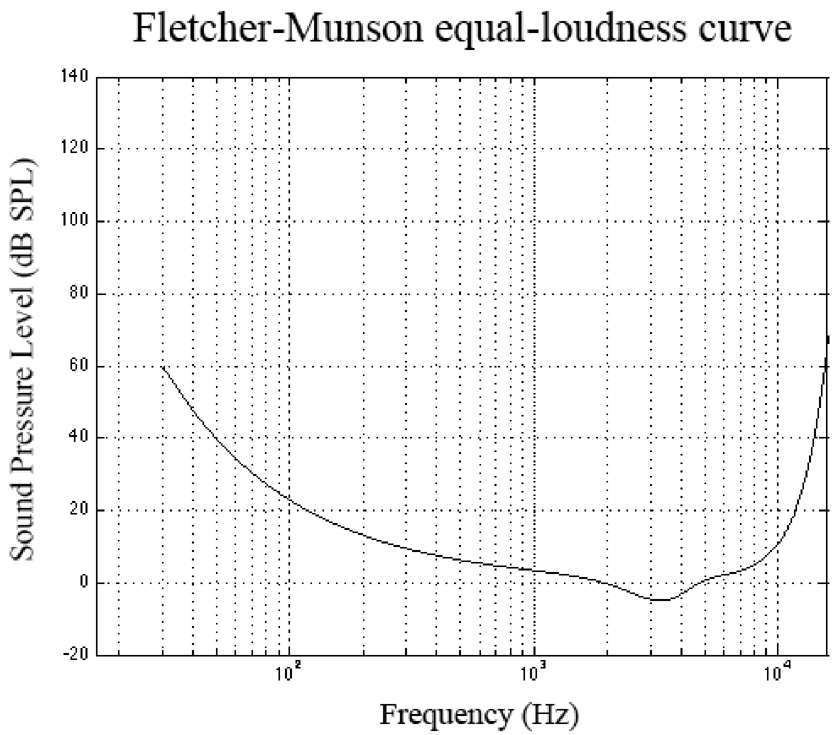


Figure 2.3: We detect frequencies between about 20 and 20,000 Hz as pitched sound. Furthermore, each of these frequencies has a minimum threshold of loudness. This graph, the Fletcher–Munson curve, shows the minimal sound pressure level in decibels (dB) required for the frequency to be heard.

message from our friend), and others none at all (like the hum of the refrigerator).

Consider seeing a relatively involved movie for the second or third time, and noticing things you didn't notice before that now make sense. We seem to prefer movies like these. We may achieve a decent understanding of the plot on the first viewing because we extract salient parts of the dialogue and action and put them in order, but a complex plot can hide clues of outcomes and their rationale all over the film that are more obvious when our brains can support them with familiar elements.

A complex piece of music can be a lot like a complex movie. We perceive both sound and light as signals. When a signal demands our attention, it is said to have a high amount of *information*. A signal with meaningless content that we don't need or want to listen to is called *noise*. ound produced by white noise machines, e.g., is random, unpitched, and trivial. It does not contain a message because it is formally disorganized, and it can even help some people sleep because of its uniform randomness. Noise is composed of so many periodic waves that we consider it aperiodic. We cannot extract individual frequencies of noise, as we can in melody or a major chord in music. A signal can be half meaningful and half noise, and our brains are powerful enough to recognize the difference and attempt to separate the two.

Although sine waves are not fun to think about, they substantiate much of the mathematics and physics behind music. The mathematical and physical equations which produced the previous graphs form a basis for the sensation of sound. Many musical concepts are results of mathematical relationships. First, let us examine the basic mathematical structure of sound. Musical form will be addressed in Chapters 3 and 4.

2.2 Simple harmonic motion

Like light, sound is traveling energy, and we can model such energy mathematically with waves. The simplest wave is a *sinusoid*, a trigonometric function such as $\sin(\omega t)$ or $\cos(\omega t)$ where t denotes time and ω specifies how often it repeats itself—its *angular frequency*.[2] A sinusoidal wave represents the *simple harmonic motion* of an object because its frequency and extreme magnitudes do not change over time.

Both a spring and a tuning fork exhibit simple harmonic motion. Below, we see two states of a vibrating tuning fork called *modes of vibration*. Both of these modes produce sound that is near in tone to a sine wave (or *pure tone*), but as you might have experienced, the tone

[2]The *frequency* f in Hz. (cycles per second) is related by $\omega = 2\pi f$.