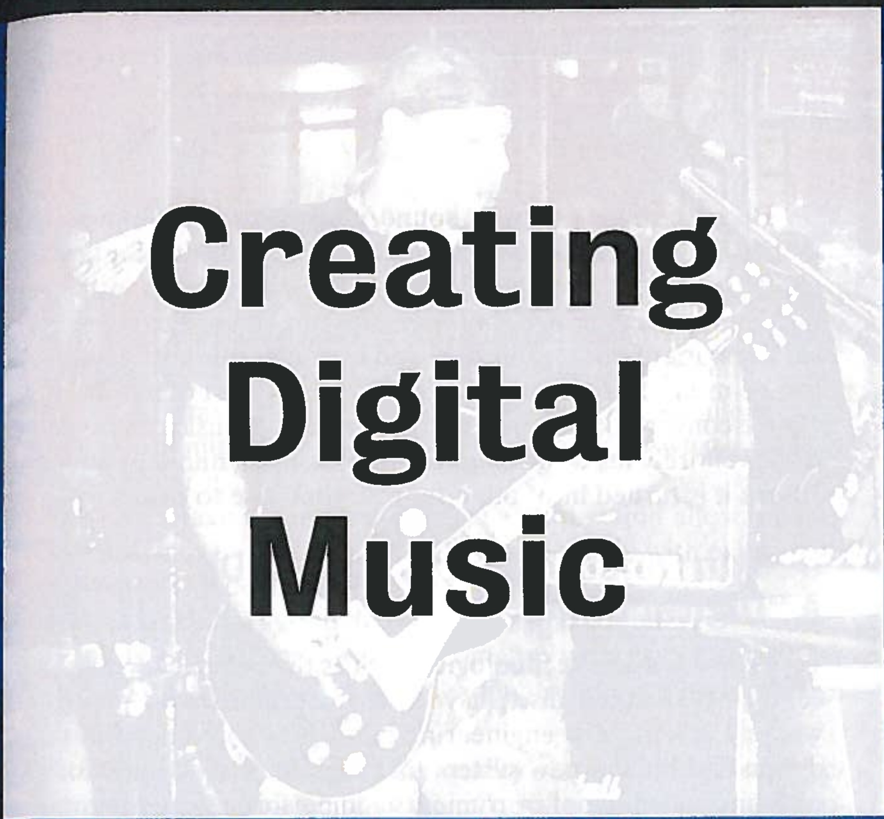


Creating Digital Music

chapter 2

Technology has changed nearly all aspects of the arts, especially the art of music. Nearly everyone listens to vocal and instrumental music on the radio, at concerts, or as part of movies and theater performances. Technology has made dramatic changes in our ability to create and enjoy a wide selection of music. Modern recording and communication methods make it possible for us to hear music performed long ago or in far-off places, using relatively inexpensive and widely available devices, such as DVD players and personal audio players.

Artists who compose and perform music now can use digital technology to enhance the sound they create by adding instruments and sound effects or even correcting performance errors. The sound of a new instrument can be created without actually constructing a new physical instrument. "Electronic compositions" are often used for the background audio of movie soundtracks. These digital techniques help make the pop stars of today sound as good as they do. Who knows—we may not be too far away from creating a "virtual singer" whose entire band shares her or his space inside of a digital device.

OUTLINE

- Introduction
- Music, Sound, and Signals
- Making Music from Sines and Cosines
- Improving the Design—Making Different Instruments

Because of the nature of sound, music is tied very closely to mathematics. We can use mathematics to create sounds that imitate traditional instruments or to create completely new sounds for instruments that have never even existed. Mathematics can be used to mimic the physical behavior of both traditional and new instruments. Mathematics also can be used to store music in digital form as a list of numbers that later can be converted back into music. A piece of music can be created and stored entirely inside a computing device, using numbers and equations, before it is turned into sound for everyone else to hear.

INTERESTING FACT:

In 1957, Max Mathews from Bell Laboratories created the first music synthesis program. An improved system named GROOVE was designed in 1970 to allow music to be created from a simple description so that the user could influence the sound as it was created. The electronics for this system was so large it would fill a walk-in closet.

Design Objective for Creating Digital Music: A "Digital Band"

Previous audio technologies, such as the compact disc and, more recently, DVD audio discs, have focused primarily on playing existing works. Let's use our engineering know-how to go beyond this task to design and build a new system that can create new music of any style, with any collection of instruments and performers at any time we want to hear it—our own "digital band."

In this book, we use the engineering design process or algorithm to help us understand and solve our problems. This algorithm begins with a series of questions to help us organize things:

- **What problem are we trying to solve?** It is always important to clearly define the problem we want to solve so that we use our time and resources effectively. In this chapter, we want to design a new device—a digital band—that will allow us to create a wide range of music without requiring us to have either extensive music training or a complete music library.
- **How do we formulate the underlying engineering design problem?** All design problems include a set of *specifications*, or features, that describe what we want our final design to do. Different approaches to the design are evaluated by how well they meet the specifications. Some of the most important capabilities of our digital band include the following:
  1. It should be able to re-create any combination of instrument sounds desired. For example, it should allow us to select known instruments or even to create the sound of new ones.
  2. It should be able to combine newly created sounds with existing recordings or modify previously recorded sounds to change the sounds' style or the type of instruments being played.
  3. It should be able to read some form of notation or score and convert performances into notation by itself without any special training.
  4. It should be able to imitate the environment in which we would like to hear the music performed so that it can sound as if the music comes from a concert hall, a recording studio, or even our echo-filled bathroom.
  5. It should be able to create music for a long time, so that we can hear a complete concert or listen to hours of background music.
  6. It should allow us to share the music we create easily with others or to take music our friends have created and change it to make it sound better to us.
  7. It should be small and lightweight so that we can take it wherever we go.

8. It should be durable and built to last a long time.
9. It should be as inexpensive and as easy to make as possible.

What will we achieve if our design meets our goals? If we do a good job in the design of our digital band, we would achieve a number of important benefits for ourselves and our friends:

1. We would have a very flexible device that would allow more people to enjoy their own music whenever and wherever they want to hear it.
2. We would have a new creative device that would allow practicing musicians to make music that is not possible with existing instruments and devices.
3. We could sell the system to others and make money.
4. By using this system, we could come up with newer and better ways to make music—perhaps leading to an even better design for the system.

How will we test our design? The quality of sound is always subjective, and the music from our digital band must sound good to the people who will use it. In order to achieve this goal, however, the design must be tested and improved at many stages of the system's development. Specific performance measures define the device's capabilities so that users will know what they can expect to do with it, and these specifications will be tested to make sure our system will function as advertised. Some of these tests might address the following inquiries:

1. What is the range of notes that the system can create? How is this range of sounds related to what we humans can hear?
2. How many different types of instruments or voices can be made? How many can be combined at one time?
3. How long can new compositions or combined recordings be?
4. How large is the device? How heavy is it? How much power does it need in order to operate? Is it a portable device, or does it need to be installed in a permanent location?
5. What devices will be able to reproduce or play the music created by the system? If the devices don't already exist, how easy will it be to make them?
6. How do listeners react to the quality of music from our system as opposed to music made by instruments that already exist? Do they enjoy what they hear when listening to our system?

2.1 Introduction

First Steps

When designing and building any device for the first time, we must make choices about the components, materials, and methods that we will use. What materials should we use to create our digital band? And what basic technology and components should it use to make music?

We begin our design task by drawing a block diagram of the system. As we learned in Chapter 1, each block has inputs, outputs, and a well-defined task. Once we have set the big picture in place with our high-level block diagram, we can then divide each of these major blocks into

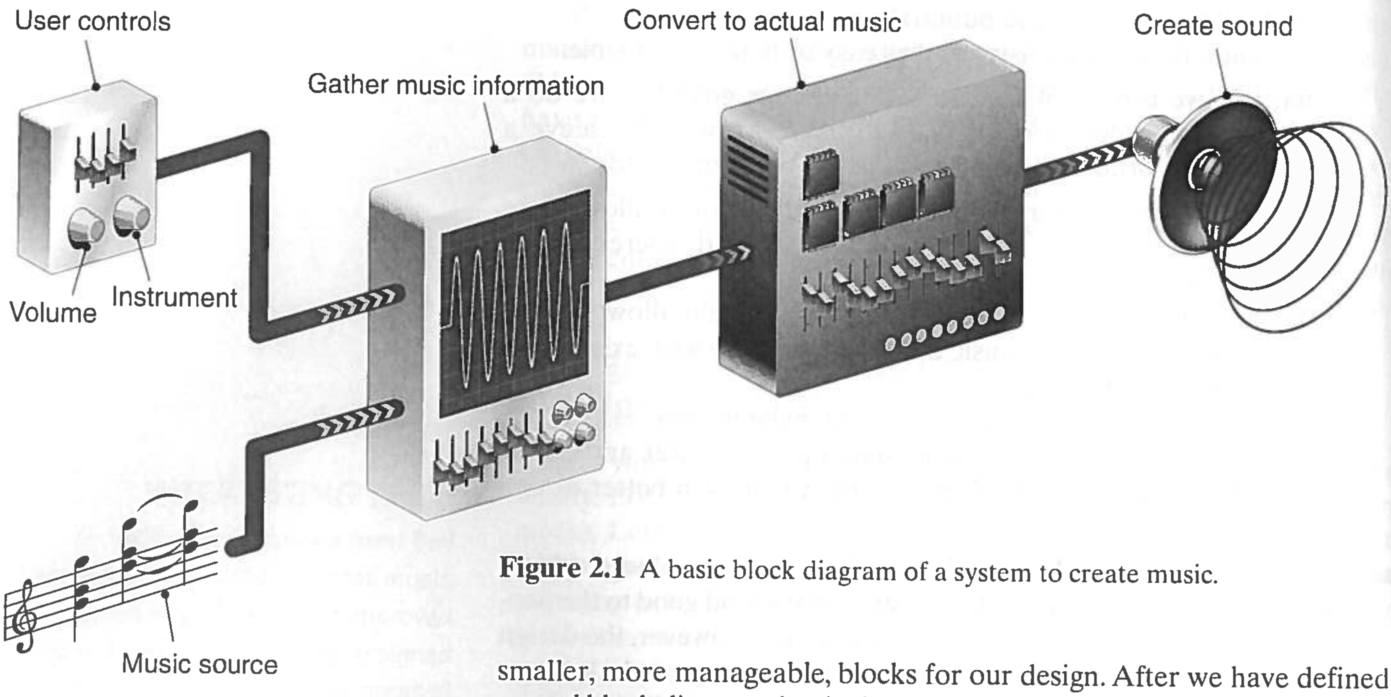


Figure 2.1 A basic block diagram of a system to create music.

smaller, more manageable, blocks for our design. After we have defined a good block diagram that includes all of our device's features, we can decide what technology to use to build each of the blocks. The choice of technology normally is based on what is currently available or expected in the near future and our estimate of the cost for the chosen technology.

A basic block diagram for our system is shown in Figure 2.1. The first block, labeled "Gather music information," has two general types of inputs. One input is the music source that describes the music we want to hear. Normally, this source would take the form of sheet music used by musicians to describe a piece of music. We will call the second type of input "user controls." These controls allow you to make changes while you are listening to the sounds of your device, such as by selecting the music from the source or adjusting the loudness with a volume control. The second block in Figure 2.1, labeled "Convert to actual music," takes the description of the music and the user controls and processes them to make an electrical form of the music before it is converted into sound waves. The third block, labeled "Create sound," translates the electrical form of the music into sound waves so that people can hear the music.

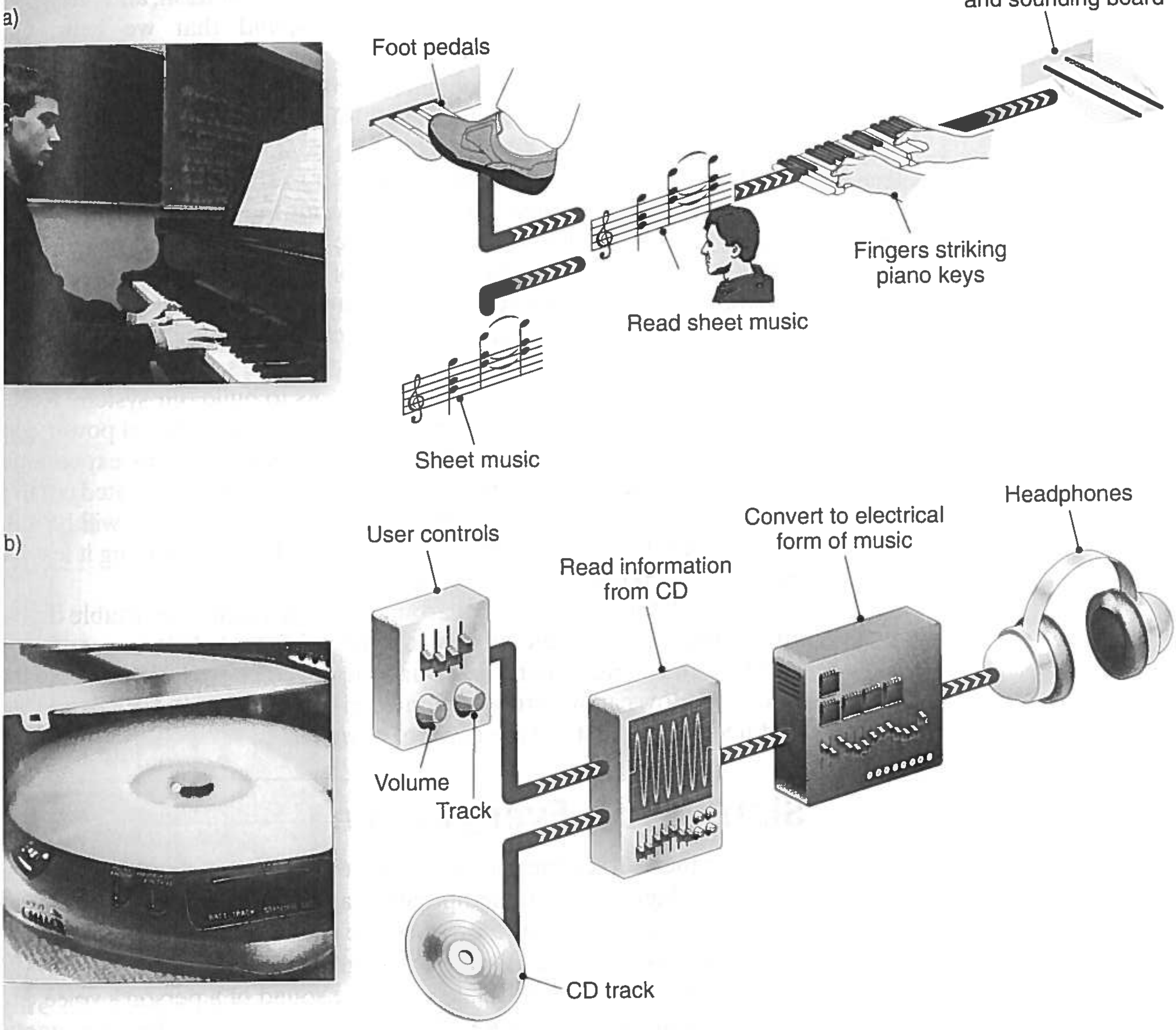
This block diagram is quite general. It can describe both simple and sophisticated music-generating systems. The details of the design of the blocks and the implementation technology will determine the capabilities of the system, its flexibility, its cost, and how well it meets our described design goals.

Engineers are always borrowing from the past and building on existing techniques and methods. They add new features and take advantage of current technology to create newer and better products. Before beginning a design, it is wise to review the development of the related products in order to understand what can be reused and to see where improvements are needed. Today, most of us listen to music from CDs or DVDs played on portable or home music systems. Because engineers are so effective at using what they already know, you might not find it surprising that CD and DVD players spin their discs in a way that is not much different from the method used by the vinyl record players that preceded them.

Vays to Make Music

gure 2.2(a) shows a person, a pianist, making music on a piano. The pi-
nist reads the notes to be played from sheet music that contains the in-
ructions for playing a song. The conversion of the sheet music's
scription into sound is done by the fingers of the pianist striking the
ano keys. The pianist can control the quality of the music by changing
e way in which the keys are struck and by changing the position of the
dals near the base of the piano. Finally, the vibrating strings and sound-
g board of the piano make the sounds we hear. So, a person playing a
ano is an example of our general block diagram of Figure 2.1.

Music can be played automatically using mechanical methods to
escribe and create the music. For example, mechanical windup musi-
l toys use a small cylinder with spokes inside to play a short melody
nce the toy is wound up. The way these windup toys work is similar to
at of the old-style music boxes that were first designed by Swiss
atchmakers in the late 1700s. The player pianos of the late 1800s and
rly 1900s worked on similar concepts. (See Figure 2.3.)



igure 2.2 Two systems for creating music and their corresponding block diagrams.
a) A person playing a piano while reading sheet music. (b) A portable CD player.

Figure 2.3 Player piano.

INTERESTING FACT:

Before Edison's phonograph, the only way to hear music was through live performances. The phonograph, and the CDs and DVDs that followed later, enabled many more people to hear high-quality music easily and inexpensively.

Another way to enjoy music is by listening to a past performance that has been recorded. The first instrument to reliably record musical sound was the phonograph, invented by the American inventor and engineer Thomas Edison. Edison's phonograph was a mechanical device that was later improved by adding an electronic amplifier. It brought much longer and more complex musical performances to people who rarely had the opportunity to listen to the great singers and symphonies of the world. Engineers working in the 1970s further revolutionized the field of recorded music by developing the compact disc (CD) and other digital technologies. This invention removed most of the mechanical components of music reproduction and dramatically improved the quality of recorded music. Improvements in manufacturing methods allowed music distribution companies to make CDs quickly and inexpensively, and the size of the discs is small enough so that they may be carried around in a backpack. All of these developments resulted in a much wider availability of systems to play CDs, along with a much wider selection of recorded styles and performers.

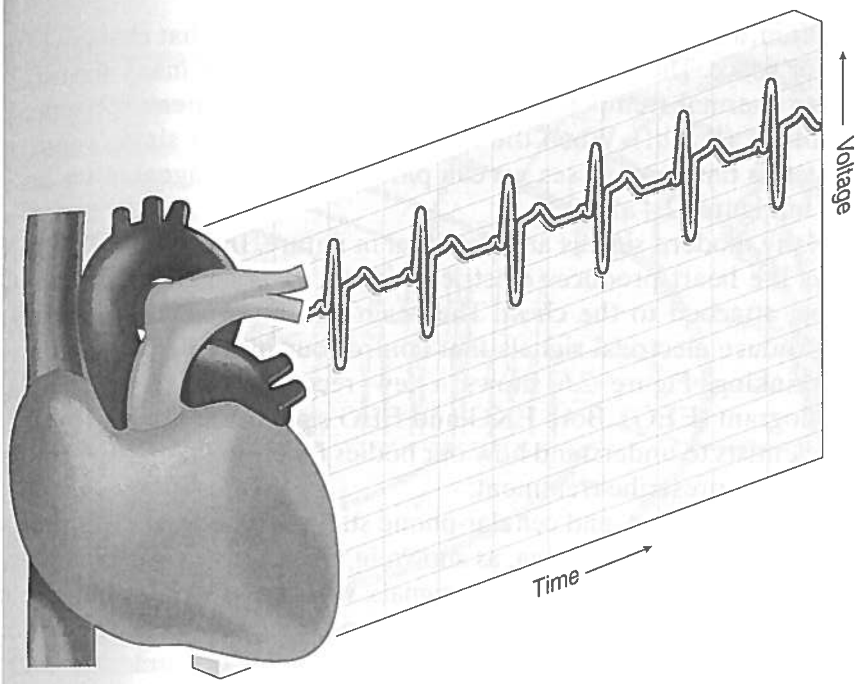
Figure 2.2(b) shows a portable CD player and its associated block diagram. In this system, the CD provides the music description, the CD player converts the music description into electrical form, and the speakers inside the headphones create the sound that we hear. Our digital-band design should do more than simply replay a piece of music; we want to be able to make significant changes to recorded music or to create completely new music. CD and DVD technologies can be part of our system, but we will also have to add new features that allow us to tell the device what we would like to hear. For example, our source may be an existing recorded piece to which we want our system to add a saxophone or for which we want to change the style by making it go faster. Or our source might be an original score for which we want our system to arrange a new or unique set of instruments to perform the piece.

As in every design task, we have choices as to how we implement our digital-band design. As we have learned, we have some good reasons to use electronic components such as transistors to build our system. Recall from Chapter 1 that the transistor density and computational power continue to increase according to Moore's law. Because of this exponential growth, if we pick an implementation method that uses integrated circuits, we can be sure that as technology progresses in the future, it will be easy to make our digital band better by adding new features, making it less expensive, or making it more portable.

From all the digital devices around us—for example, portable digital audio players, electronic keyboards, and miniature digital recorders—we know that sound can be made and improved upon digitally. But what is sound? How can we process sound digitally? To understand sound, we first need to understand the concept of *signals*.

Signals Are Everywhere

What do the voltage measurements produced by an electrocardiogram (EKG; see Figure 2.4), the temperature at the Dallas–Fort Worth International Airport (see Figure 2.5), and the sound of a guitar have in common? The answer is that they are all signals! Some signals, such as the chemical concentration within a cell or the sound of a person's voice, are created inside of our bodies. Other signals, such as the electromagnetic waves produced by a radio transmitter or cellular telephone, the sound

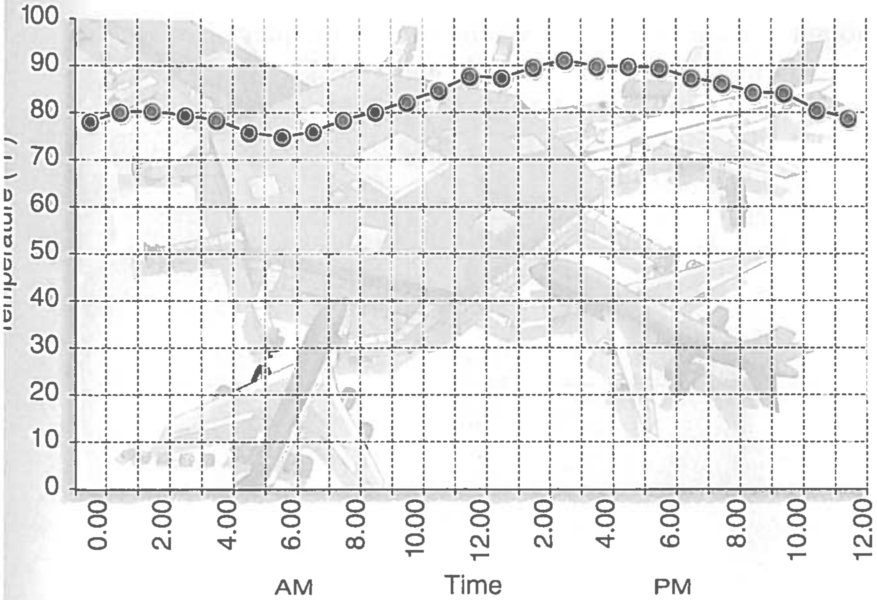


igure 2.4 An EKG of a human heart.

f thunder, the vibration of an earthquake, or the light from the sun, are reated in our surrounding environment.

What exactly is a signal? A **signal** is a pattern of variation over time hat contains information. To understand this concept, let's return to he example of the temperature at Dallas–Fort Worth International Airport. What information can we get from this signal? You can see rom the plot in Figure 2.5 that the temperature does in fact vary at the irport over the course of a day. The coolest time of day is 6:00 AM, and he hottest time of day is 3:00 PM.

Signal: A pattern or variation that contains information, usually denoted as $s(t)$.



igure 2.5 Temperatures recorded over a 24-hour period at the Dallas–Fort Vorth International Airport. The range of the temperatures, the difference etween the lowest and the highest, is less than 20°F.

Often, a signal represents some physical quantity that changes with time or space. The physical quantity can be in one of many forms; including thermal (temperature), electrical (voltage), pressure (sound), or altitude (height). When the quantity depends on a single value or changes as time progresses, we can plot or graph the signal as we have done in Figures 2.4 and 2.5.

Many modern signals are electrical in nature. In Figure 2.4, for example, the heart produces electrical signals that can be measured by sensors attached to the chest. The brain and human nervous system also produce electrical signals that control our movement, sensations, and thinking. Figure 2.6 shows a few recordings of an electroencephalogram (EEG). Both EKG and EEG signals are used by doctors and scientists to understand how our bodies function, to diagnose problems, and to prescribe treatment.

Radio, television, and cellular-phone stations broadcast electromagnetic signals from an antenna, as shown in Figure 2.7 for a cellular base station antenna and signal. These signals, which represent variations in voltage at the antenna over time, can be captured by a receiving antenna and then converted back to sound or images for our entertainment. These same signals also can be viewed directly by technicians or engineers on instruments that graph the signals as a function of time.

Music is also a signal. By capturing music with a microphone, music can be placed in an electronic form just like an EKG or a radio signal, and its variations can be plotted as a graph as in the previous examples. Like any graph, the signal can also be saved as a list of numbers and re-created later as was the case in Figure 2.4. In Figure 2.8, we see a group of musicians making music sounds. Each instrument is creating its own characteristic signal. These signals are sounds that combine in the air so that we hear them all at once. When each of these signals is converted to a list of numbers, we can use modern digital technology to re-create the sound. More importantly, we can also make changes to the numbers to create new sounds. Or, if we are clever enough, we can use basic mathematical techniques to create lists of numbers directly and convert them to musical sounds. We will explore these possibilities in detail in later sections of this chapter.

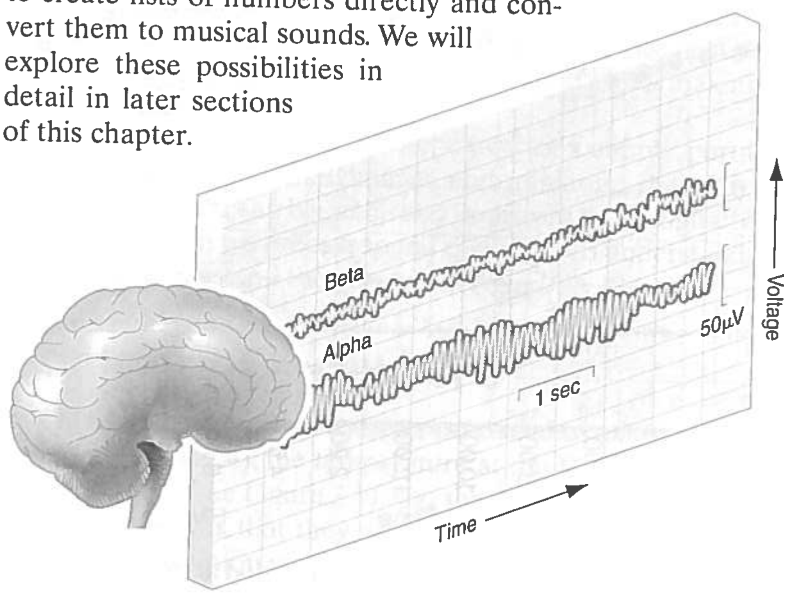


Figure 2.6 EEG signals from the human brain.

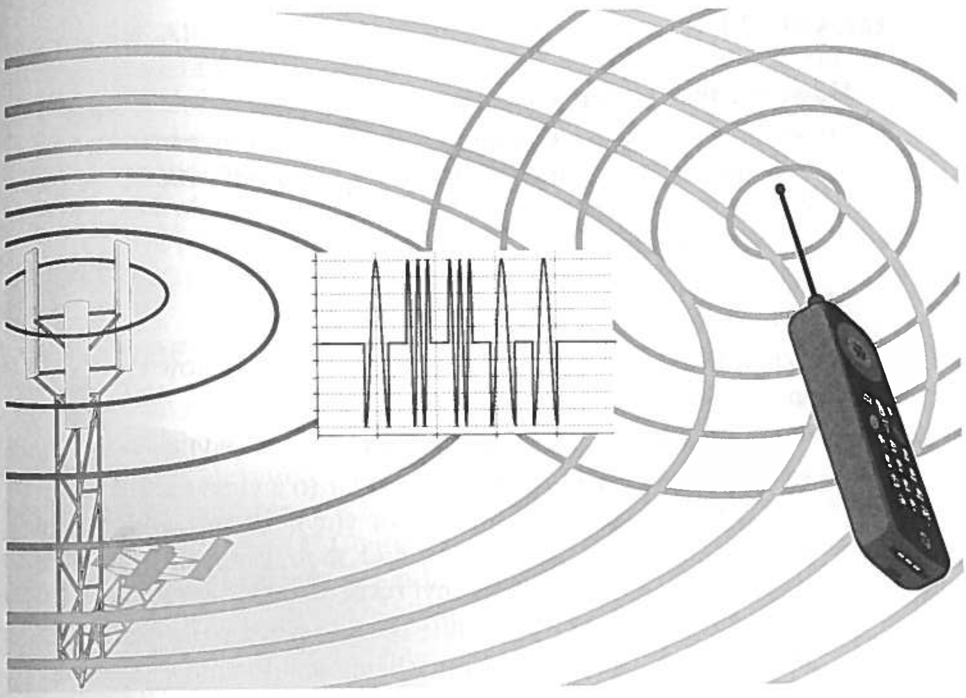


Figure 2.7 Cellular telephone signal.

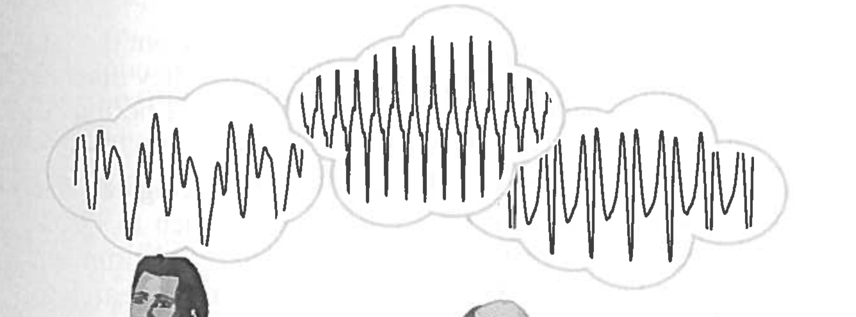


Figure 2.8 A group of musicians and the signals they create.

EXERCISES 2.1

Mastering the Concepts

1. When formulating a solution to an engineering problem, how are specifications used? Write a set of specifications for the following designs:
   a. a step stool
   b. a personal vehicle to go to and from school
   c. a writing instrument
2. When first formulating an engineering design, how can existing designs help the designer?
3. What is the generic block diagram for a device that makes music?
4. Describe how a CD or DVD is similar to a vinyl record. What ideas did engineers borrow from vinyl records and record players to make CDs, DVDs, and their players? What are some differences between vinyl records, on the one hand, and CDs and DVDs, on the other?
5. What is a signal? What is a mathematical function? How are they similar? How are they different?

Back of the Envelope

6. Plot examples of five different signals that come from the real world. These signals can be from anywhere, but they must be able to be represented by a plot. Be sure to label both axes. How are these signals used in a particular application?
7. Find pictures of systems or devices that process signals. The sources of the pictures can be printed media such as magazines, newspapers, or catalogs, or you can find and print out pictures from the Web. For each device, do some research to find out the following:
   a. what technology the device uses
   b. whether the device contains or processes digital representations of signals

2.2 Music, Sound, and Signals

What Is Sound?

Before designing a new product or system, an engineer must understand how people will use it and understand the related science and technology that can be used in the design process. So, before we can make music, we have to understand what makes sound and how we hear sound.

We can't see sound, but we can learn about sound through a visual analogy. We are all familiar with waves on the surface of water, such as those that move across a lake or crash on a beach. These waves are caused by wind or some other disturbance. Like waves on water, sound is a wave that travels through a physical medium, such as the air. Sound waves can travel only when there is some physical material through which they may be carried, so they can't travel in the near-vacuum of

INTERESTING FACT:

Sound causes air molecules to move back and forth. Wind causes air molecules to move forward in only one direction.

outer space. Although sound waves can travel in gasses, liquids, and solids, because of our interest in music we'll focus on sound waves in air.

Sound in air is created when a small disturbance causes the air molecules to move back and forth quickly. These disturbances come from mechanical vibrations such as the motion of our vocal cords when we talk or the vibrations of a plucked guitar string. These small "ripples" of motion actually move from one place to the next in a coordinated pattern called a "traveling wave," similar to the ocean waves we see at the beach.

INTERESTING FACT:

Two astronauts cannot talk to each other through the vacuum of outer space without some assistance. They use wireless radios to convert their voices to electromagnetic waves and back to sound again.

The Speed of Sound

Anything that moves has a number or value for its speed, and sound is no different. The speed of sound in air depends on several things, such as temperature, elevation, and humidity. The speed of sound at sea level in the Earth's atmosphere is about 340.4 meters per second (m/s) when the air temperature is 15° Centigrade (C). We can convert this number to miles per hour (miles/hour) as follows:

$$(340.4\ \text{m/s}) \times (60\ \text{s/min}) \times (60\ \text{min/hour}) \times (1\ \text{mile}/1609\ \text{m}) = 761.6\ \text{miles/hour}$$

This speed is faster than most normal passenger jets fly.

Since 761.6 miles per hour is faster than most of us have ever traveled, let's see if we can put this speed into a more understandable context. If sound travels 761.6 miles every hour, then it will take 1/761.6 hours, or 0.001313 hour, for sound to travel 1 mile. We can convert this number to seconds per mile as follows:

$$(0.001313\ \text{hour/mile}) \times (60\ \text{min/hour}) \times (60\ \text{s/min}) = 4.727\ \text{s/mile}$$

So, it takes sound about 5 s for sound to travel 1 mile. It takes us about 65 s to travel 1 mile in a car at highway speeds. However, the speed of sound is very slow when compared with the speed of light and electricity, which is about 186,000 miles per second! Light and electricity take only 0.0000054 s = 5.4 μs to travel 1 mile—a time frame that is almost instantaneous compared with that of sound. The difference between the speeds of sound and light can be used to figure out how far away lightning is in an approaching thunder storm.

Figure 2.9 A group of fans doing "the wave."

INTERESTING FACT:

The way sound moves is similar to the way that "the wave" in Figure 2.9 is created by people as it moves through a stadium at an athletic event. In this analogy, the people are likened to air molecules, and their movements are likened to the sound waves moving through the air.

Sounds and Signals

If we had the ability to view air molecules and were to put ourselves in the middle of a sound wave, we would see that the density of the air molecules surrounding us would be changing all the time. We call this pattern of variation a **sound signal**. It is a signal that our ears or a microphone respond to and is different for each type of sound that we hear. This variation in density or air pressure causes our eardrums to move with the same pattern of variation. We perceive this movement as sound because our inner ears translate these small movements of the eardrum into nerve impulses that our brain interprets as sound.

Sound Signal: A pattern or variation in the motion of air molecules that a sound makes.

What is so amazing is that the sound waves from many voices or musical instruments combine at each of your ears into a single sound wave. This fact means that we can use simple devices to sense and replay such signals. A **microphone** is a device much like our eardrums in that it converts sound energy into electrical energy allowing the energy

Microphone: A device that converts sound energy into electrical energy.

Infinity Project Experiment: Plots of Speech

Like music, speech can be represented by a signal that can be plotted easily on a computer screen. Examine the shape of your own voice as you talk into a microphone. How does it look over several seconds? What happens if you zoom in on your speech signal over a fraction of a second? Try making different speech sounds by singing into the microphone or by talking in a different language, if you can, and look at the resulting signals. How do the signals change? How does the signal change if you whistle into the microphone?

Loudspeaker: A device that turns electrical energy into sound energy.

to be stored, changed, or displayed inside of an electronic device. Similarly, a **loudspeaker** is a device that converts electrical energy into sound energy so that it can be heard.

If we were to plot a sound signal, it might look like the plot shown in Figure 2.10, which depicts the signal picked up by a microphone when someone was saying "Hello, my name is. . . ." This plot of the sound signal mathematically represents what our ears respond to when we hear the sound of this particular person's voice. Mathematicians would call these signals *functions*, but we'll use the name "signal" because it is more common in engineering.

INTERESTING FACT:

The next time you are in a thunderstorm, watch for any bright flashes of lightning. As soon as you see one, count the number of seconds before you hear the sharp clap of thunder. By dividing the number of seconds you have counted by five, you can roughly calculate how many miles away the lightning is from you. Try it!

Most music is a combination of sounds created by several musicians playing instruments or singing together, as shown in Figure 2.8. Each instrument sounds different because the sound signal each creates is different. The characteristics of the sound we hear from a single instrument will depend on what instrument is used, what notes are being played, and the capabilities of the instrument player.

Figure 2.11 illustrates three sound sources—a tuning fork, a guitar, and a bird. Part of the signal recorded by a microphone for each is shown on the right. The signal produced by each source is different because the sound that each makes is different.

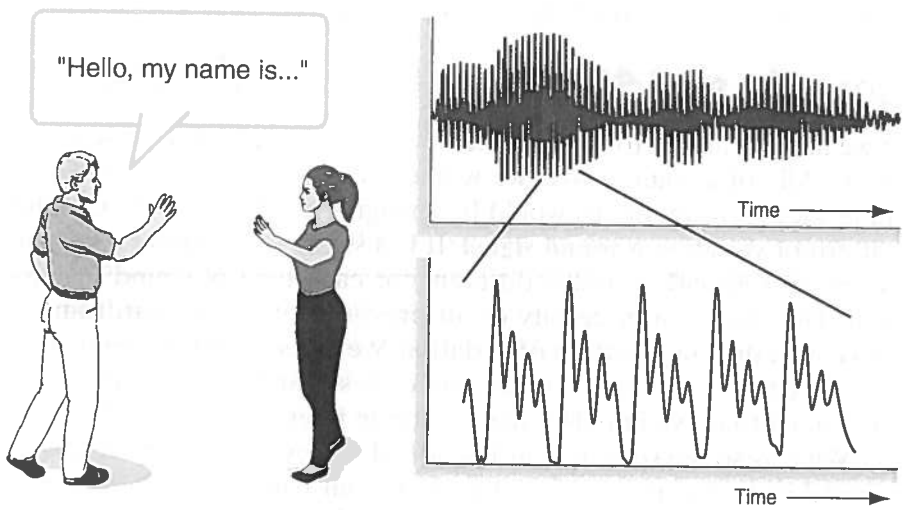


Figure 2.10 Person saying "Hello, my name is . . . ," and sound signals of the person's voice plotted on two different scales. The upper plot shows the whole signal, while the lower plot shows the time interval highlighted in red.

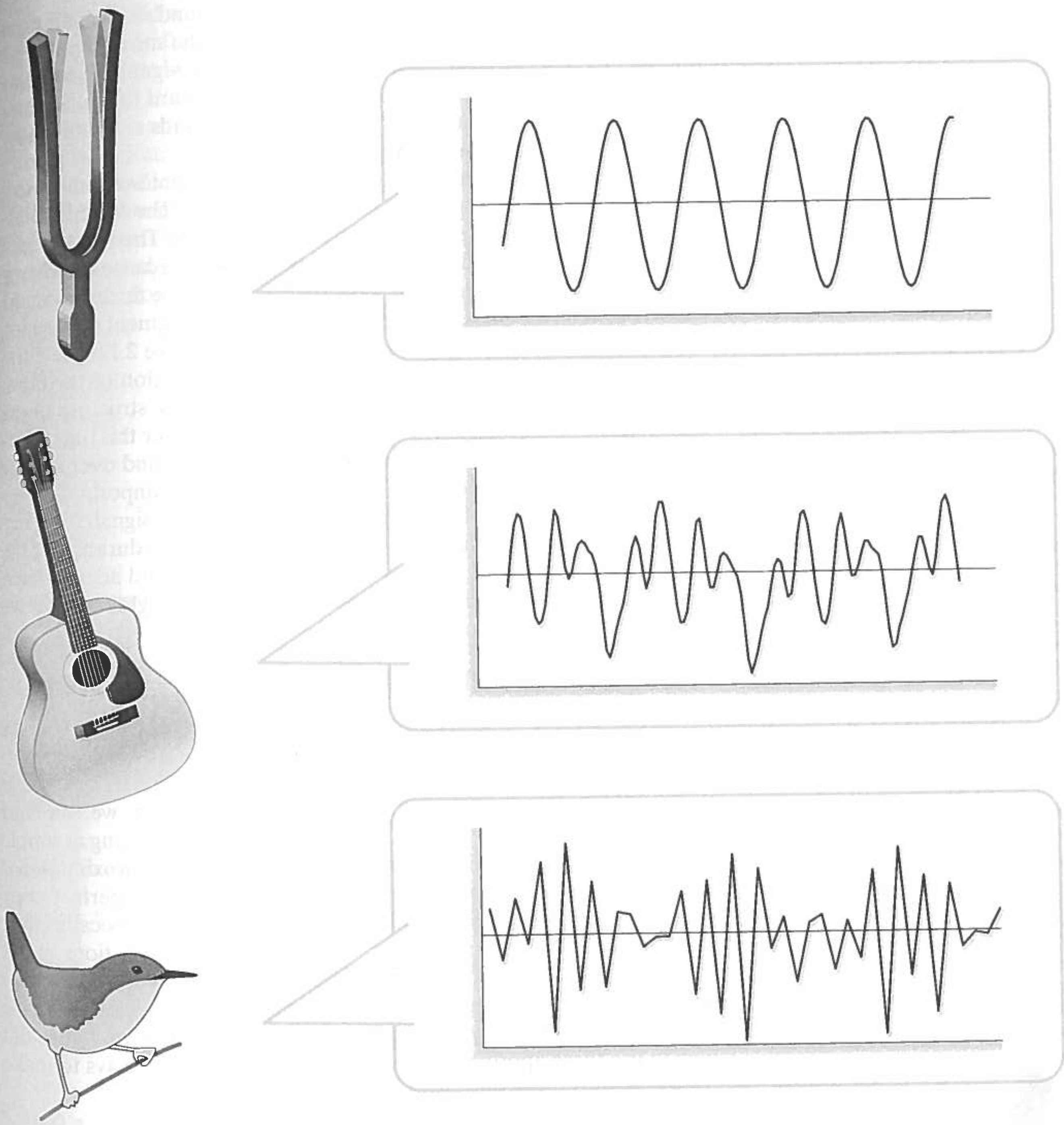


Figure 2.11 Three sound sources—a tuning fork, a guitar, and a bird—and the sound signals they create.

The design objective we've chosen for this chapter is to make a device that can create combinations of sounds from several instruments. Since sounds from instruments combine in the air, we can first design and test a device to re-create the sound of a single instrument. If we can figure out how to make the signal associated with any particular sound, such as those shown in Figures 2.8, 2.10, and 2.11, then we only have to convert those sound signals into vibrations in the air in order to re-create the sound itself. Once we have solved this problem for one instrument, we can explore methods for creating sounds from several instruments at the same time.

If we are designing our system to create the sound of an instrument that already exists, the quality and enjoyment of the music it produces will depend on how closely our signal matches the signal produced by the original instrument. For this reason, it is important for us to better understand musical signals and how instrument sounds are both similar to and different from one another.

Figure 2.12 shows an example of a single instrument's sound—in this case, the sound of a single note played on a guitar. The left-hand side shows the sound over a duration of several seconds. The sound signal looks like a solid ink blob. The signal looks this way because it is changing so quickly that the plotted lines of each variation are thicker than the spaces between the lines. If we zoom in on the small segment of this plot, we see the signal shown on the right-hand side of Figure 2.12. The duration of this signal is several milliseconds, a small fraction of the duration of the signal on the left. Here, we notice the basic structure of the sound signal. It has a distinctive pattern of variation over this time interval. Moreover, the signal appears to repeat itself over and over as time goes on. From an engineering perspective, this is a very important observation, because it simplifies our job of making a similar signal. If we can re-create this signal at this level of detail over the entire duration of the guitar sound and play it through a loudspeaker, we would actually hear the guitar sound. This fact gives us our first important insight into how we will go about designing our digital band.

Using Mathematics to Create a Signal

In this chapter, the signals shown so far for instruments and other sources are complicated functions of time. In fact, there are no simple mathematical formulas that perfectly describe their shape, so there's no simple mathematical method for making these sounds. However, we can start the design of our system to re-create musical sounds by looking at simple approximations for our complicated musical signals. An **approximation** is a calculation or procedure that makes a reasonable, but not perfect, copy of what we want or expect to hear. We use approximations because they make our life easier in some way. For example, approximations might make the overall cost of our digital band affordable to everyone, or they might make the digital band smaller and lighter to carry. We can describe these approximations by using simple mathematics. After we figure out how to use these approximations, we can explore possible ways to make the sounds more rich and realistic.

Approximation: A calculation or procedure that makes a reasonable, but not perfect, copy of what we want or expect to hear.

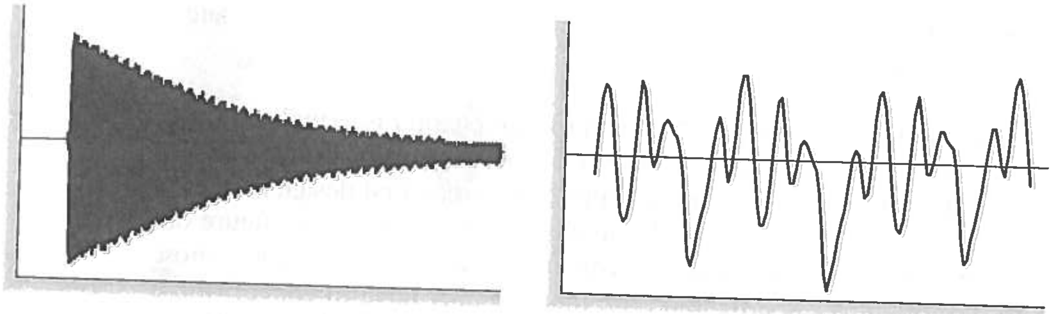


Figure 2.12 The guitar signal in Figure 2.11 shown over two different time scales. The plot on the left shows several seconds of the signal, while the plot on the right shows several milliseconds of the signal.

To proceed with our design, we will first need to figure out how to plot functions from their mathematical descriptions, so that we can compare the plot of a sound's approximation to that of the actual sound signal. If we have a formula for a signal, how do we plot it? Any signal that is a function of time is written as $s(t)$, where t has units of time. The value of $s(t)$ at any given time instant t is called the signal's **amplitude**, which is the height of the signal as measured from the time axis. When we plot such signals, if $s(t)$ is positive the value of $s(t)$ is the height of the curve above the t-axis, or time axis; if $s(t)$ is negative, it will be plotted below the time axis. Figure 2.13 shows three examples of simple functions and their corresponding plots.

Amplitude: The height of a signal, *s*, at time, *t*.

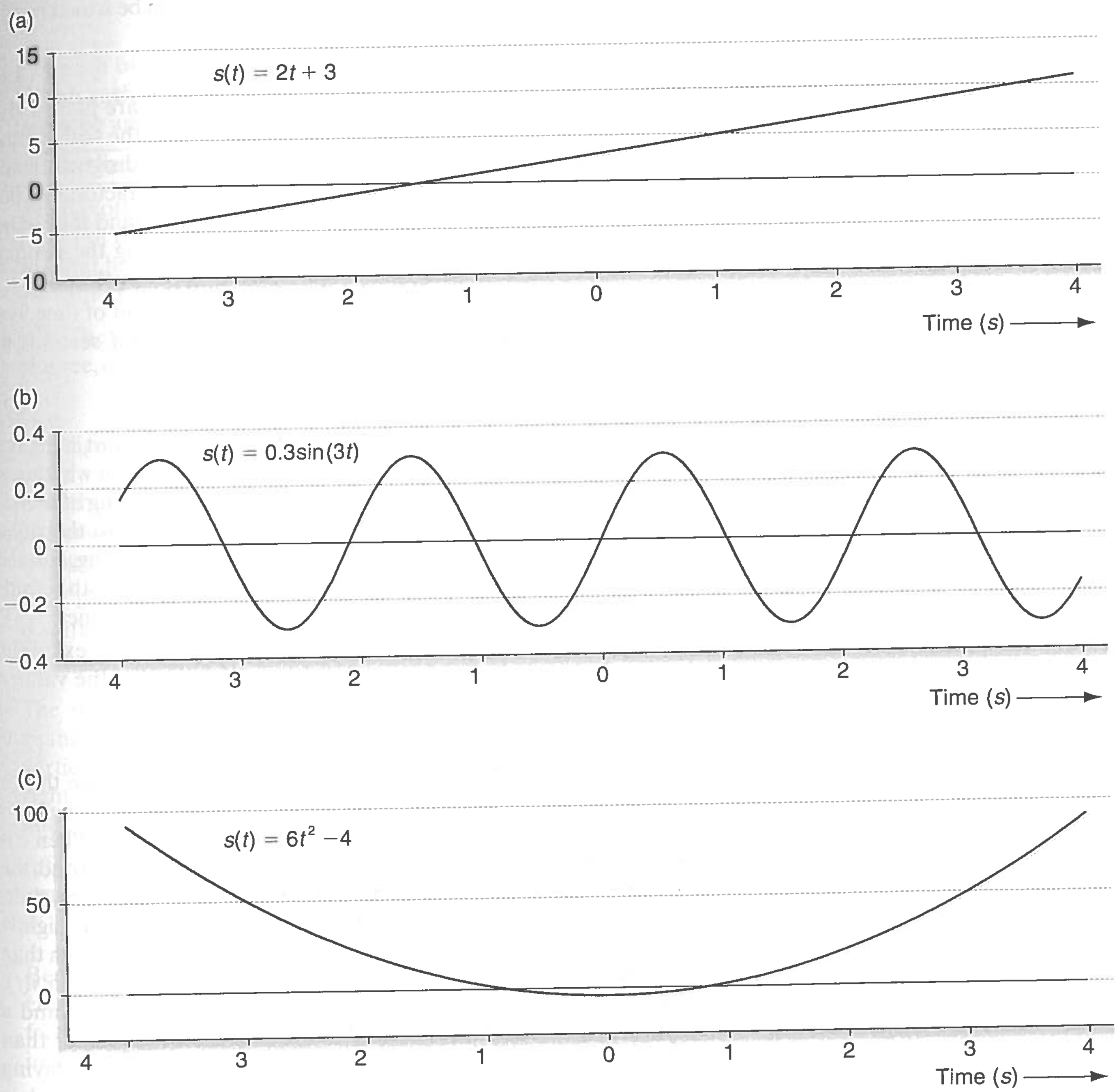


Figure 2.13 Plots of three different equations for $s(t)$ for values of t ranging from −4 to +4 seconds.

If we compare the plot of the function in Figure 2.13(b) with the plot in Figure 2.12(b), we see that they look very similar, which means that when they are converted to sound by a speaker, they will sound similar. This realization gives us the hope that we might be able to approximate a musical instrument's sound with some simple signals that we can compute easily. To produce realistic-sounding music from mathematics, our mathematical function should look as much as possible like the instrument's signal. Therefore, we will need simple ways to change the mathematical function to make it look more like an instrument signal. There are three basic changes we can implement:

1. *Scale the amplitude:* Multiply the height, or amplitude, of a signal by a value. For example, the signal $s(t)$ in Figure 2.13(b) can be scaled by the value A to make a new signal $x(t)$ as follows:
$$x(t) = A \times s(t) \tag{2.1}$$
When we listen to the sounds of $s(t)$ and $x(t)$ as they are played, the latter signal will sound either louder or softer to us. If the scale factor A is greater than one, scaling the amplitude of a sound signal makes the plot taller and thus the sound louder. If the scale factor A is between zero and one, the height of the plot is reduced, and the sound gets softer. We could create the same effect by turning the volume control on an audio sound system up or down, respectively.
2. *Shift the time:* Move a signal left or right by some amount of time. For example, the signal $s(t)$ can be shifted by a value of d seconds to make the signal $y(t)$ as follows:
$$y(t) = s(t + d) \tag{2.2}$$
The sounds of $s(t)$ and $y(t)$ will be the same to us; we just hear the sound of $y(t)$ earlier or later than that of $s(t)$, depending on whether d is positive or negative. When d is negative, $y(t)$ starts playing after $s(t)$ starts playing, and the plot of $y(t)$ looks like $s(t)$ shifted to the right. When d is positive, $y(t)$ starts playing before $s(t)$ starts playing, and the plot of $y(t)$ looks like $s(t)$ shifted to the left. In either case, both sounds play for the same length of time; they just start at different times.
3. *Scale the time:* Multiply the time variable t by a value. For example, the signal $z(t)$ can be made from $s(t)$ by multiplying t by the value c as follows:
$$z(t) = s(ct) \tag{2.3}$$
When we compare the sounds of $s(t)$ and $z(t)$, we notice three things. First, the duration, or length, of the two sounds is different. When c is greater than one, the duration of $z(t)$ is less than that of $s(t)$. When c is less than one, the duration of $z(t)$ is greater than that of $s(t)$. Second, the two sounds may start at slightly different times, as if $z(t)$ has been shifted to the left or right. Third, the characteristics of the sounds are slightly different as well. When c is less than one, $z(t)$ sounds lower in pitch than $s(t)$. When c is greather than one, $z(t)$ sounds higher in pitch than $s(t)$.

These effects can be demonstrated easily with a turntable and a vinyl record. If we spin the turntable too fast, then c is greater than one and the sound will last for a shorter time than if it were playing at a normal speed. In addition, the sound will seem to have a higher pitch. For example, a man's voice may sound more like a child's voice. If we spin the turntable too slowly, the sound takes longer to play and is much lower in pitch. These notions of lower and higher pitch are directly connected to how we perceive musical information. We'll learn more about pitch shortly.

EXAMPLE 2.1 Plotting Signals

Suppose we have the signal $s(t) = 2t + 4$, which is similar to the signal shown in Figure 2.13(a). Three different signals $s_2(t), s_3(t)$, and $s_4(t)$ are created from $s(t)$ as follows:

$$s_2(t) = 3s(t)$$
$$s_3(t) = s(t - 1)$$
$$s_4(t) = s(2t)$$

Plot each of these new signals on its own set of axes, and then plot $s(t)$ on each set of axes. Verify that each signal is a line, and identify the slope and the vertical and horizontal intercepts of each signal.

Solution

When plotting equations or signals, it helps to get an idea of what to expect in the plot. For example, we could say that $s(t) = 2t + 4$ is a line that has a slope of 2 and that, when $t = 0$, it crosses the s-axis at the value $s = 4$. When $t = -2, s(t) = 0$.

(a) The signal $s_2(t) = 3s(t)$ changes the amplitude of $s(t)$ by a factor of three, or

$$s_2(t) = 3s(t) = 3(2t + 4) = 6t + 12$$

This signal is a line that has a slope of 6 and crosses the s-axis at the value $s = 12$. Because the amplitude of $s(t)$ is scaled by a factor of three, both the slope and the vertical intercept are increased by a factor of three. However, the horizontal intercept (where $s = 0$) does not change. Plots of both $s_1(t)$ and $s(t)$ are shown in Figure 2.14(a).

(b) The signal $s_3(t) = s(t - 1)$ is a delayed version of $s(t)$. The line is shifted to the right by 1 second, because

$$s_3(t) = s(t - 1) = 2(t - 1) + 4 = 2t + 2$$

The slope is still 2, because we only shifted the signal; we did not change its amplitude. However, the shift changes both intercepts. The vertical intercept is changed from 4 to 2, because $s_3(t)$ is shifted to the right of $s(t)$. And now $s_3(t) = 0$ at $t = -1$, which is one time unit later than when $s(t) = 0$. This plot is shown in Figure 2.14(b).

(c) The signal $s_4(t) = s(2t)$ is a time-scaled version of $s(t)$. This operation compresses time by a factor of two, such that

$$s_4(t) = s(2t) = 2(2t) + 4 = 4t + 4$$

Because time has been compressed, the slope of the original signal increases by a factor of two and the horizontal intercept is reduced by a factor of two. Now $s_4(t) = 0$ when $t = -1$ instead of -2. The vertical intercept is not changed. This plot is shown in Figure 2.14(c).

This example illustrates the three basic operations we can use to change simple mathematical functions to better match real instrument signals.

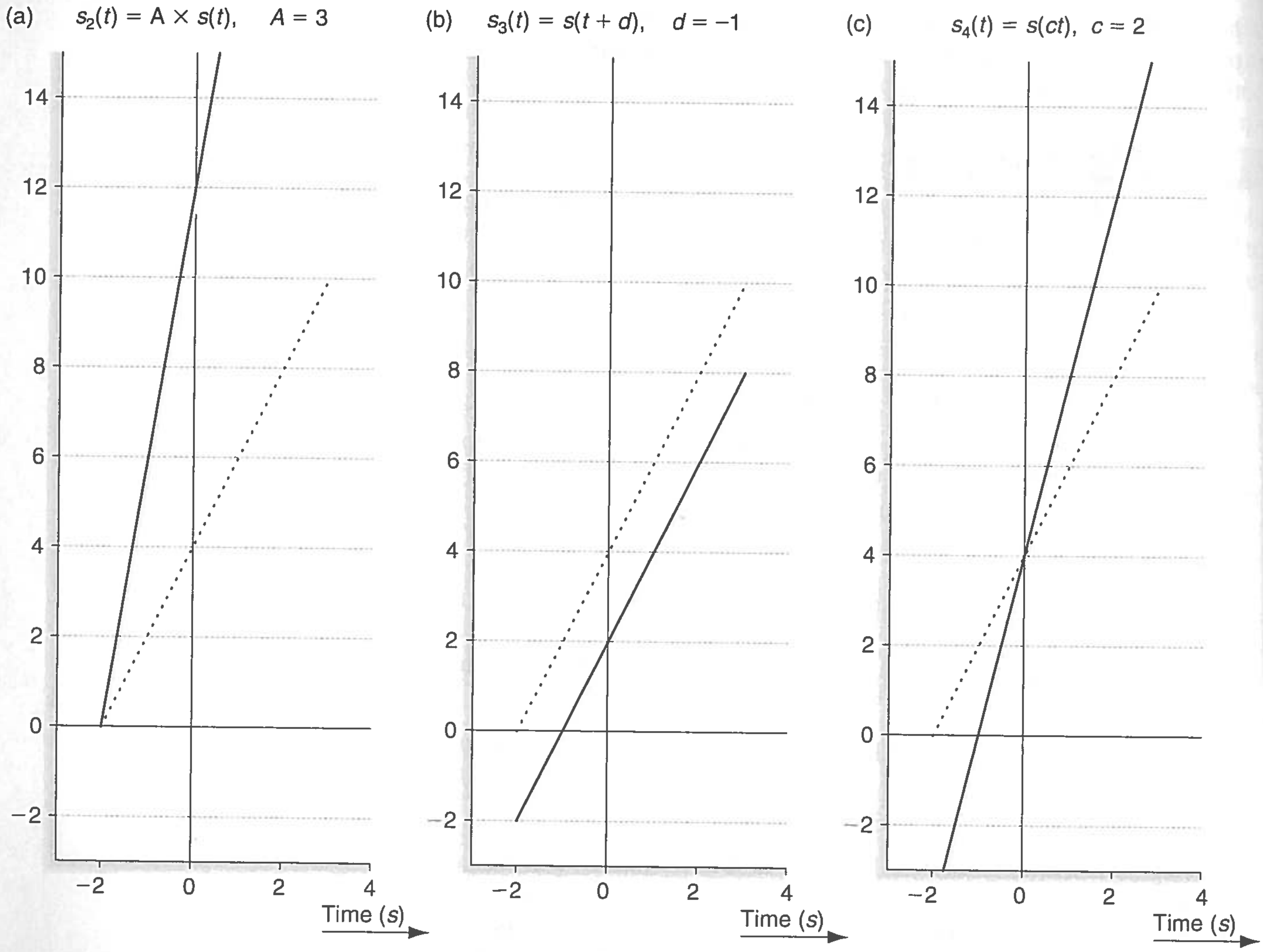


Figure 2.14 Plots of a signal $s(t)$ for t between −2 and 2 after (a) scaling its amplitude, (b) shifting in time, and (c) scaling the time variable. The original $s(t)$ is plotted as a dotted blue line for reference.

Pitch, Frequency, and Periodic Signals

To make music, we need to construct signals that have the special characteristics of signals produced by instruments. What are these special characteristics? We all know that music sounds different from other sounds, such as the roar of an engine. What makes music different? The answer lies in how instruments make sound.

Notes: The signs with which music is written.

Pitch: The perceived frequency of the sound of a note when we hear it.

When a musician plays an instrument—by striking an individual key of a piano, for example—each sound that is played is called a **note**. Some notes sound high, whereas others sound low. This perceptual difference is called **pitch**. Pitch is tied closely to a particular characteristic of a musical instrument's sound signal: the rate at which the sound *oscillates*, or how fast the amplitude of the signal moves up and down over time. We will need to learn how our perception of pitch is related to the signal of our device so that our system will to be able to create different notes.

Figure 2.15 shows the signals from two different notes that have been played on a piano keyboard. On the left, both signals are displayed over

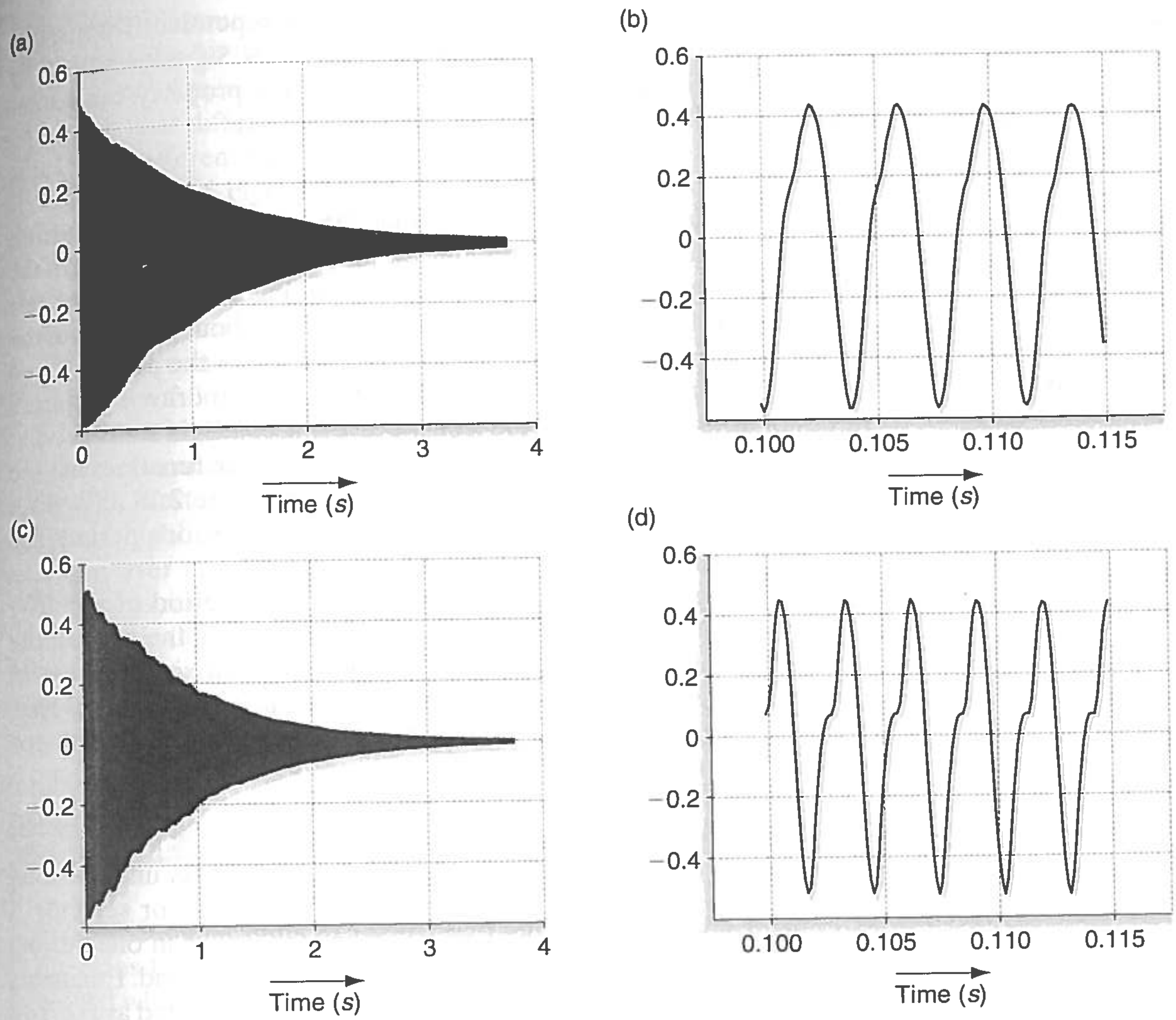


Figure 2.15 Plots of two piano notes are shown on the left for a time interval of 4 seconds each. The 15-ms interval highlighted in red in (a) is shown in (b). The 15-ms interval highlighted in red in (c) is shown in (d). The note shown in (c) has a higher pitch than the one shown in (a), so the highlighted part plotted in (d) looks compressed horizontally compared with (b).

4 seconds so that we can see the complete sound from when it starts to when it ends. On the right, we've zoomed in on a short portion of each sound over a time interval of 0.015 seconds (s), or 15 milliseconds (ms). This small 15-ms interval is highlighted in red on the signal plots on the left. The plots on the right show the characteristics of the sound created by the piano and of the specific note being played. Both signals on the right have a similar shape because they were both made by a piano as opposed to some other instrument. One of the signals, however, is compressed in time relative to the other, so it oscillates up and down at a different rate. This difference in time scale is what causes us to hear one note at a higher pitch than the other. Our observation gives us a key idea for making musical sound mathematically: Once we know the characteristic signal of an instrument for one note over a short interval, we can approximate other notes from the same instrument by scaling the time variable, thus expanding or compressing the signal in time.

Like the guitar sound in Figure 2.12, each signal on the right side of Figure 2.15 has an oscillating, or vibrating, quality to it. Over the time

Periodic Signal: A signal that exactly repeats at regular intervals.

Period: The repeating interval of a periodic signal.

INTERESTING FACT:
High-fidelity music systems are designed to produce sounds between 20 Hz and 20 kHz because that is the range of pitches humans can normally hear.

Fundamental Frequency: A mathematical quantity describing the repetition rate of a periodic signal.

Hertz: The units of frequency, represented in periods/second, or seconds^{-1}.

interval shown, each signal is almost perfectly repetitive. That is, one small part of the signal repeats at regular intervals. Signals that exactly repeat are called **periodic signals**. This repeating property can be described mathematically for a periodic signal $p(t)$ as

$$p(t) = p(t + T) \tag{2.4}$$

where T is called the **period** of the signal. Described in words, Equation (2.4) means the following: Shifting a periodic signal by T seconds to the left gives the same signal again. We can see in Figure 2.15 that the two notes have different periods. The period of one is about 3.8 ms, and the period of the other is about 2.9 ms. And if we know the period T and the values of $p(t)$ over one complete period, we can draw $p(t)$ for all values of time by repeating the signal every T seconds.

Let's draw a simple block that generates periodic functions for the block diagram of our digital band. As shown in Figure 2.16, this block has two inputs: the period T, and the shape of $p(t)$ for one period. The output is the signal $p(t)$ for all values of time.

The pitch of these signals is determined by the period of $p(t)$: The shorter the period, the higher is the pitch; the longer the period, the lower is the pitch. Since pitch gives us only an approximate sense of how high or low a sound is, engineers have come up with a more precise measure of pitch called the **fundamental frequency**. It can be computed from the period T of the sound as follows:

$$f = 1/T \tag{2.5}$$

What are the units of frequency? Since the period T has units of time (seconds), the units of frequency are periods/second (s), or seconds^{-1}. Since the value of T tells us how many seconds there are in one period, the value of f tells us how many periods there are in 1 second. Engineers have named this unit of frequency **Hertz**, which is abbreviated as Hz. You may have seen the units of Hz before when reading about high-quality audio equipment that can produce frequencies as low as 20 Hz and as high as 20 kHz. A 20-Hz sound has a period of $\frac{1}{20}$ s $= 0.05$ s, or 50 ms, whereas a sound with a frequency of 20 kHz $=$ 20,000 Hz has a period of $\frac{1}{20{,}000}$ s $= 0.00005$ s, or 0.05 ms. This is the range of frequencies that a young man or woman can normally hear.

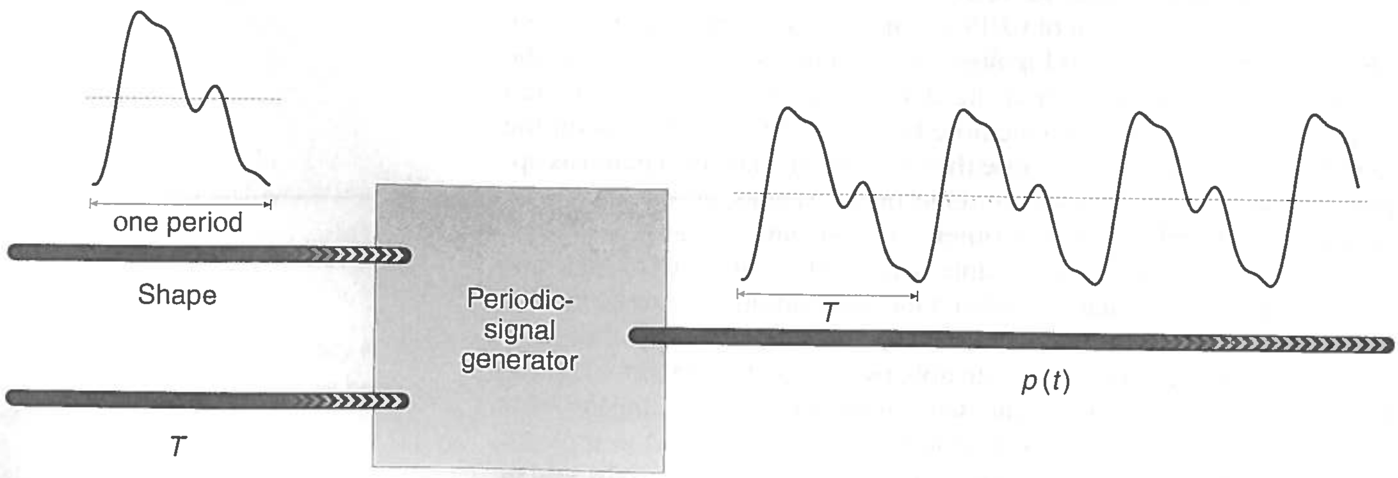


Figure 2.16 Block diagram of the periodic-signal generator.

When an instrument is making a sound, it is producing a periodic signal that has a fundamental frequency. Our perception of pitch comes from this fundamental frequency of a sound. When we hear periodic signals, we hear differences in the pitches of notes because the different notes have different fundamental frequencies.

The center of Figure 2.17 shows the names of notes assigned to eight of the keys in the middle of a piano keyboard, as well as the fundamental frequencies of the keys. The leftmost highlighted key in this cluster of eight shaded keys is called middle C. This key is often used as a reference key for piano music. If we compare the F-above-middle-C note, with a frequency of 349.23 Hz, with the middle-C note, which has a frequency of 261.63 Hz, we find that the F note sounds higher, because it has a higher frequency. We also say that its pitch is higher.

There are only seven unique note names, from the letter A to the letter G, in Western music. For notes that are above and below those on the piano highlighted in Figure 2.17, the letters A through G are used over and over again to specify the frequencies we hear. Every time a note name is reused above this range, the note's frequency is twice that of the identically named note below. Likewise, every time a note name is reused below this range, the note's frequency is half that of the identically named note above. In the figure, the frequencies of all the C notes are shown. As we move to the right, the frequency of each C note is double the frequency of the previous C note.

INTERESTING FACT:

Heinrich Hertz (1857–1894) was a German physicist. An experimentalist, he demonstrated that electromagnetic waves exist and showed how they could be made. His discoveries paved the way for all modern wireless communications technologies, including radio, television, cell phones, and radio astronomy.

EXAMPLE 2.2 Find the Frequency of a Periodic Signal from its Plot

From the signal plots in Figure 2.15 and the definition of the frequencies for piano keys in Figure 2.17, determine which notes were played when these signals were recorded.

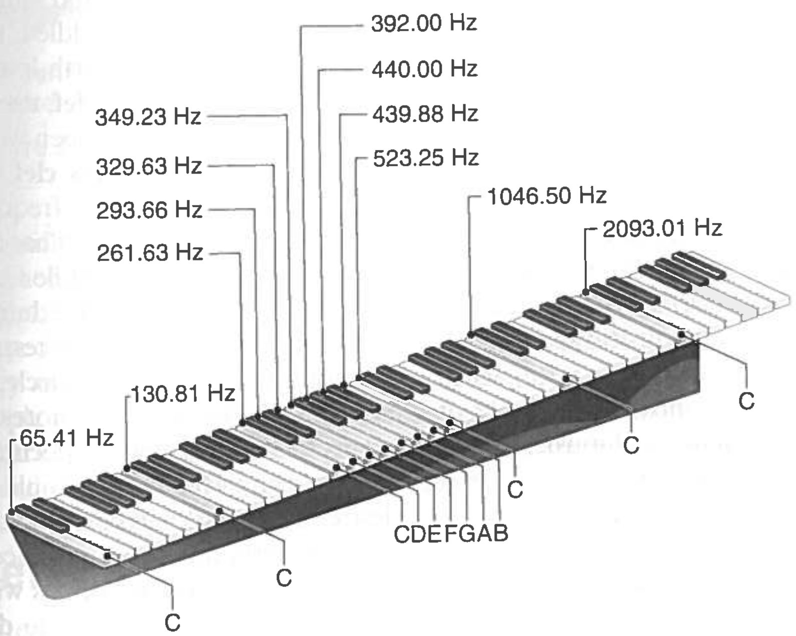


Figure 2.17 Frequencies for notes on the piano keyboard. The frequency labels shown in the upper portion of the figure correspond to the white keys labeled with the names of the respective notes in the lower portion. The frequencies of all notes named C increase by a factor of two moving from left to right.

T = Period from Plots	$f = 1/T$	Closest Piano Key
2.9 ms	3.4×10^2	F at 349 Hz
3.8 ms	2.6×10^2	C at 262 Hz

Solution

From the plots, we already estimated that the periods of the two signals were 2.9 ms and 3.8 ms. Using Equation (2.5), we can compute the corresponding fundamental frequencies from the period values. The frequencies are shown in the center column of the table to the left. Our measurements from the plot are not accurate enough to compute the frequencies to the numerical accuracy shown in Figure 2.17, but since the notes came from a piano, we can choose the piano key that has the frequency closest to our result. From this calculation, we see that the first note corresponds to middle C and the second note corresponds to F above middle C.

Melodies and Notes

Melody: A sequence of notes that make up a piece of music.

Score: Notation showing all parts or instruments.

Clef: The sign written at the beginning of a staff to indicate pitch.

Treble Clef: Also called G clef. Indicates that the note on the second line of the staff is G' (G above middle C).

Bass Clef: Also called F clef. Indicates that the note on the fourth line is F below middle C.

Tempo: The speed of a piece of music.

How interesting a song is to us depends on many factors, but perhaps the most important factor is the song's melody. A **melody** is a sequence of notes that make up a piece of music. When you whistle a song, you are whistling a melody. Notes are the "letters" that make up the "language" of the melodies that we hear in music. To describe melodies so that others can reproduce them with instruments, we need a formal method to list the sequence of notes in the music. The description we use must specify the frequency of each note as well as its time duration. The traditional method by which music is specified uses the notation and graphical presentation of music scores.

A complex composition such as *Don Giovanni,* by Wolfgang Amadeus Mozart, can be performed from its sheet music, or score. A small part of this score is shown in Figure 2.18(a). The **score**, which is a kind of recipe for creating music, represents in a graphical manner what notes should be played and when they should be played. The system of lines and spaces is known as a staff or stave. Treble clef, also called G clef, represents a G, and indicates that the note on the second line of the staff is G' (G above middle C). The bass clef, also called F clef, indicates that the note on the fourth line is F below middle C. Figure 2.18(b) shows a picture of the **treble clef**, used for higher frequency notes, in which the names of the notes have been written on the corresponding lines and spaces of the figure. The **bass clef** (pronounced "base clef") shown in Figure 2.18(c) is used for lower frequency musical notes. The C that appears in the lower part of the bass clef has a frequency of 130.8 Hz, which is exactly half of the frequency of middle C.

Figure 2.18(d) shows the note markings used to indicate the duration of notes in a musical score. The duration of notes is described in terms of fractions of a whole note, which is the note that looks like a circle. The other notes shown are called half, quarter, eighth, and sixteenth notes. The speed of a piece of music, more commonly called its **tempo**, is specified at the beginning of the piece and sets how long any whole note should last. To make music, we simply place the desired note markings corresponding to the durations of the notes we want to play onto the lines and spaces of the bass or treble clef. The vertical positions of the note markings within the clef specify what frequencies are to be played, while the note durations are specified by the shape of the note. The order of the notes proceeds from left to right on the page just like this text.

If we want to re-create a piece of music from a score, we first identify each note and its frequency and then create a corresponding periodic sig-

Figure 2.18 Music notation.

nal with the fundamental frequency of each note to be played. The shape of each note on the score tells us how long to play its corresponding signal, and the sequence of notes tells us the order in which they should be played. When two or more notes are aligned vertically in a musical score, they are played at the same time. This structure is called a **chord**. An example of a chord is shown in the highlighted part of the score shown in Figure 2.19 (repeated from Figure 2.18(a)). In the figure, each note in the chord is highlighted on the keyboard shown below the score.

INTERESTING FACT:
The sheet music we use today employs a form of musical notation that dates from 1600 A.D. in Europe. It was used mainly in churches and cathedrals where music was performed for ceremonies. Music notation was invented as a way to communicate ideas, much like a language.

Chord: A collection of simultaneously played tones.

Refining Our Design of the Digital Band

After our study of musical sound and signals, we are in a better position to specify the blocks within our digital-band design in Figure 2.1. For the moment, we shall consider a design that plays only one note at a time. While this design may not produce the high quality of music that we ultimately want, it will allow us to test our approach in a controlled

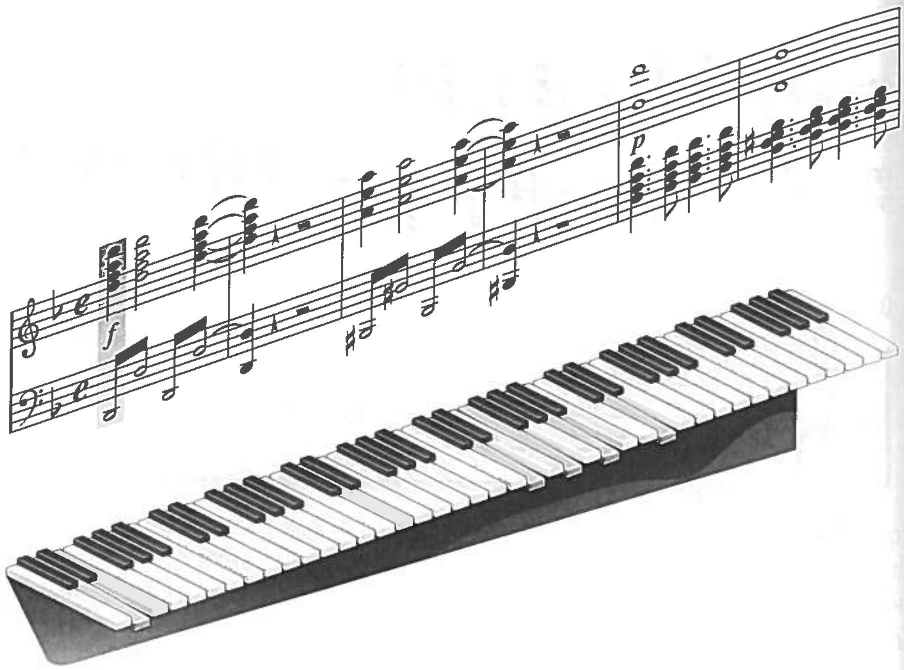
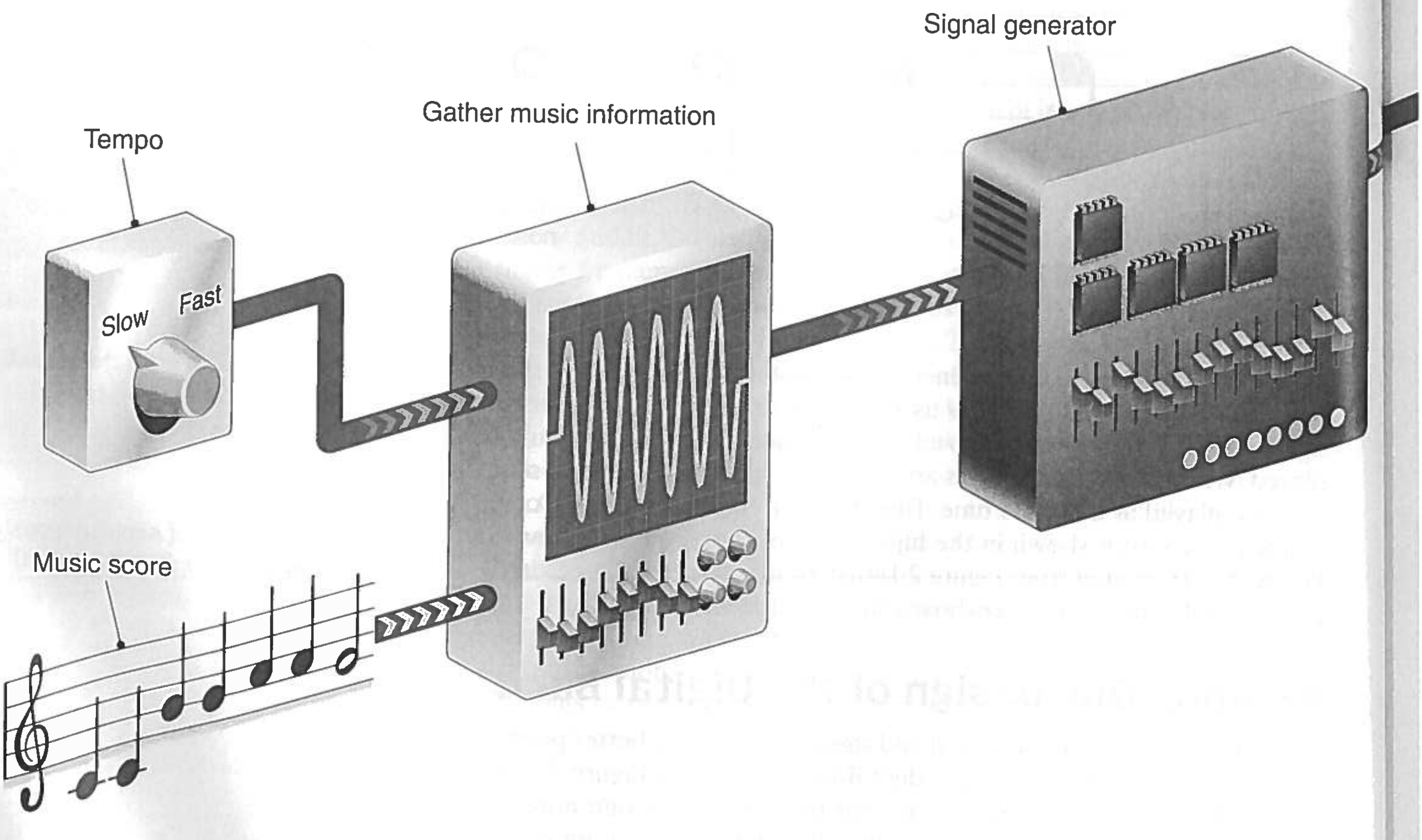


Figure 2.19 Five notes played on the keyboard at the same time to create the first chord.



manner. Our improved design is shown in Figure 2.20. The input to our system is the musical score that specifies the notes of the song we want to hear. The frequencies of these notes are used by the sound-generation block to create signals for the specified notes. The sound-generation block also uses the shape of a periodic signal to define the signal's sound. We will use $T = 1/f$ to compute the period from the frequency f. The output of this signal generator is sent to a sound-creation device, which usually contains a loudspeaker and amplifier to boost the amplitude of the signal so that we can hear it at our desired volume.

As in all engineering designs, specifying the system can be easy when it is broken down into its parts or individual blocks. But how do all of these parts work? And how can we increase the capability of our system to enable it to play the sounds of different instruments or more complex sounds and music? In the next section, we'll address these questions, and, in the process, we will improve our design significantly.

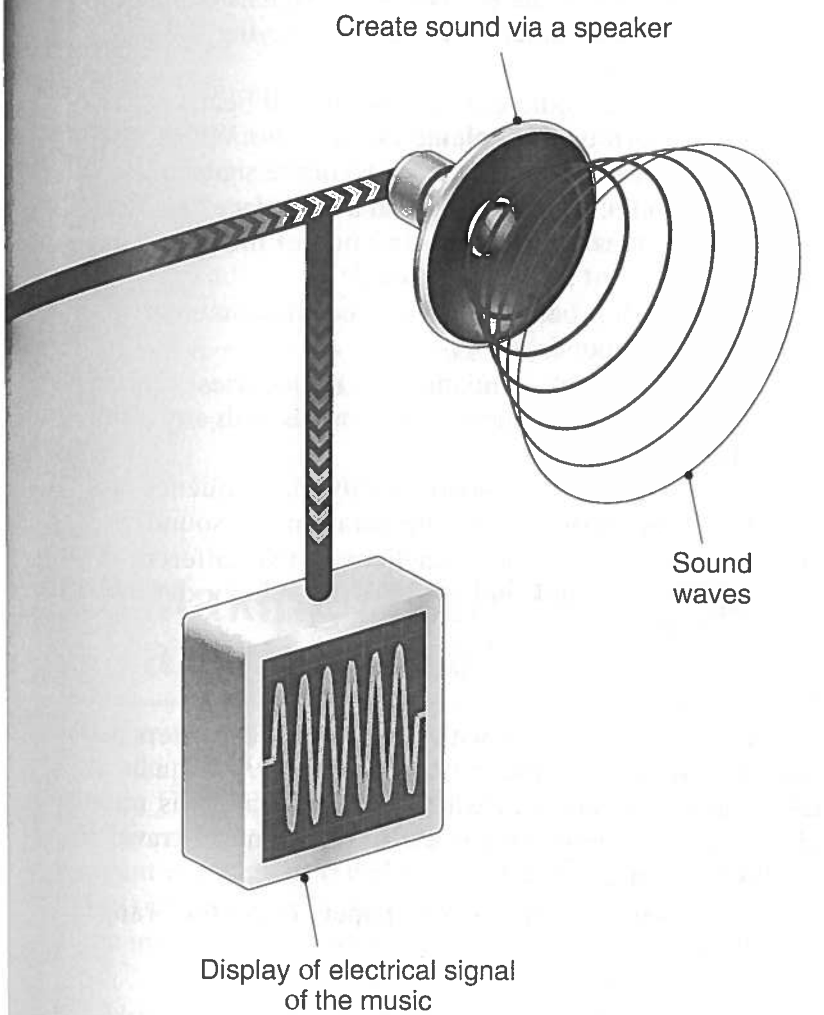


Figure 2.20 Block diagram for a system that can create the sound of a single instrument. The signal-generator block in the center is the periodic-signal generator.

EXERCISES 2.2

Mastering the Concepts

1. One of your friends says that loud sounds move air molecules like a race car zooming down the highway. Another friend says that loud sounds move air molecules like bumper cars bouncing back and forth in an amusement park ride. Who is right?
2. Can really complicated sounds, such as the sound of a 100-instrument orchestra, be represented by a single signal?
3. Name three ways that you can modify a sound signal.
4. Delaying a signal amounts to shifting the signal to the right on a plot. If $s(t)$ is the signal to be delayed and T is a positive number indicating the amount of time delay, is the delayed signal represented by $s(t + T)$ or $s(t - T)$? Draw an example so that you may be sure of your answer.
5. If the amplitude scaling factor A of a signal is negative, how will $s(t)$ be changed by scaling?
6. How are pitch and frequency related?
7. You hear two musical instruments playing the same tune. What can you say about the periods of the signals being produced by the two instruments if they are playing the same notes at the same time?
8. While listening to the radio one afternoon, you hear your favorite song and turn up the volume of the radio. Which have you changed, the amplitude or the period of the signal?
9. What is the period of the hour hand on a clock face? In other words, how long does it take for the motion of the hour hand to repeat? What about the minute hand?
10. Give the relationship between period and fundamental frequency of periodic sounds.
11. What determines which fundamental frequencies can be played on a piano? Can a piano make sounds with any given fundamental frequency?
12. What is a score? How does a score specify the frequency of a sound? How does a score specify the duration of a sound?
13. Would you expect the periodic functions of two different instruments playing the same note to look the same or different when plotted? Why?

Try This

14. The speed of sound increases with temperature. In meters per second, the speed of sound is approximately equal to ($331.4 + 0.6 \times$ temperature) when the temperature is measured in degrees C. How long does it take sound to travel 1 mile at the freezing point, 0°C?
15. Plot the following functions on paper over the range $0 < t < 0.1$ s:

    a. $s_1(t) = 0.7 \sin(2\pi 45t)$
    b. $s_2(t) = 11t - 0.07$
    c. $s_3(t) = t - 0.25t^3$

16. Using the functions from Exercise 2.2.15, plot the following functions on paper over the range $0 < t < 0.1$ s:
    a. $s_4(t) = s_1(t - 2)$
    b. $s_5(t) = s_2(t + 3)$
    c. $s_6(t) = s_3(t + 3)$
17. Plot $s_1(t)$ for the following signals:

$$s_1(t) = -t - 2$$

$$s_1(t) = |t| = \begin{cases} t, & \text{if } t \geq 0 \\ -t, & \text{if } t < 0 \end{cases}$$

$$s_1(t) = 3 - 4t + t^2$$

    In each case, plot the corresponding signals $s_2(t)$, $s_3(t)$, and $s_4(t)$ as defined in Example 2.1.
18. Plot the following signals either by hand or on your calculator:
    a. $s_1(t) = 0.5\cos(0.3t + 0.2)$
    b. $s_2(t) = 2\,s_1(t)$
    c. $s_3(t) = s_1(t - 2)$
19. What is the fundamental frequency of the periodic signal in Figure 2.13(b)?

In the Laboratory

20. Have a friend bring a musical instrument to the lab. Use laboratory equipment from the "Plots of Speech" Infinity Project Experiment on page 44 to display the instrument's sound. Measure the signal produced by the instrument for several notes across the frequency range of the instrument and player. How do these signals differ? How are they similar?
21. When you whistle, you make a nearly periodic sound. Take a bottle, blow across the top of it, and measure its sound signal. How periodic is the signal? Can you measure its fundamental frequency?

2.3 Making Music from Sines and Cosines

The knowledge that we have gained about sound and music has enabled us to add many important and specific details to the design of our digital band in Figure 2.1. We are much closer to our design objective, but we need to provide more details than in our current design in Figure 2.20 if we want our system to make music. To do so, we will study each block, figure out exactly what it will do, and determine its inputs and outputs. Some important questions to help us in this task are as follows:

- How do we get our information into the device?
- How will this information be converted into a sequence of notes?
- What signals will we use to convert the notes into sound?

Using MIDI to Specify Information

Let's start by specifying the first block of our digital band in Figure 2.20. We've learned that music is described by a sequence of notes that define the melody we want to hear. Since the sound of each note can be made by a periodic function, this first block has to specify the fundamental frequency of each note. It also has to specify the order in which the notes are played and how long each note will last.

There are many ways to specify information of this sort. Sheet music contains a graphical presentation of this information that has been used by live performers for centuries. The following are some advantages of sheet music:

- The graphic layout of sheet music can be learned easily by most people, to the point that musicians can read it as fast as they can play.
- The use of pictures for the notes does not depend on any particular spoken language. Thus, anyone in any part of the world can learn to read and make music for others to hear and play. Sheet music represents a **standard** that allows musicians from anywhere in the world to play together.
- The paper on which sheet music is written is durable, portable, easy to make, and easy to copy.

Standard: A description for a method or process for using or building something that a group of people has agreed to use. By establishing a standard, people can use, enjoy, and even build on other people's work to make it better. Standards are important in music, because they allow us to make, share, and enjoy more music together.

Even with these advantages, sheet music is probably not the best solution for the first block of our digital band. If we were to use sheet music as our musical input, we would have to design a system to read the notes from the page. We could build such a system, but the cost and size of the scanning device would probably be too great for us to meet the other design goals for our digital band. Is there a better way if we start from scratch?

Let's think again about a piece of music and how it is specified. We know that each note corresponds to a frequency and lasts a specific length of time. Why not just list the frequencies and lengths of time in the order that they are played? We could create a document or computer file that has this information. Then, an electronic device could read the file and play the piece from the instructions contained in the file. The advantages of such a solution are many and include the following:

- Since the piece of music is in electronic form, it can be manipulated easily by a computer, saved on a CD or hard disk, and transmitted by electronic means such as the Internet.
- As with sheet music, we could convince others to use our description, allowing us to share the piece. In this way, we will have created a standard that allows us to communicate and share music more easily.
- Designers of electronic instruments can use our description to communicate information between other electronic devices and instruments, such as from a computer to an electronic piano, saxophone, or trumpet, and back again.

The biggest drawback to our method is that we would have to convert existing pieces of music to our new format—but given the ease with which we can communicate electronically, we'd have to convert each piece only once.

Our method of specifying a piece of music would allow us to create simple melodies, like "Mary Had a Little Lamb." There are a couple of problems with our methods, though:

1. It wouldn't enable us to make more complicated pieces of music that have periods of silence, called *rests*.
2. Most pieces of music have more than one note playing at a time, and there's no way for us to specify when more than one note should be played.

We need to change our method to add this information.

Fortunately, designers of electronic instruments have come up with a similar system to represent music electronically that overcomes the two problems we've identified. This system is called Musical Instrument Digital Interface, or **MIDI** for short. Music in this form is stored in MIDI files on a computer, often denoted by the extension ".mid". These files can contain many types of information, but the most critical information for our digital band is the note information. Music in a MIDI file is stored as a list of instructions to turn notes on and off. That way, two notes can be turned on at the same time, and we can turn all the notes off at any time as well. MIDI is also a great choice for our digital band, because it is already a widely accepted standard among companies that make and sell systems for creating and manipulating music.

MIDI (Musical Instrument Digital Interface): A specification for how information is communicated between electronic instruments. It was developed in the 1980s by engineers from Sequential Circuits, Roland Corporation, and Oberheim Electronics—three electronic-instrument companies. The specification was first published in August 1983 and has since been modified to include certain types of non-musical-performance information.

Figure 2.21 shows a simple score on the left and a corresponding pseudo-MIDI-file list of instructions on the right. On the left of the MIDI list is a column labeled "timestamp" that describes when an instruction occurs. For each timestamp value, we see a note value in the form of a letter and number, such as C4 and G4, and an "event" on the same line. The letter is the note name, and the number specifies the exact note with that name. In Figure 2.17, the note designated as C4 (middle C) has a frequency of 261.63 Hz. In our simple example, the events are either "On" or "Off," corresponding to whether we turn the note on or off, respectively. Other information that might be contained in a MIDI file includes the channel of the instrument being played, the loudness of the note, and other performance information. A more complicated example might show such information as well as multiple notes being played simultaneously.

For fun, play the notes in Figure 2.21 on a piano or some other instrument and see if you can identify this familiar children's song.

Making Signals

Now that we have a way of getting music into our digital band using a widely accepted standard, we shall turn our attention to the next block of the block diagram in Figure 2.20. This block takes the information contained in the MIDI file and turns it into signals. We will first explore simple and well-understood ways for making these signals.

The periodic signals shown in Figures 2.12 and 2.15 each have distinctive plots that look like they might be hard to describe mathematically. Moreover, we don't have a simple way to calculate them. Have you ever seen a saxophone or tuba key on a calculator?

A mathematically simple periodic function is one that you might already have seen plotted in a mathematics textbook: the **sinusoid**.

Sinusoid: A simple oscillating waveform created from the sine and cosine function.

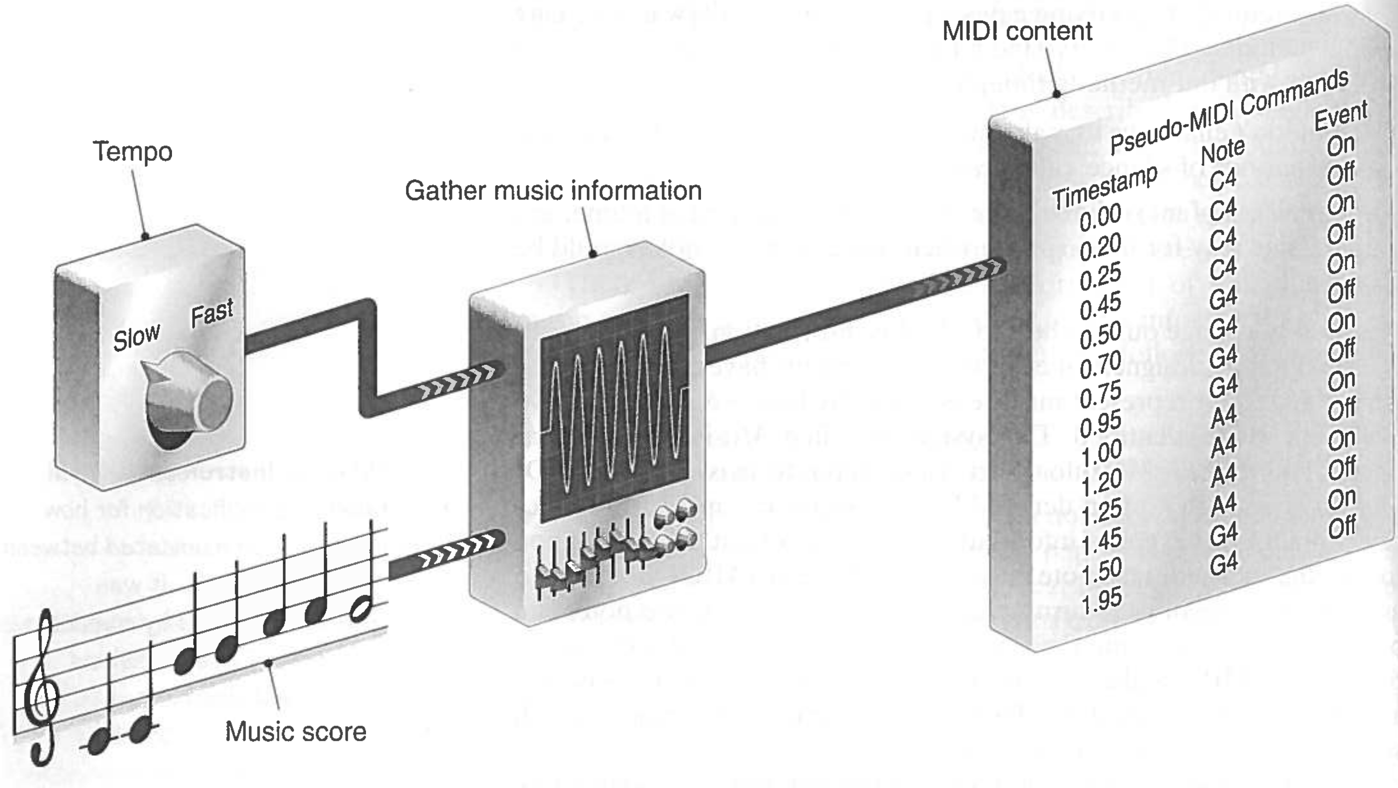


Figure 2.21 An example of how Musical Instrument Digital Interface (MIDI) can be used to specify information for a piece of music.

Figure 2.22 shows a sinusoidal signal being played by a MIDI file. Sinusoids are very common signals, and all of them have the same simple shape. The tuning fork in Figure 2.11 produces a change in acoustic pressure that is a nearly perfect sinusoid. A guitar string produces a signal that is approximately a sinusoid and becomes more and more like a sinusoidal signal as the sound decays away.

The Sine and Cosine Functions Sinusoids are a type of mathematical function that includes both the cosine and sine functions, using standard definitions from right triangles. Figure 2.23(a) shows a circle with radius one. Any point on the circle can be said to have coordinates (x, y), which are at a distance of one from the origin of the coordinate system, because the radius of the circle is one. The specific location on the circle can be described by the angle measured from the horizontal axis. The symbol for this angle is the Greek letter θ. When the angle θ is known, the Cartesian (x, y) coordinates can be computed from the sine and cosine of the angle. The x value is $\cos(\theta)/1$, or $\cos(\theta)$, and the y value is $\sin(\theta)/1$, or $\sin(\theta)$, where the first three letters of "sine" and "cosine" are used as abbreviations.

We can get a good idea of what $\cos(\theta)$ and $\sin(\theta)$ look like as θ increases by studying the diagram in Figure 2.23. When $\theta = 0$, we are at the rightmost part of the circle, where $x = \cos(0) = 1$ and $y = \sin(0) = 0$. Similarly, at the top of the circle, $\theta = 90°$, so $x = \cos(90°) = 0$ and $y = \sin(90°) = 1$. Continuing around the circle to the leftmost point, we

INTERESTING FACT:

A tuning fork makes a sound that is almost a perfect sinusoidal signal. The signals from the sound of a flute and a person whistling are also very close to sinusoidal.

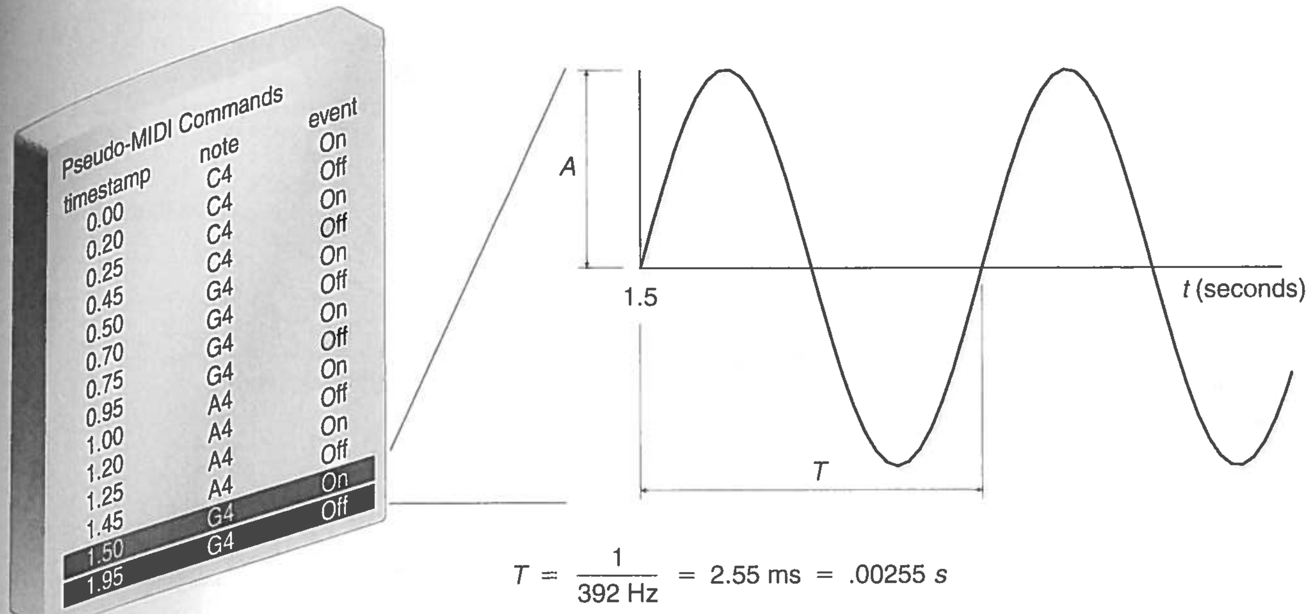


Figure 2.22 MIDI file producing a sinusoidal signal for the note G4 with a frequency of 392 Hz. The amplitude A and period T of the signal are shown.

have $x = \cos(180°) = -1$ and $y = \sin(180°) = 0$. At the bottom of the circle, $x = \cos(270°) = 0$ and $y = \sin(270°) = -1$. When we complete our trip around the circle to 360°, we find ourselves back where we started, and the same sequence of values for sine and cosine will be produced as the angle continues to increase. The values for some intermediate angles can be measured from the coordinate values on the circle, since $x = \cos(\theta)$ and $y = \sin(\theta)$. These values are summarized in Table 2.1.

We can compute the Cartesian coordinates of any point on this circle by using the cos() and sin() functions on a scientific calculator.

EXAMPLE 2.2 Computing Cartesian Coordinates

Using your calculator, determine the x and y coordinates of the point A on the plot in Figure 2.23(b).

Solution

After making sure that your calculator is set to angular measurements in degrees, the values of x and y can be computed to three decimal places as

$$x = \cos(72) = 0.309 \quad \text{and} \quad y = \sin(72) = 0.951$$

These numbers seem to agree with the points along the x and y axes of the plot. If we measure $\cos(\theta)$ and $\sin(\theta)$ for all values of θ between 0 and 360°, we can plot $\cos(\theta)$ and $\sin(\theta)$ as a function of θ, as shown in

INTERESTING FACT:

MIDI can also be used to transmit information from one location to another, such as through a cable from a piano keyboard to a separate sound module or even through a connection over the Internet. With MIDI, several musicians can play music together without being in the same room, participating in a "virtual jam session."

INTERESTING FACT:

Mathematicians often use Greek letters for symbols so they will not be confused with text. The most common symbol used for an angle is the Greek letter theta, θ.

Table 2.1 Values of $x = \cos(\theta)$ and $y = \sin(\theta)$ for different angles in Figure 2.23(a)

θ in degrees	0°	45°	90°	135°	180°	225°	270°	315°	360°
θ in radians	0	$\pi/4$	$2\pi/4$	$3\pi/4$	π	$5\pi/4$	$6\pi/4$	$7\pi/4$	2π
$x = \cos(\theta)$	1	0.707	0	−0.707	−1	−0.707	0	0.707	1
$y = \sin(\theta)$	0	0.707	1	0.707	0	−0.707	−1	−0.707	0

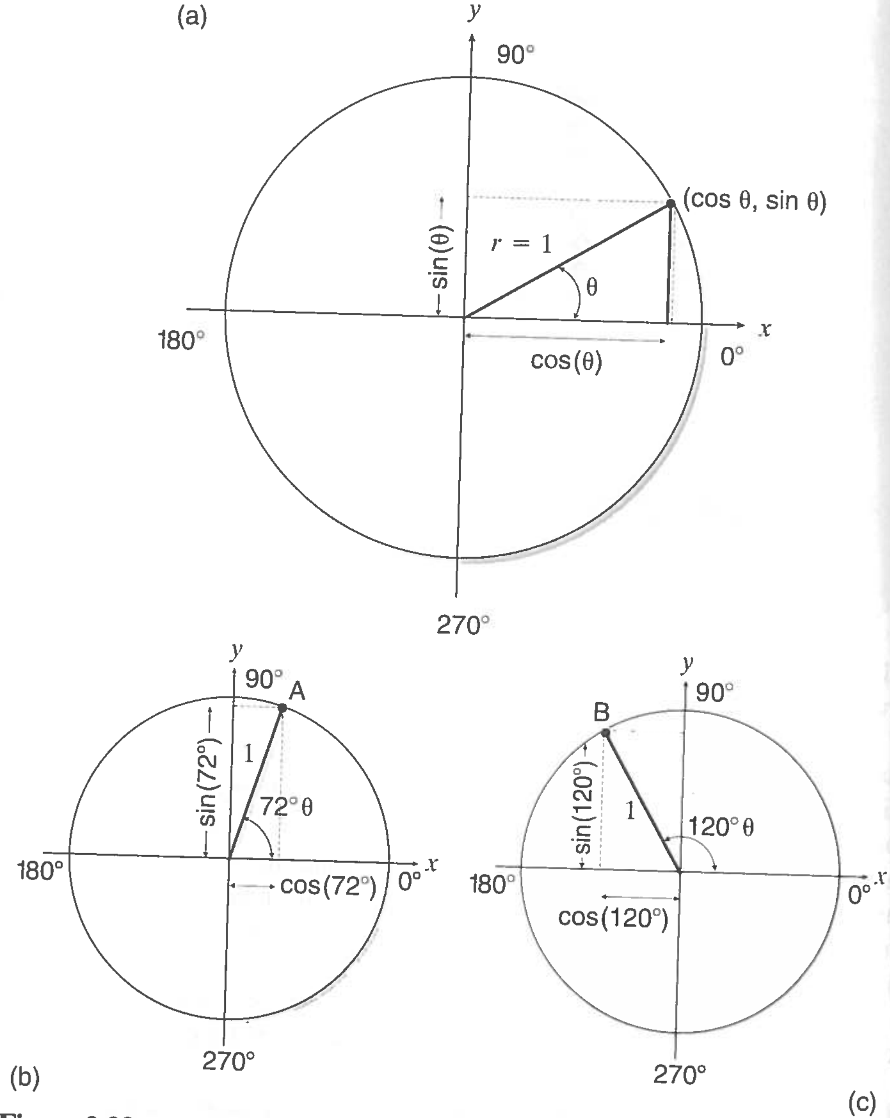


Figure 2.23 A circle with a radius of one, showing the sine and cosine of an angle.

Figure 2.24. You can verify that the plots and the table have the same values of $\cos(\theta)$ and $\sin(\theta)$ for the angles listed in Table 2.1. From this figure, we clearly see that $\cos(\theta)$ is the same as $\sin(\theta)$ shifted to the left by 90°. This relationship is described mathematically as follows:

$$\cos(\theta) = \sin(\theta + 90°)$$
$$\sin(\theta) = \cos(\theta - 90°) \tag{2.6}$$

The effect of adding or subtracting 90° from the angle θ is the same as that of shifting a signal in time—the plots move to the left or right, respectively.

The sine and cosine functions are periodic functions with a period of 360°. Since we know the period and we know the value of the functions for one full period, we can extend the plots to as large a value of θ as we wish by repeating the values of the function every 360°:

$$\sin(\theta) = \sin(\theta + 360°)$$
$$\cos(\theta) = \cos(\theta + 360°) \tag{2.7}$$

INTERESTING FACT:

The human hearing system is not very sensitive to time shifts in sinusoids or in any other periodic signals. A human listener cannot hear the difference between a cosine and a sine function when played through a loudspeaker.

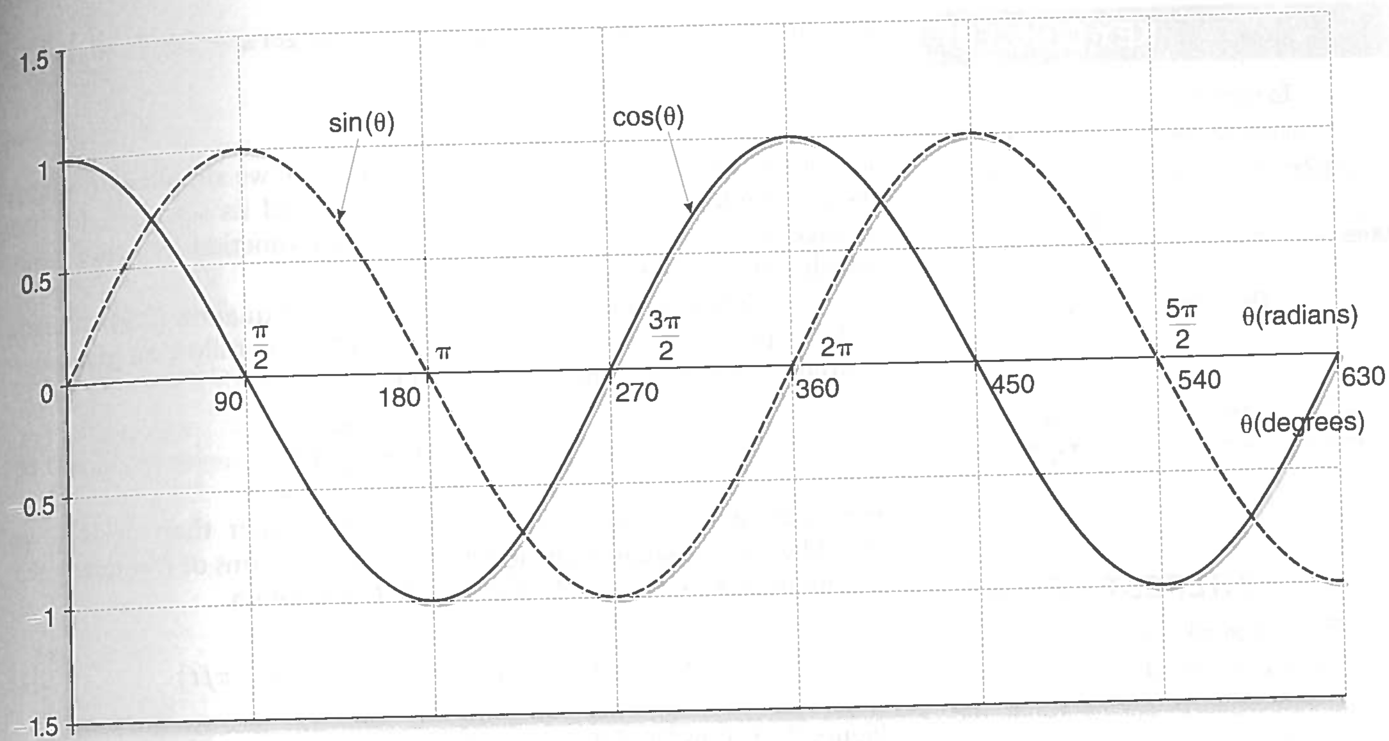


Figure 2.24 Sine and cosine as a function of angle. The angles are shown in both degrees and radians.

Infinity Project Experiment: Generating Sine and Cosine Signals

Sine and cosine signals are directly related to how a wheel rotates. From the spinning "dot" on your computer, can you see how a sinusoid is created? How about a cosine signal? Try changing the speed of rotation. What happens to the signal produced?

Using Radians to Measure Angles We use degrees for the units of an angle when we first learn about sines and cosines, but mathematicians and engineers often use a different unit of measure called a **radian**. For angles, 2π radians means the same thing as 360°. It turns out that using radians makes many of our equations simpler.

Radian: An alternative way to measure angles that is often used by engineers and mathematicians. One radian is about 57.3°, and 360° is 2π radians.

Cosines and Sines as Sound Signals Now that we have the cosine and sine functions as possible candidates for making signals, we would like to turn one of these functions into a sound signal that is a function of time. The simplest way to do this is to make θ equal to some constant multiplied by time so that the angle will increase as time increases. In other words,

$$\theta = c \times t \tag{2.8}$$

where c is our scaling factor and is constant for each note. We want the signal to be periodic with period T seconds, but sinusoids are periodic with period 2π radians (360°). So, we should choose the scaling factor

KEY CONCEPT

To convert degrees to radians, multiply the angle in degrees by $(2\pi/360)$ radians per degree, or

$$\text{radians} = (2\pi/360)\frac{\text{radians}}{\text{degree}} \times \text{degrees}$$

Degrees	Radians
0°	0
90°	$\pi/2$
180°	π
270°	$3\pi/2$

INTERESTING FACT:

Besides amplitude and frequency, there is a third parameter that defines a sine or cosine signal. The *phase* of a sine function is the amount in radians that the sinusoid is offset with respect to the origin. When the angle of a sine function is made proportional to time, changing the phase of the sine function is equivalent to time-shifting the sinusoidal signal by some amount. Phase is important for specifying the exact forms of sinusoidal signals and their combinations when matching a plotted signal. It is also very important for communications systems. Phase is less important when making music, because our ears are not at all sensitive to small changes in the phase of a sinusoid.

so that $\theta = 2\pi$ when $t = T$. Solving for c, we get $c = 2\pi/T$, so the correct relationship is

$$\theta = (2\pi/T) \times t \tag{2.9}$$

In addition to specifying the sinusoid's period, we should also specify its amplitude so that we can control how loud its sound is. For this change, we simply multiply the sine or cosine function by a number A, which represents the amplitude.

For the function $\cos(\theta)$, we choose θ as in Equation (2.9) and multiply by the amplitude scaling factor A to get the following general sinusoidal signal with amplitude A and period T.

$$s(t) = A \times \cos\left(\frac{2\pi}{T}t\right) \tag{2.10}$$

Because music is specified by frequencies rather than periods, we should write the sinusoidal-signal expression in terms of frequency. Recalling from Equation (2.5) that $f = 1/T$, we obtain

$$s(t) = A \times \cos\left(\frac{2\pi}{T}t\right) = A \times \cos(2\pi f t) \tag{2.11}$$

Figure 2.25 shows a plot of two cosine functions, one with a frequency of 625 Hz and one with a frequency of 250 Hz. The period T and amplitude A of each is marked on the plot. We can now see the meaning of both parameters: A corresponds to the maximum height of the sinusoid above or below from zero, and T is the repeating interval. The frequency of the sinusoid corresponds to the number of times the function repeats itself every second. We can think of the frequency f as the number of periods per second, and the period T as the number of seconds per period.

A third way to change a signal $s(t)$ is by shifting it in time, as shown in Equation (2.12). Shifting a cosine 90°, or $\pi/2$ radians, to the right produces a plot that is exactly the same as a sine. We can shift a sine or cosine signal that is a function of time by any amount. When we do, it will not start or end exactly at a value of zero (as does a sine function) or a value of one (as does a cosine function). Equation (2.12) shows the most general expression for a cosine signal that has been modified using amplitude scaling, time scaling, and time shifting:

$$s(t) = A \times \cos(2\pi f(t + d)) \tag{2.12}$$

EXAMPLE 2.3 Plotting Sines and Cosines

Plot the sine and cosine functions from $t = 0.0$ ms to $t = 6.0$ ms if each has a frequency of 400 Hz and an amplitude of 3.2.

Solution

We first specify the axes for our plot before drawing the functions. The horizontal axis goes from 0 to 6 ms. Since the amplitude of the sinusoid is 3.2, we know that the final scaled sinusoid will always be between −3.2 and 3.2 on the vertical axis. We next convert the frequency of the sinusoid into its period, using $T = 1/f$:

$$T = 1/400\text{ Hz} = 0.0025\text{ s}$$

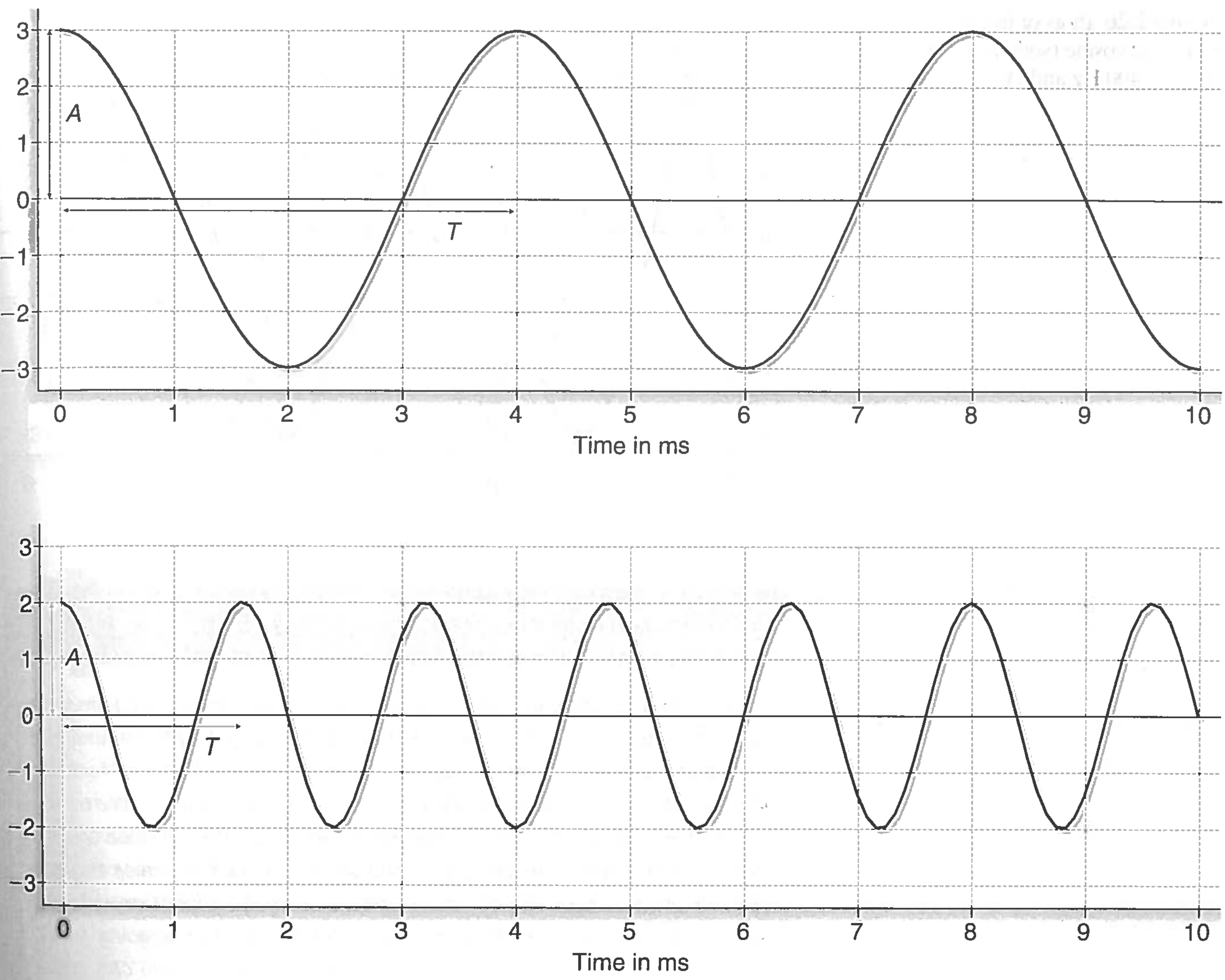


Figure 2.25 Plots of two cosine functions for 10 ms. In both plots, the amplitude A is shown by the vertical arrow and the period T is shown by the horizontal arrow. In the upper plot, the amplitude is 3 and the frequency is 250 Hz, so the period is 4 ms. In the lower plot, the amplitude is 2 and the frequency is 625 Hz, so the period is 1.6 ms.

There are 1000 ms in 1 second (s), so $T = 0.0025\,\text{s} \times 1000\,\text{ms/s} = 2.5\,\text{ms}$. The cosine will have its maximum value at $t = 0$, $t = T$, $t = 2T$, and so on. Let's mark the values of $s(t) = 3.2$ at times $t = 0.0$, 2.5 ms, and 5 ms on the graph. These values are shown in green in Figure 2.26. The cosine will have its minimum value at $t = T/2$, $3T/2$, $5T/2$, and so on. The values of $s(t) = -3.2$ at $t = 1.25$ ms and 3.75 ms are marked in blue on the plot. Finally, we can sketch the function $A \times \cos(2\pi ft)$, where $A = 3.2$ and $f = 400$ Hz, using our knowledge of what a cosine function looks like. This function is shown as the solid red line in Figure 2.26. We can repeat this process for the sine function, shown with a dotted blue line in Figure 2.26, by simply shifting all the points $T/4 = 0.625$ ms, or one-quarter period, to the right.

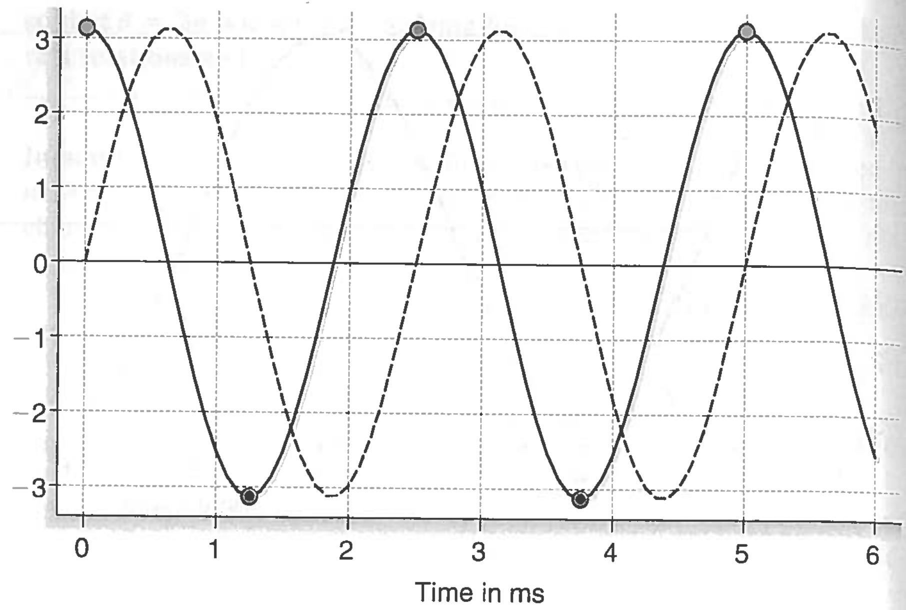


Figure 2.26 Plots of the sine (dashed blue) and cosine (solid red) functions for $f = 400$ Hz and $A = 3.2$.

INTERESTING FACT:

Hipparchus of Rhodes (190 BC to 120 BC), a Greek mathematician, is considered by some historians to be the inventor of trigonometry.

Infinity Project Experiment: Listening to Sines and Cosines

Digital devices can calculate easily the value of a function thousands or even millions of times every second. If we amplify and play the sound of a function using a device's loudspeaker, we can hear what the function sounds like. Try creating the sound of a sinusoid, and listen to the result. What does the function sound like? What happens when you raise the frequency of the sinusoid? What happens to the sound if you raise the amplitude of the sinusoid? Look at the plot of a sinusoid function while you change its amplitude or frequency, and determine what is going on. Do these plots make sense? What is the highest frequency you can make? What is the highest frequency you can hear? What is the lowest frequency you can make? What is the lowest frequency you can hear?

EXAMPLE 2.4 Writing the Equation of a Sine or Cosine Function from Its Plot

From the plot of the sinusoid in Figure 2.27, determine the values of the amplitude A and the frequency f, and write the equation for the sinusoid, using either the sine or cosine function.

Solution

By finding the highest and lowest values of the signal, we see that $A = 3.0$. We next need to find the period T by finding the interval between repetitions of the signal pattern. The maximum values occur at $t = 0.2$ s, $t = 1.0$ s, and $t = 1.8$ s. We can calculate the period as the time interval between maximum values, so $T = 1.0 - 0.2$ s $= 0.8$ s. Using Equation (2.5), we can compute the frequency $f = 1/T = 1/0.8$ s $= 1.25$ Hz. We could have picked any point on the curve to measure the period, but choosing two maximum values or two minimum val-

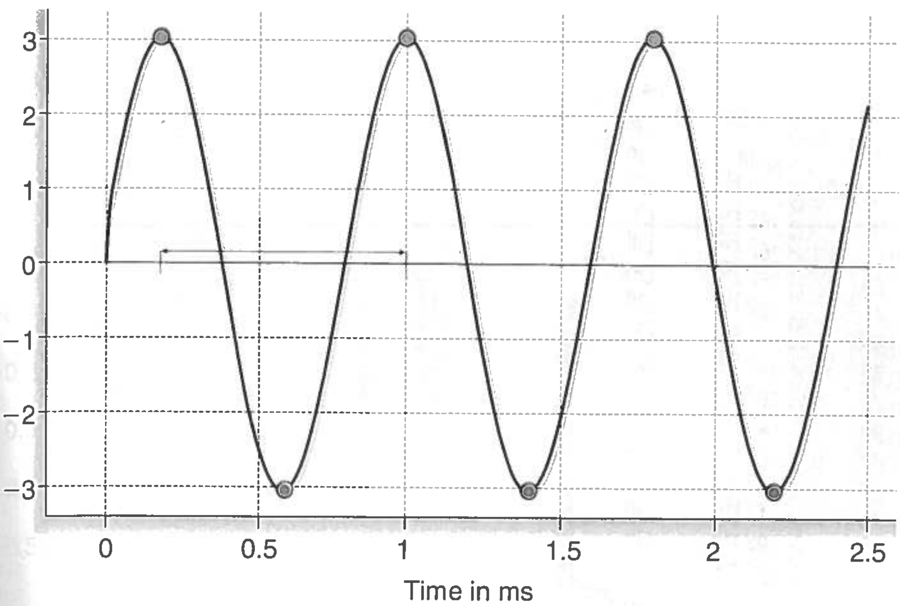


Figure 2.27 Plot of a sinusoidal function.

Infinity Project Experiment: Measuring a Tuning Fork

A tuning fork makes a sound that is nearly sinusoidal. The frequency of this sinusoid is determined by the tuning fork's size and physical makeup. By measuring the signal of a particular tuning fork's sound, we can determine the period of that sound. What is the frequency of the tuning fork? How close to sinusoidal is its signal? What is the signal's amplitude? If we know both the frequency and amplitude of the tuning fork's signal, we can re-create the signal by using a sinusoidal-function generator. Can you use such a generator to produce a sinusoid that sounds like the one produced by the tuning fork?

ues makes it easy to find the same point in the next period. The signal starts at 0 when $t = 0$ and has its first maximum value at $T/4 = 0.2$ s. Comparing this signal to the values in Table 2.3, we see that it has the pattern of a sine. Our final expression is

$$s(t) = A\sin(2\pi f t) = 3\sin(2\pi \times 1.25t)$$

Making Melodies with Sinusoids

A melody is a sequence of notes. Our first digital-band design will use simple melodies with only one note played at any one time. Each individual note can be played by making a periodic signal whose fundamental frequency corresponds to the pitch of the note. Therefore, all we have to do to make music is to play these periodic signals, one after the other, in correspondence with the notes to be played. Let's consider a simple example.

Figure 2.28 shows a block diagram for the creation of the first few notes from a familiar children's song, starting with the input of the information and ending with the production of sound. Suppose this piece of music has already been converted into a MIDI file and saved. The MIDI-file information shows the notes being played as well as their frequencies, which can be figured out using Figure 2.17.

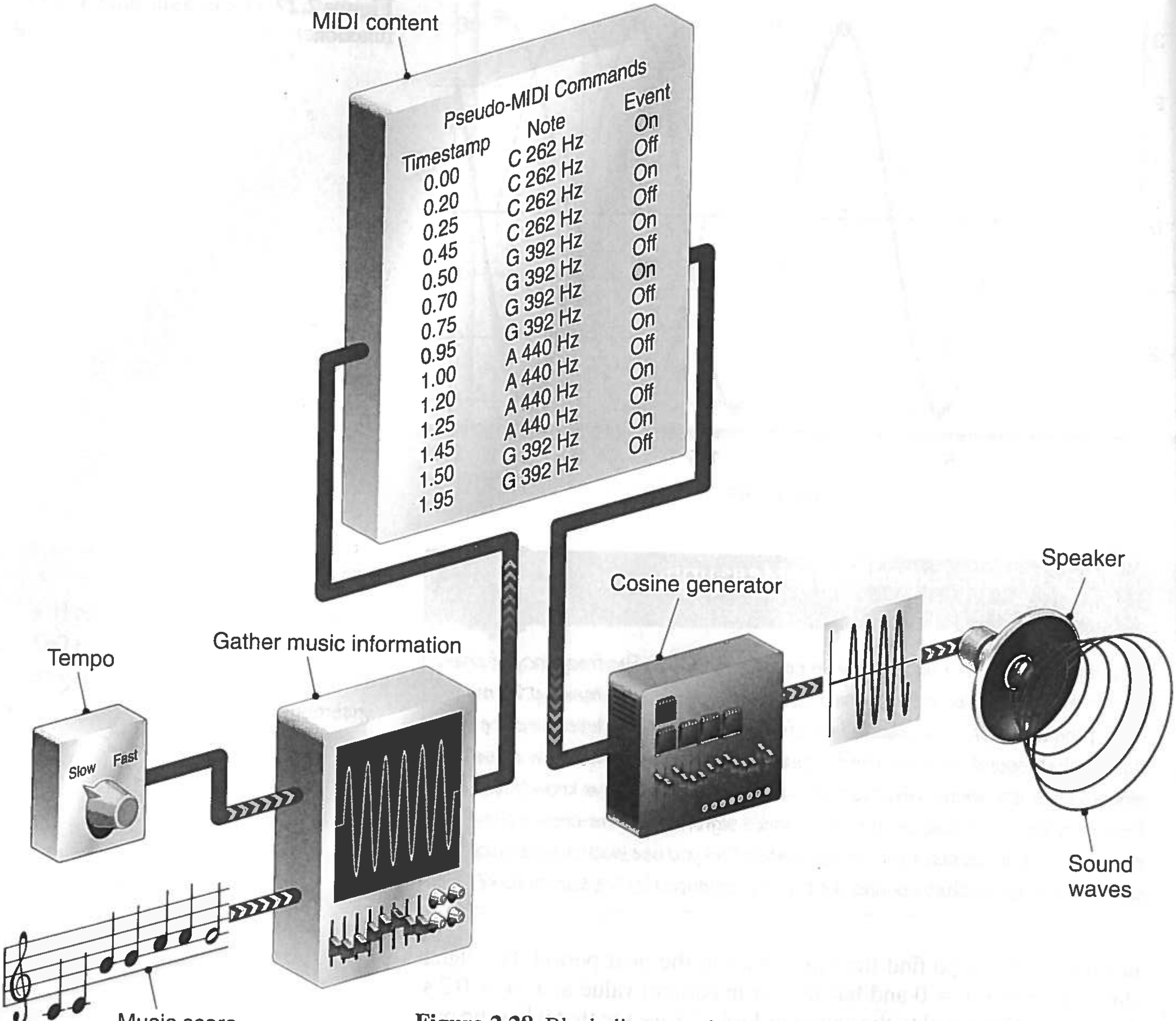


Figure 2.28 Block diagram of a system to create a familiar children's song from a MIDI file.

The cosine generator will compute the values of a cosine function at the specified frequency and will turn each cosine function on and off according to the time information in the MIDI file. Each cosine has the form of Equation (2.11), with amplitude $A = 1$ and frequency f determined by the MIDI file. A very short segment typical of part of the music signal is also shown in Figure 2.28. The sequence of cosine waves generated at the selected frequencies and times is our music signal that will be connected as the input to the "Convert to sound" block. This last block makes sounds from the signal, using a device that you probably already know about—a loudspeaker.

The short melody in Figure 2.28 should last for almost 2 seconds (s). Since the periods of our sinusoidal notes are between 2.3 ms and 3.8 ms, we cannot look at the entire music signal in one plot and expect to see the detail of the individual periods of the sinusoids. In Figure 2.29, we have

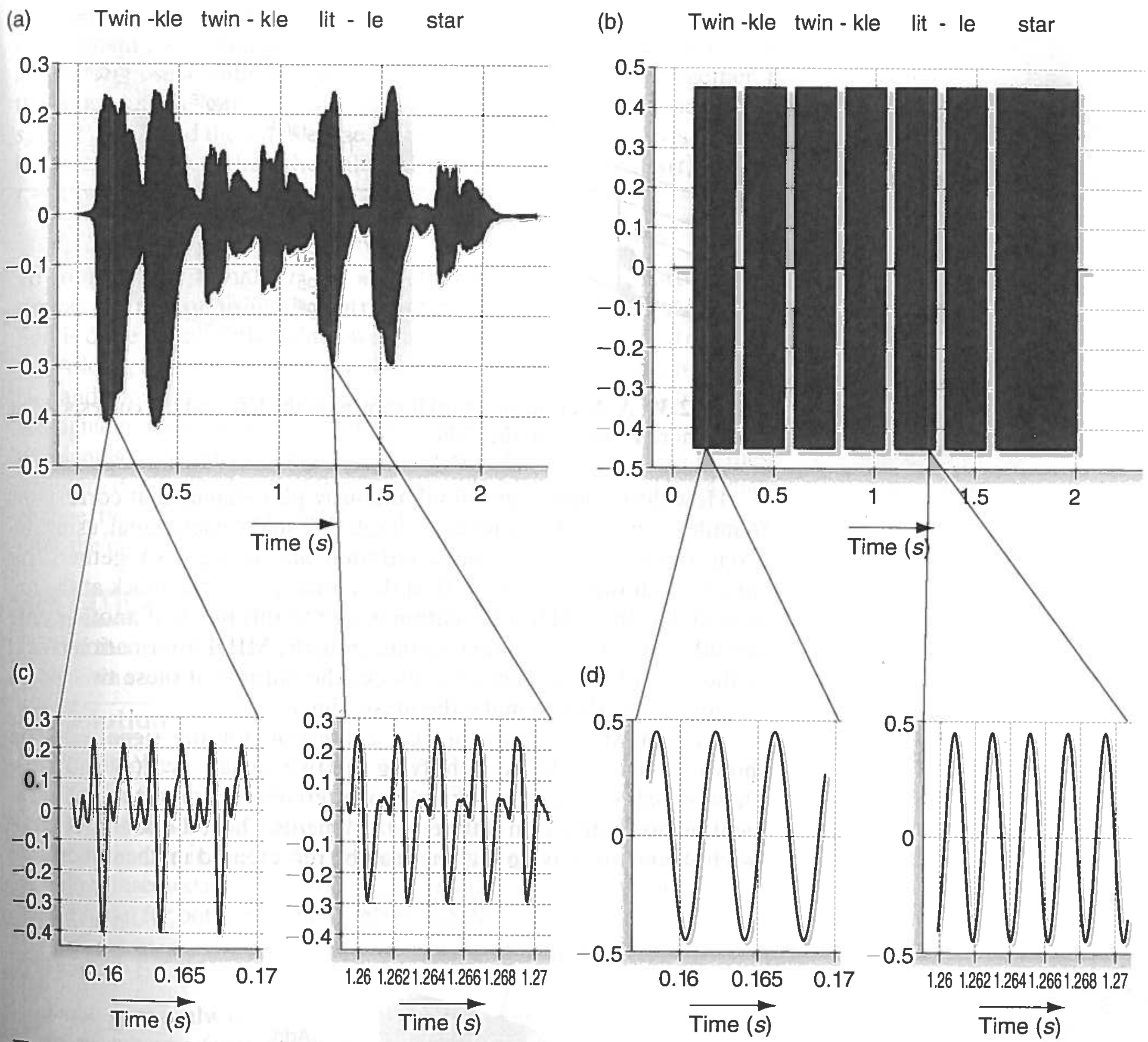


Figure 2.29 The plot of the signal for a saxophone playing a popular children's song in (a) can be compared with the version from our sinusoidal-signal generator in (b). Plots of short intervals of two notes from each (highlighted in red) are shown in (c) and (d), respectively.

taken a few periods out of each note and plotted the notes in sequence to give you an idea of what the music signal looks like. If the whole signal were plotted at a scale of 20 ms per inch, it would be over 8 feet long.

Making Music with More than One Note at a Time

Most interesting music is made up of more than just simple melodies. Usually, two or more notes should be played together, often by more than one instrument at a time, as in the chords of Figure 2.18. Figure 2.30 is an example of a simple chord and its MIDI file description.

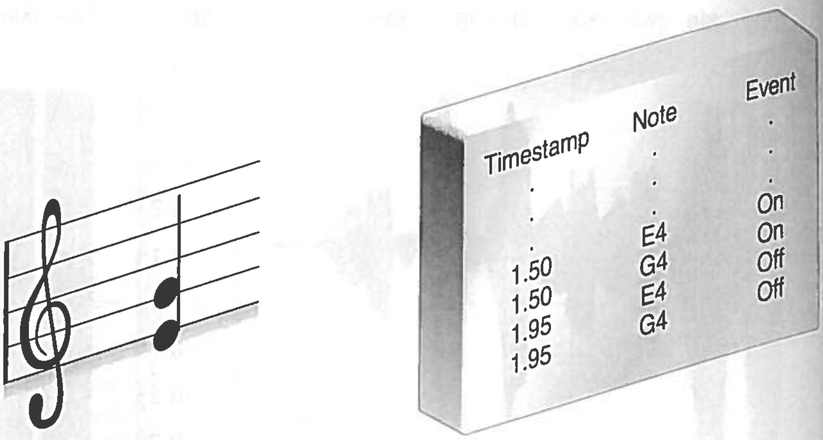


Figure 2.30 A sheet-music chord is shown on the left, and the corresponding MIDI chord is shown on the right.

How do we make and simultaneously play signals that correspond to multiple notes? The answer is simple: We make each signal, using different cosine-generator blocks, and then *add* the signals together. This process is shown in Figure 2.31. If the cosine-generator block at the top is available, the MIDI information is sent to this block. If another note should start before the previous one ends, the MIDI information is sent to the second cosine-generator block. The outputs of these two blocks are added together to make the music signal.

We can write a mathematical expression for the signal addition shown in Figure 2.31 by identifying the two signals as $s_a(t)$ and $s_b(t)$. These signals can be signals from two different notes on the same instrument or two notes from different instruments. Then, the sound of these two instruments playing together can be represented mathematically as

$$s(t) = s_a(t) + s_b(t)$$

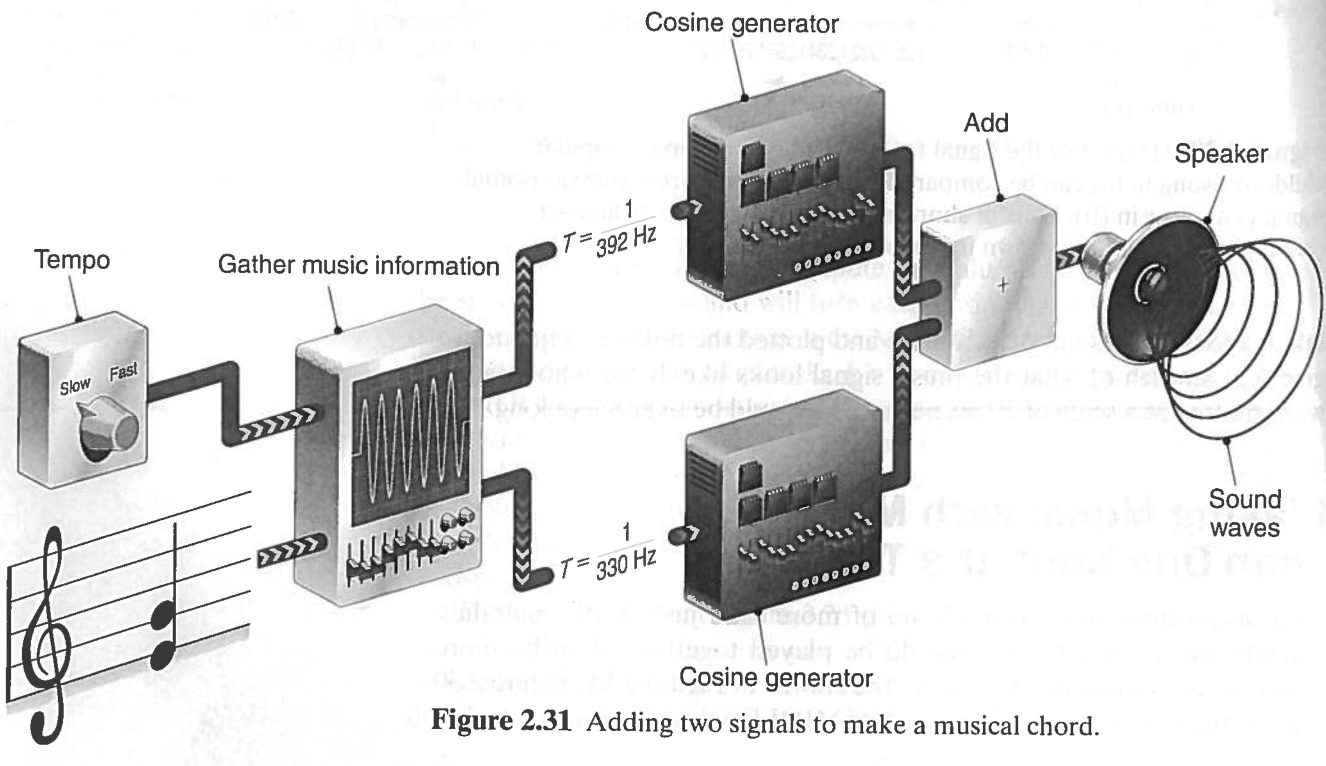


Figure 2.31 Adding two signals to make a musical chord.

The addition is done by finding the value of each signal at a specific time instant $t = t_0$ and adding the values of the signals together. For example, suppose a guitar sound and a singing voice are added together. If the value of the guitar sound signal at time $t = 1.5$ s is found to be $s_a(1.5) = 0.7$ and the value of the voice signal at time $t = 1.5$ s is found to be $s_b(1.5) = -0.4$, then the value of the combined signal $s(t)$ at time $t = 1.5$ s is

$$s(1.5) = s_a(1.5) + s_b(1.5) = 0.7 + (-0.4) = 0.3$$

To compute the signal $s(t)$ for any other time instant, we would do a similar calculation using the corresponding numbers. When this operation is done for all time values, we have the complete signal $s(t)$.

Adding two signals over a time interval to create a plot can be done by hand for simple signals. In such cases, significant points, such as when the signal is at a maximum, a minimum, or zero, are identified for each of the signals to be added. Then the values of both signals at each of these times are determined so that they can be added. When a signal has a sharp change, then the value of the sum can be computed just before and just after the change.

EXAMPLE 2.5 Adding Two Signals Together

Add the two signals in Figure 2.32(a) and (c) together, and draw the resulting signal.

Solution

We will solve this problem in two ways. The first method takes advantage of the particularly simple form of $s_1(t)$. The second way, using a table, will work for all sums of signals. The two signals to be added are plotted on the same axes in Figure 2.32(b), and the sum is plotted in (d). If one signal represented a clarinet and the other a flute, the sum would represent the signal for both instruments played together.

Method 1:

Since $s_1(t)$ is always $+1$ or -1, we will divide the time axis into segments where $s_1(t)$ is not changing and then plot over those segments.

Time Interval	$s_1(t)$	$s_2(t)$	For $s(t) = s_1(t) + s_2(t)$, Plot
$t = 0$ to $t = 0.5$	1	$s_2(t)$	$s(t) = 1 + s_2(t)$
$t = 0.5$ to $t = 1.0$	-1	$s_2(t)$	$s(t) = -1 + s_2(t)$
$t = 1.0$ to $t = 1.5$	1	$s_2(t)$	$s(t) = 1 + s_2(t)$
$t = 1.5$ to $t = 2.0$	-1	$s_2(t)$	$s(t) = -1 + s_2(t)$

Method 2:

This method works for all types of signals. To start, make a table of values for the two functions, and add corresponding pairs of values.

t	0.05	0.25	0.45	0.55	0.75	0.95	1.05	1.25	1.45	1.55	1.75	1.95
$s_1(t)$	1.0	1.0	1.0	−1.0	−1.0	−1.0	1.0	1.0	1.0	−1.0	−1.0	−1.0
$s_2(t)$	−1.8	−1.0	−0.2	0.2	1.0	1.8	1.8	1.0	0.2	−0.2	−1.0	−1.8
$s(t) = s_1(t) + s_2(t)$	−0.8	0.0	0.8	−0.8	0.0	0.8	2.8	2.0	1.2	−1.2	−2.0	−2.8

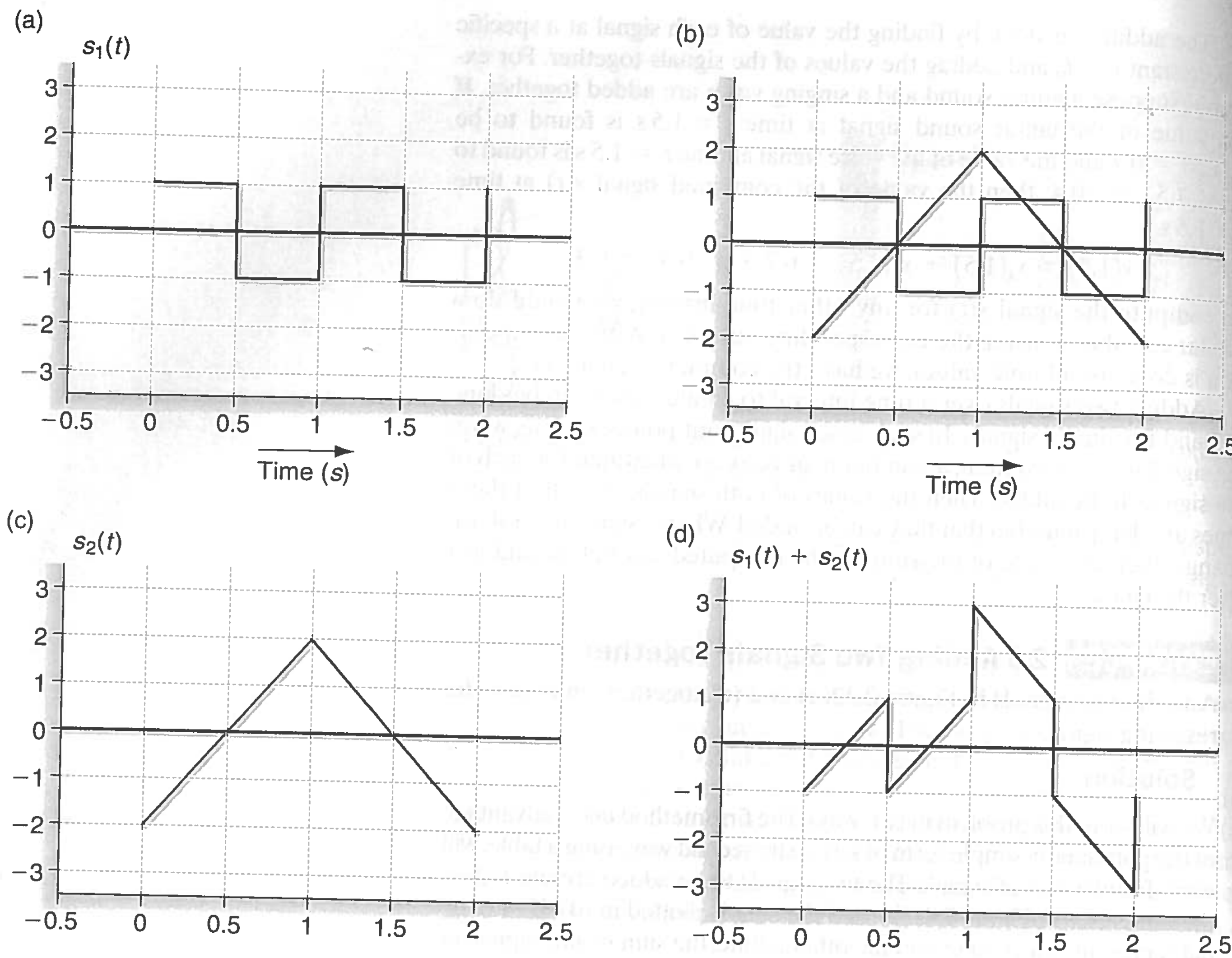


Figure 2.32 Sum of two plots of signals. The individual signals are shown in (a) and (c), the two signals are plotted on the same axis in (b), and the sum is shown in (d).

Once we have the sums of the pairs, we can sketch out the desired signal. The points used in the table are shown in both the individual signals and the sum in Figure 2.32(d).

The results using Method 1 and Method 2 are exactly the same.

Adding Two Sinusoids Together The first design of our digital band used sinusoids for the periodic functions. If we want the band to play two different notes together, we sum two sinusoids together, where each sinusoid corresponds to each note. Consider the sum of two cosines with the same amplitude $A = 1$, but different fundamental frequencies f_1 and f_2, so that

$$s(t) = \cos(2\pi f_1 t) + \cos(2\pi f_2 t)$$

In this sum, the values of f_1 and f_2 could correspond to two notes that are being played by an instrument, such as a note we'll call concert A (440 Hz) and another note we'll call the F below concert A (349.23 Hz).

These two cosine functions could be added together using the table method from the previous example, although this operation is more easily done with a calculator. Figure 2.33 shows the two individual signals and the resulting sum $s(t)$. In general, it is hard to guess the appearance of the plot of the sum of two sinusoids. In most cases, the sum is not periodic, because the individual signals will never start over again at the same time. But in this case, the ratio of the frequencies is approximately 4 to 5, and the sum will repeat almost perfectly after only a few periods of each signal. When we hear sounds with this characteristic, they usually sound good to us.

INTERESTING FACT:

Why do the sounds of certain pairs of notes sound good when they are played together? The ratio of the periods of the sounds is the key to the answer. If this ratio contains two small integers, such as $\frac{3}{2}$, $\frac{4}{3}$, and $\frac{5}{3}$, the sound signal produced is periodic and has a pleasing sound. This chordal or vertical structure of music (as opposed to melodic or horizontal structure) is known as harmony.

Reverse Engineering the Score

Up to this point, we've been focused on building a system that takes a description of music and turns it into sound. Suppose we already have a piece recorded by somebody else that we want our digital band to play.

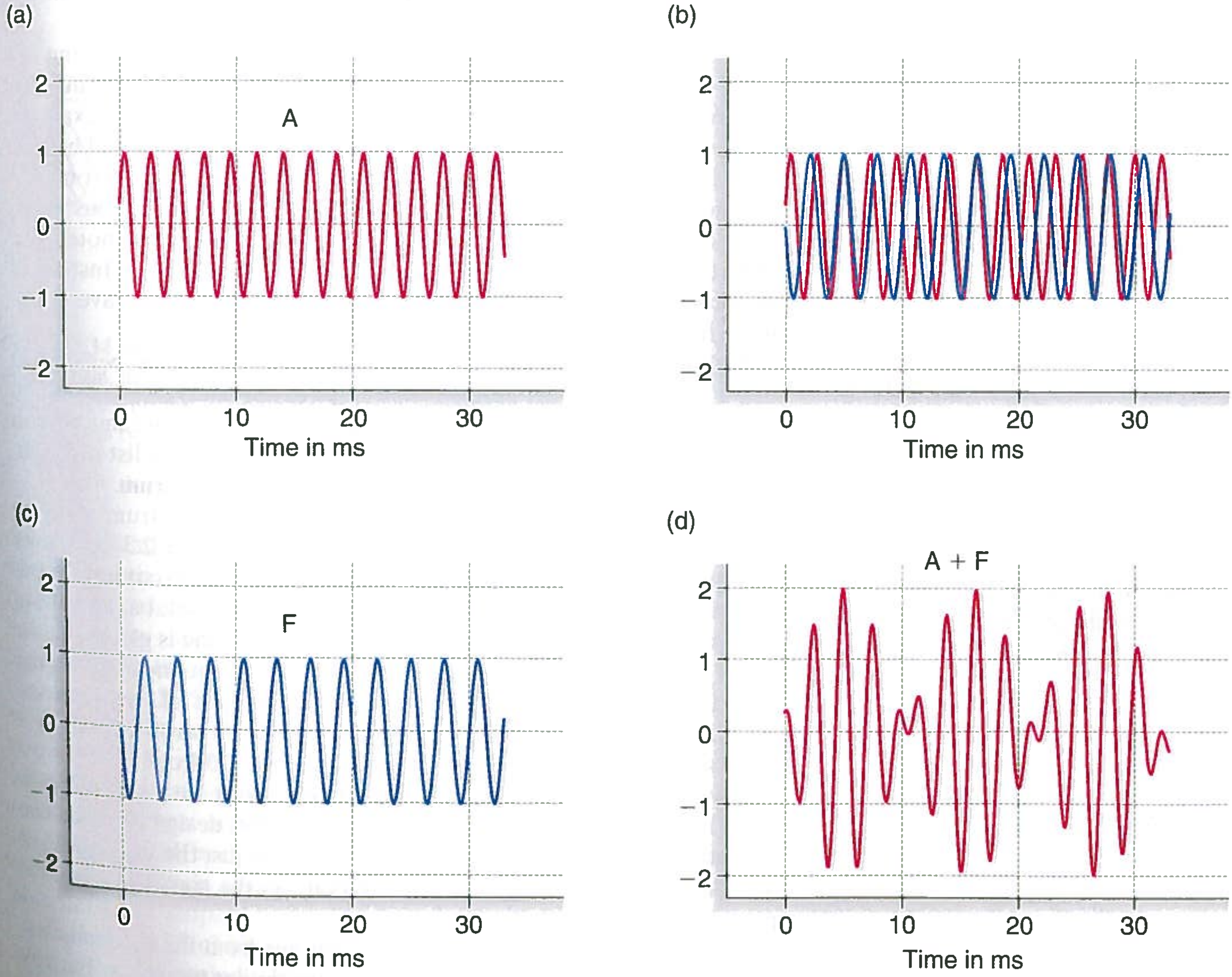


Figure 2.33 Sum of two sinusoidal signals at the frequencies of A and F. The individual signals are shown in (a) and (c), the two signals are plotted on the same axis in (b), and the sum is shown in (d). The sum is not a sinusoidal signal.

Infinity Project Experiment: Building the Sinusoidal MIDI Player

This experiment allows you to construct your own virtual keyboard inside of a computer, using sinusoidal-function blocks, adder blocks, and a MIDI message-reader block. Once you've built your synthesizer, go to the Web and get a MIDI file of your favorite piece of music by your favorite artist and listen to it. How does your favorite piece of music sound like when played using sinusoids? What do the signals look like as the piece is playing?

Is there some way to create the score or the MIDI file from the sound of the music itself? If so, there would be many things that we could do with such a system. For example, we could use the system to translate all the pieces we like into MIDI files automatically without our help. We could also study the score that the system produces to learn how a particular piece was made.

It might seem like we could design such a system by redrawing our digital-band block diagram to reverse all the inputs and outputs. The functions of all the blocks, however, would have to change. For example, the speaker that converts an electronic signal into sound would have to be replaced by a microphone that converts sound into an electronic signal. The block that converts notes into frequencies and durations would have to convert frequency and duration values back into notes. The biggest change in the system would be in the middle block. Instead of producing periodic signals from frequency values, it would have to calculate frequency values from the periodic signals it is fed.

Spectrum Analyzer: A device that determines the amplitudes and frequencies of the sinusoids in any signal.

Spectrum: A list of the amplitudes and frequencies of all the sinusoids in a signal.

Engineers have come up with an important tool for performing such a calculation. This tool is called a **spectrum analyzer**. A spectrum analyzer determines the amplitudes and frequencies of the sinusoids in any signal through a set of mathematical operations. The list of amplitudes and frequencies produced is called a signal's **spectrum**.

Figure 2.34(a) shows the block diagram for a spectrum analyzer block when the input signal is just one sinusoid. Figure 2.34(b) shows the same block diagram when the input is the sum of two sinusoids. The output of the spectrum analysis is plotted on a horizontal axis showing the range of possible values of frequency. A vertical line is plotted on the axis at the value of each of the input sinusoids' frequencies. Since the input in Figure 2.34(b) has two cosines, the output of the spectrum analysis is two vertical lines. The complete block diagram for creating a music score from music sound is shown in Figure 2.34(c).

Most spectrum analyzers are designed to look at a signal over a short time interval and calculate its spectrum. This design is a good one for analyzing music, because the spectrum shows just the frequencies of the notes being played during the interval. As the signal changes, the spectrum changes with it.

It is also possible to accumulate the outputs from the spectrum analyzer and make a graphic display that is very similar to a score. The axis of possible frequencies is horizontal for the spectrum analyzer and vertical for the score's treble and bass clefs, so we will have to rotate the spectrum and make it a vertical strip to match the score. When we accumulate

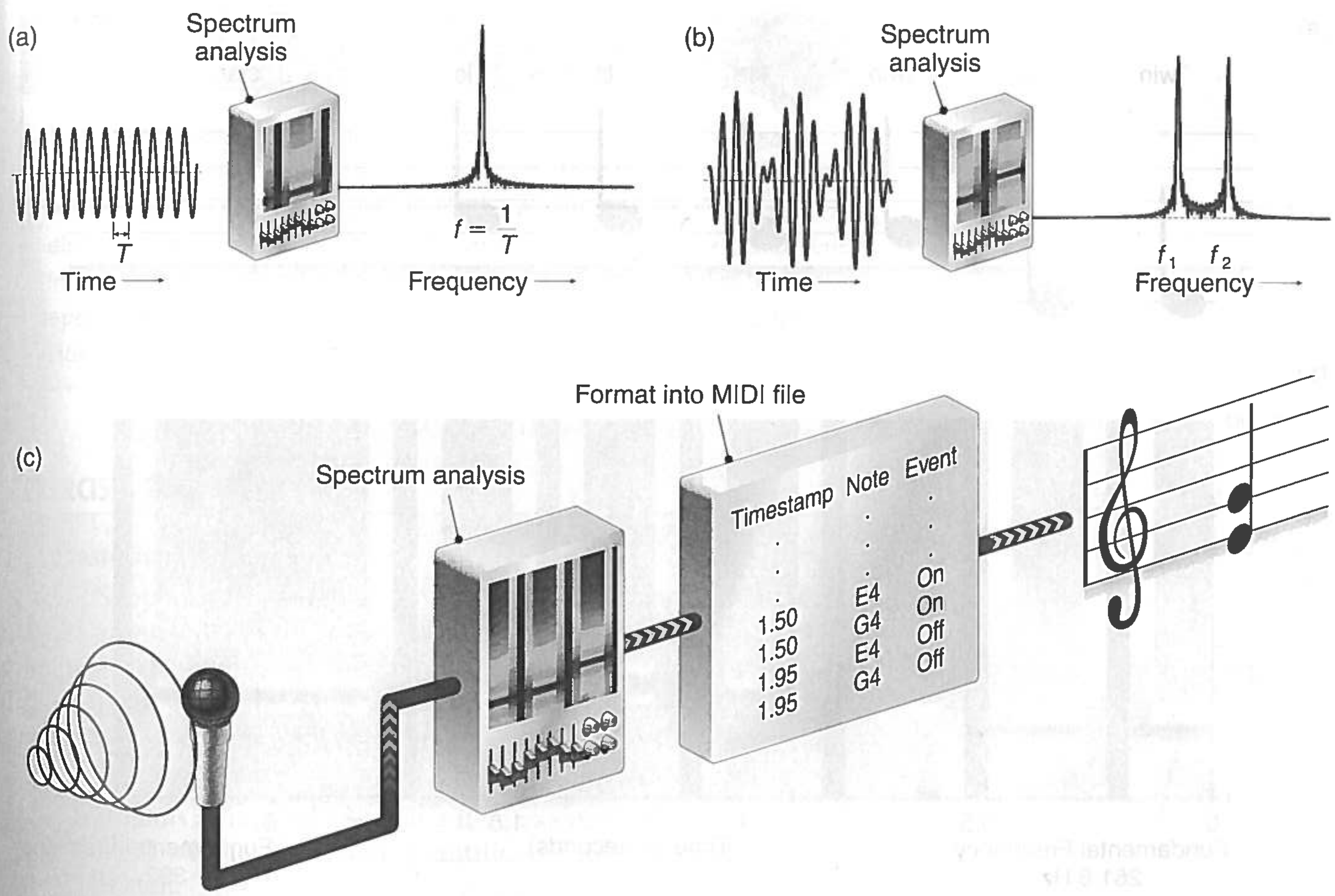


Figure 2.34 Block diagram showing conversion of music sound to MIDI information, using a spectrum-analyzer block. In (a) a signal with a single sinusoid is identified as one note. In (b) a signal with two sinusoids is identified as two notes and the frequency of each is shown.

these strips and place them side by side, we will have lines that show us what notes were being played. This graphic display is called a **spectrogram**. It will not be as stylized as a score, but it will show the frequencies of the notes as they are played and when each began and ended. In Figure 2.35(a), the score for the children's song is shown again for reference. In Figure 2.35(b), we can see the output of the spectrogram of the music signal created by our device, using sinusoidal signals. The dark red horizontal bars show the frequency and duration of each of the notes. Color is used to show the amplitude of each sinusoid with dark red for the highest or loudest values and dark blue for the lowest values. The vertical bars that are completely dark blue represent the quiet time between notes, and the vertical yellow bars show when a note starts and stops.

Spectrogram: A two-dimensional image describing the spectrum of a sound over time.

In Figure 2.35(c) we see the spectrogram of the same sequence of notes played by an instrument with a more complex sound. The periodic function generated by this instrument is the sum of a sinusoidal signal at the fundamental frequency of the note and several other sinusoids which are called harmonics.

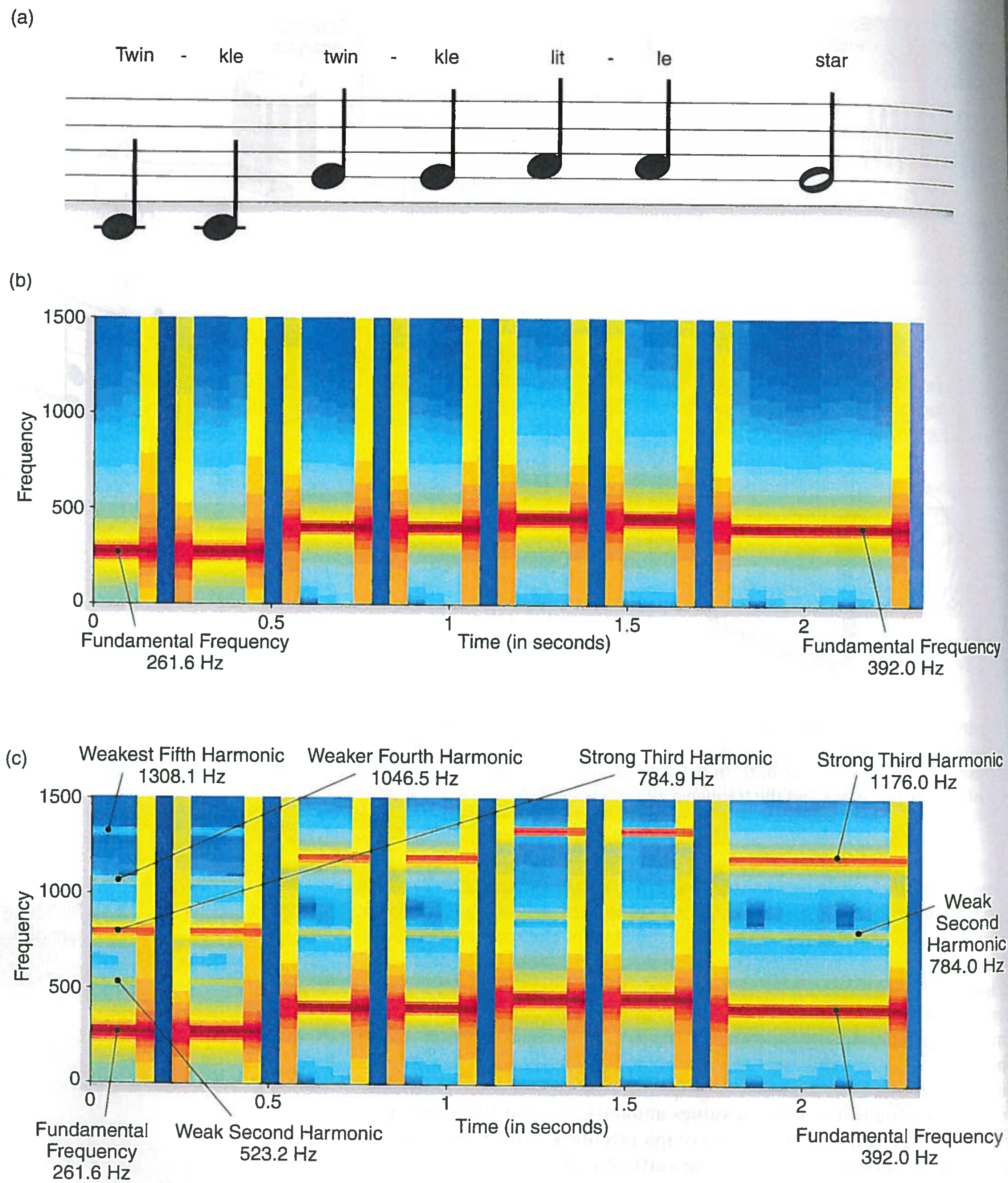


Figure 2.35 The music shown in (a) produces the spectrogram in (b) when played by our sinusoidal music generator. The dark-red horizontal bars show the frequency of the note played at each time, and the blue background corresponds to the quiet intervals between notes. The spectrogram also displays a yellow vertical stripe at the beginning and end of each note. An instrument with a more complex sound produces the spectrogram in (c). It has a weak second harmonic, a stronger third harmonic, and much weaker fourth and fifth harmonics that can be seen only for the first two notes.

Infinity Project Experiment: The Spectrogram

The spectrogram is a useful tool for understanding the frequency structure of any sound—not just music. The spectrogram can be applied to any signal, not just music. Examine the spectra of several simple sounds, and write down the amplitudes, frequencies, and time instants of the sinusoids that are contained in them. What does the spectrogram of your voice look like? Does it have a simple description? How about that of a recording from your favorite performer?

EXERCISES 2.3

Mastering the Concepts

1. Suppose you want to create the sound of a piano digitally. Can you use a MIDI file to create an accurate piano sound? Why or why not?
2. How are songs described using MIDI information? If you were to open up a MIDI file and read its instructions, what would you see?
3. What is a sinusoid? Sketch out an example of a sinusoid, and label all the important quantities of the sinusoid on your sketch.
4. How is the sine function related to a circle with a radius of one? How is the cosine function related to this circle?
5. What's the difference between radians and degrees? Which measure of angle is most often used by engineers and mathematicians?
6. What is the frequency of a sinusoid? How do we change a sinusoid's frequency?
7. What is the amplitude of a sinusoid? How do we hear the effect of changing the amplitude of a sinusoid?
8. Suppose we want to play the sounds of several sinusoids at the same time. How do we do that?
9. Engineers have come up with a way to calculate the frequencies of the sinusoids in a piece of music and display them in a plot. What is this type of plot called? Describe how the frequency information is contained in the plot.

Try This

10. A piece of music contains the following note sequence: E4, D4, C4, D4, E4, E4, E4. Each note lasts 0.25 second, and no two notes are played at the same time. Write the list of instructions that would be contained in a MIDI file that describes this piece of music.

11. Convert the given angle values to Cartesian coordinates, using $(x, y) = (\cos(\theta), \sin(\theta))$. Then sketch the (x, y) points on a graph.
    a. $\theta = 15°$
    b. $\theta = \pi/3$ radians
    c. $\theta = 4\pi/5$ radians
    d. $\theta = 4\pi$ radians
12. Plot the following sine and cosine signals on a graph over 7 ms.
    a. $s(t) = 2\cos(2\pi \times 675t)$
    b. $s(t) = 3\sin(2\pi \times 432t)$
    c. $s(t) = 6\cos(2\pi \times 500(t - 0.00025))$
13. Look at the plots of the sinusoids you drew in Exercise 2.3.12. From those plots, calculate following:
    a. the amplitude of each sinusoid
    b. the period of each sinusoid
    c. the frequency of each sinusoid
14. Plot the sum of $s_1(t)$ from Figure 2.32 and $s_2(t) = 2\sin(\pi t)$ for 2 seconds (s).
15. Sketch out a picture corresponding to the spectrogram for the piece of music in Exercise 2.3.10 if the frequencies of C4, D4, and E4 are 261.6 Hz, 293.7 Hz, and 311.1 Hz, respectively. Carefully label the axes of your plots.
16. Sketch the outputs of a spectrum analyzer that is analyzing the following signals:
    a. $s(t) = 2\cos(2\pi \times 675t) + 3\cos(2\pi \times 544t)$
    b. $s(t) = 1.3\sin(2\pi \times 182t) + 2.4\cos(2\pi \times 215t)$

In the Laboratory

17. A tuning fork makes a signal that is almost exactly a sinusoid. How close are other signals to sinusoids? Have a musician come into the lab and play her or his instrument for you while you measure the signal produced by the sinusoid. Try to make a sinusoid that has the same frequency as the note the musician plays. How similar do the two signals sound? How different do the two signals look?

Back of the Envelope

18. Would it make sense for a MIDI file to have out-of-order timestamps for its notes? Why or why not?
19. Figure out how many note entries a typical MIDI file might have. To do so, suppose a MIDI file describes a 3-minute piece of music and that no more than five instruments are being played at any one time. Also, suppose each instrument plays an average of five different notes each second. What is the largest number of note entries that the MIDI file will have? Remember, it takes two entries to create each note (one to turn the note on and one to turn the note off).

2.4 Improving the Design—Making Different Instruments

Instrument Synthesis

We like variety in music. Just look at all of the different instruments used in making music today—trumpets, violins, snare drums, piccolos, bassoons, guitars, etc.—and we haven't even considered the different styles of music, such as rock, R&B, jazz, country, and so on. The lists seem endless. At this point, our digital band can play only sinusoidal signals. With these signals, we can hear the melody and chords very clearly, but the sound it creates is often described as "hollow," or worse yet, boring. If this type of music were all we listened to, the music world wouldn't be nearly as rich and textured as it actually is. We need to improve our making-music block in our digital band so that it produces real-world instrument sounds, as well as the sounds of new instruments that have never existed. But how can we do this?

To help us figure out a better making-music block, we will first study what makes instrument sounds different from sines and cosines. You will recall that we chose sine and cosine signals because they were simple. Unfortunately, very few instruments make a signal that is an exact sinusoid, even if we choose the period of the sinusoid to be the same as that of the instrument's sound. For example, the guitar sound in Figure 2.12 is not exactly sinusoidal, nor are the two piano sounds shown in Figure 2.15. Another important feature of an instrument's sound is the way its loudness changes over time. For example, each of the two piano sounds at the top of Figure 2.15 has a large amplitude at the start of the note, and the sound gradually gets softer over time. We will have to take these signal properties into account if we want to make high-quality instrument sounds.

There are several methods we might explore to create realistic instrument sounds. One simple approach might be to record all the instruments we want to use. We would have to record each instrument playing every possible note. Then we would combine signals we've recorded from each instrument to make the music we want to hear. From an engineering perspective, this approach has several problems. First, it would require storing all of those sounds and then finding them when they were needed. This method also wouldn't enable us to make new sounds that weren't combinations of our saved sounds. A better approach is to find a way to actually create these sounds as we need them.

Synthesis is the creation of useful and complicated items from more basic ones. **Sound synthesis**, then, is the creation of useful and complicated sounds from more basic sounds. In the next section, we will focus on two of the most popular techniques for creating periodic waveforms that do a very good job of matching the characteristics of instrument sounds. We will also figure out ways to control how these sounds turn on and off so that they more closely mimic the sounds of instruments. Combining these two ideas will help us make a much better digital band.

Synthesis: The creation of useful and complicated items from more basic ones.

Sound Synthesis: The creation of useful and complicated sounds from more basic sounds.

Waveform Synthesis: A synthesis technique that stores a characteristic signal of an instrument for one specific note and then shrinks or stretches the signal in time to play several different notes with different fundamental frequencies.

Additive Synthesis: A synthesis technique that creates a periodic signal by adding sinusoids together.

INTERESTING FACT:

Besides wavetable and additive synthesis, engineers have a third way of making a realistic instrument sound. Called *physical modeling synthesis*, the technique uses a mathematical model of the physics of an instrument to compute the instrument's sound. This synthesis method is the most sophisticated and the newest. When it is properly implemented, this method can provide extremely realistic sounds.

Making Periodic Signals

The two most common digital techniques for making periodic signals with a specific shape are called **waveform synthesis** and **additive synthesis**:

1. *Waveform synthesis,* or *wavetable synthesis,* stores a characteristic sound signal of one specific note for the actual instrument being re-created and then shrinks or stretches it in time to play different notes with different fundamental frequencies. While this synthesis method is the more straightforward of the two methods, it does use significant amounts of memory to store the characteristic sound of each instrument.
2. *Additive synthesis* (also known as *Fourier synthesis*) uses the sinusoidal functions we've put in our digital band as building blocks to create sounds. Each characteristic sound is created by starting with a cosine signal whose period matches that of the note we want to play. We then improve upon this signal by adding more sinusoids at higher frequencies so that the shape of the signal follows that of the desired sound more closely. This synthesis method was used in most early music synthesizers, although it is less popular in modern-day synthesizers.

Waveform Synthesis Suppose that we know the characteristic signal $p_A(t)$ of an instrument for one note with a fundamental frequency f_A and a corresponding period $T_A = 1/f_A$. We would like to create that same instrument sound for a new note with a fundamental frequency f_{new}.

One way to change the period T in a periodic signal is by time scaling the signal as described in Equation (2.3). With time scaling, a signal is "stretched" or "compressed" by expanding or shrinking, respectively, the period of the sound without changing the signal's overall shape. Then the new signal $p_{\text{new}}(t)$, with new period T_{new}, can be created from the original periodic signal as follows:

$$p_{\text{new}}(t) = p_A\left(\frac{T_A}{T_{\text{new}}}t\right) = p_A\left(\frac{f_{\text{new}}}{f_A}t\right) \tag{2.13}$$

Either the ratio of the periods or the ratio of the frequencies can be used to scale the time variable in this calculation. A block diagram for waveform synthesis using time scaling is shown in Figure 2.36.

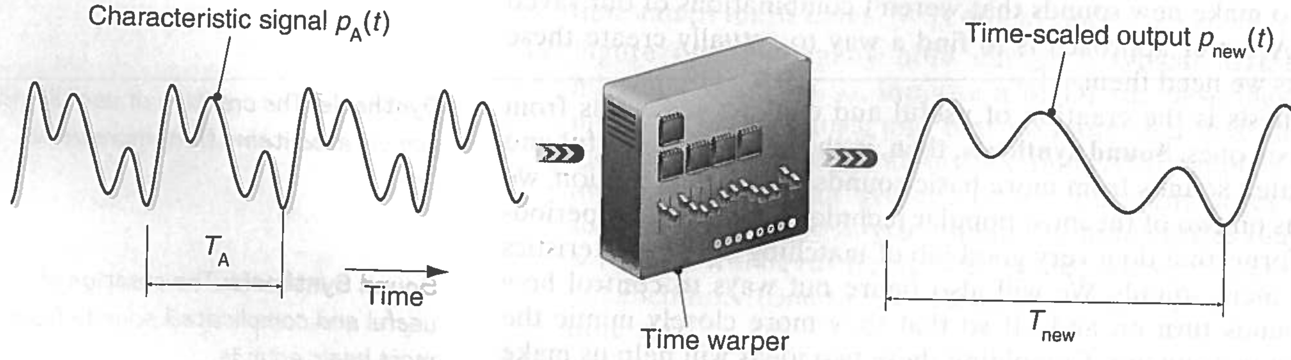


Figure 2.36 Block diagram for time scaling of signals.

EXAMPLE 2.6 Using Time Scaling to Create Mathematical Expressions

Figure 2.37(a) shows the characteristic signal $p_{250}(t)$ for a new instrument being played at a fundamental frequency of 250 Hz over 8 ms. Use time scaling to create mathematical expressions for signals from this instrument when it plays notes with fundamental frequencies of 300 Hz and 400 Hz, respectively. Plot the signals for both notes over 8 ms, and plot the 250-Hz signal on each plot. Also, plot the 300-Hz signal over 6.6 ms, and plot the 400-Hz signal over 5 ms. How do these last two plots compare with that of $p_{250}(t)$?

Solution

Using Equation (2.13), we can write the following expressions for the new periodic functions:

$$p_{300}(t) = p_{250}((300/250)t) = p_{250}(1.2t)$$
$$p_{400}(t) = p_{250}((400/250)t) = p_{250}(1.6t)$$

We can check that these two answers are correct by first noting that $p_{250}(4\,\text{ms}) = p_{250}(0)$, because the period of the original signal is $\frac{1}{250}$ Hz $= 4$ ms. Since the 400-Hz signal has a period of 2.5 ms, we must have $p_{400}(2.5)\text{ms} = p_{400}(0)$. Using the expression above, $p_{400}(2.5\,\text{ms}) = p_{250}(1.6 \times 2.5\,\text{ms}) = p_{250}(4\,\text{ms}) = p_{250}(0) = p_{400}(0)$. Therefore, our expression for $p_{400}(t)$ is periodic with a period of 2.5 ms. A similar check can be made for $p_{300}(t)$.

All of the requested plots are shown in Figure 2.37 (b–e). When plotted over the same time scale as that for the original signal, $p_{250}(t)$, the new signals clearly have different periods. All of the signals look similar, however, when plotted over time intervals that are scaled according to their periods in (c) and (e).

Using waveform synthesis to make sound requires a technique for specifying the function $p_A(t)$ with a fundamental frequency f_A. We could develop a database of recorded waveforms and load each one in when it is needed. An even simpler way would be to draw the waveform on the computer, given a hard-copy picture of it. The next Infinity Project Experiment allows us to try our hand at using waveform synthesis to make sounds and to hear the music that it creates.

Additive Synthesis Engineers are always looking for new and useful ways to do things. Waveform synthesis is just one way to realistically re-create the sounds of instruments. Are there other ways? In particular, is there any way we can use the sinusoid as a simple building block for creating interesting sounds?

A mathematical technique called *additive synthesis* is another way to create periodic signals that closely match those of instruments. Additive synthesis is built on the following fundamental concept:

The shape of any periodic signal can be approximated to any desired accuracy by adding together enough different sines and cosines with the right amplitudes, frequencies, and time delays.

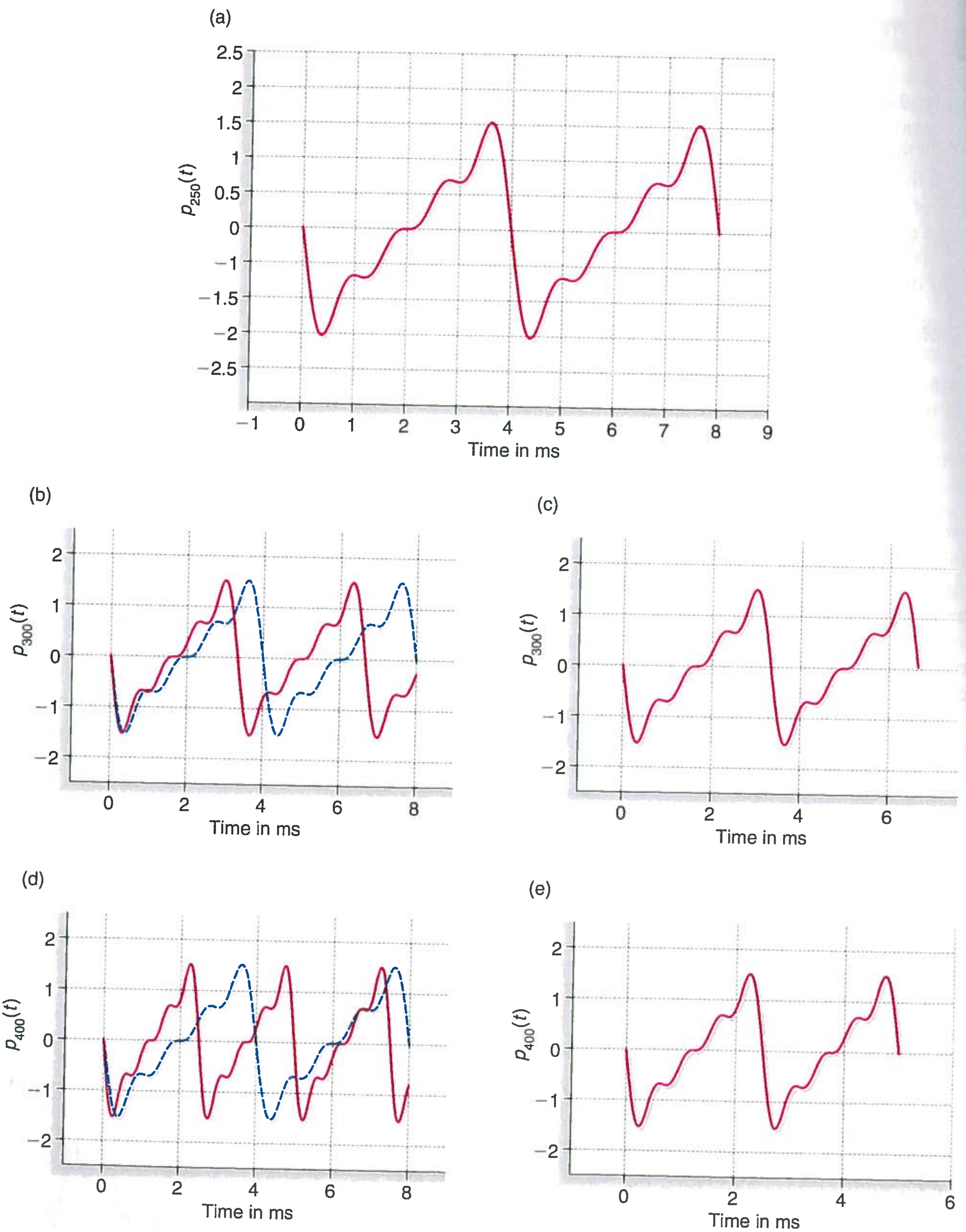


Figure 2.37 Two new periodic signals, $p_{300}(t)$ and $p_{400}(t)$, are created from $p_{250}(t)$ by scaling the time variable t. In (b) and (d), the two new signals are plotted for 8 ms on the same axes as the original $p_{250}(t)$, which is shown by the dashed blue plot. The effects of the time scaling are easily noticed. In (c) and (e), each of the new signals is plotted for two periods, which are 6.6 ms and 5.0 ms, respectively. Since just the time scales are different, the new signals look exactly like the original signal $p_{250}(t)$ plotted for two periods.

Infinity Project Experiment: SketchWave with MIDI

Sketch Wave is a system that allows you to sketch one period of the function $p(t)$ and hear the resulting sound $s(t)$ while it is being created. Try sketching different periodic parts and listening to the music as the periods of the signals are being changed by the MIDI block. Then try to imitate the waveforms shown in Figures 2.12, and 2.15. Can you re-create the sounds of these instruments to make the music you want to hear?

This idea allows us to construct a periodic signal such as that produced by an instrument, using the same building blocks that we used to make sums of sinusoids in the last section. Instead of using the frequencies, amplitudes, and note on–off commands defined by the score, we can use the frequencies, amplitudes, and time delays to re-create the periodic signal of any instrument. The values of the frequencies and amplitudes for a periodic signal can be calculated using a spectrum analyzer. An example shows how this process works.

EXAMPLE 2.7 Finding Amplitudes and Frequencies of Sinusoids in Periodic Signals

Figure 2.38 shows four simple periodic signals on the left and their corresponding spectra on the right. For each periodic signal, write the list of the amplitudes and frequencies of its sinusoids. Find the period of each signal and identify its fundamental frequency.

Solution

(a) The spectrum analyzer shows an amplitude of 1 at both 200 Hz and 400 Hz. The period is $T = 5$ ms $= 0.005$ s, so the fundamental frequency is 200 Hz.

(b) The spectrum analyzer shows an amplitude of 1 at 400 Hz and an amplitude of 0.5 at 1200 Hz. The period is $T = 2.5$ ms $= 0.0025$ s, so the fundamental frequency is 400 Hz.

(c) The spectrum analyzer shows an amplitude of 1 at 300 Hz, 900 Hz, and 1500 Hz. Three periods take 10 ms, so the period is $T = 3.33$ ms $= 0.00333$ s. The fundamental frequency is 300 Hz.

(d) The spectrum analyzer shows an amplitude of 1 at both 200 Hz and 300 Hz. The period is $T = 10$ ms $= 0.1$ s, so the fundamental frequency is 100 Hz.

Figure 2.39 shows how additive synthesis works. Part (a) shows a plot of a saxophone sound in blue. The signal is periodic with period 3.8 ms and a sinusoid with that period is shown with a dashed red plot. In (b)–(d), we have attempted to reconstruct the shape of the saxophone sound, using sums of 2, 4, and 11 sinusoids, respectively. The sum of sinusoids is shown in blue and the dashed red plot of the true signal is also shown for reference. As we add more sinusoids together, we see that the signal produced by this sum begins to take on the shape of the original

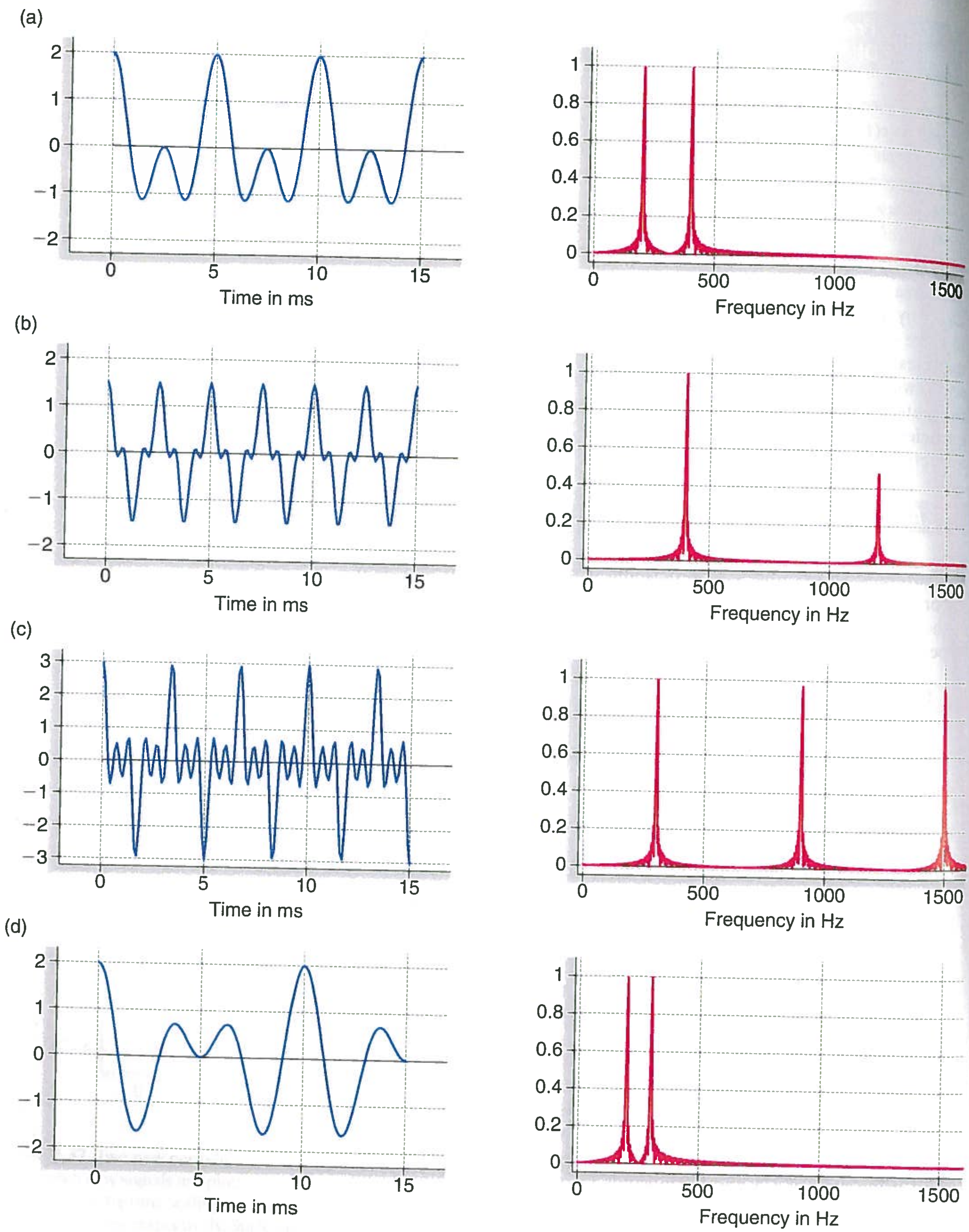


Figure 2.38 Spectrum-analyzer information is shown on the right for the four different periodic signals shown on the left.

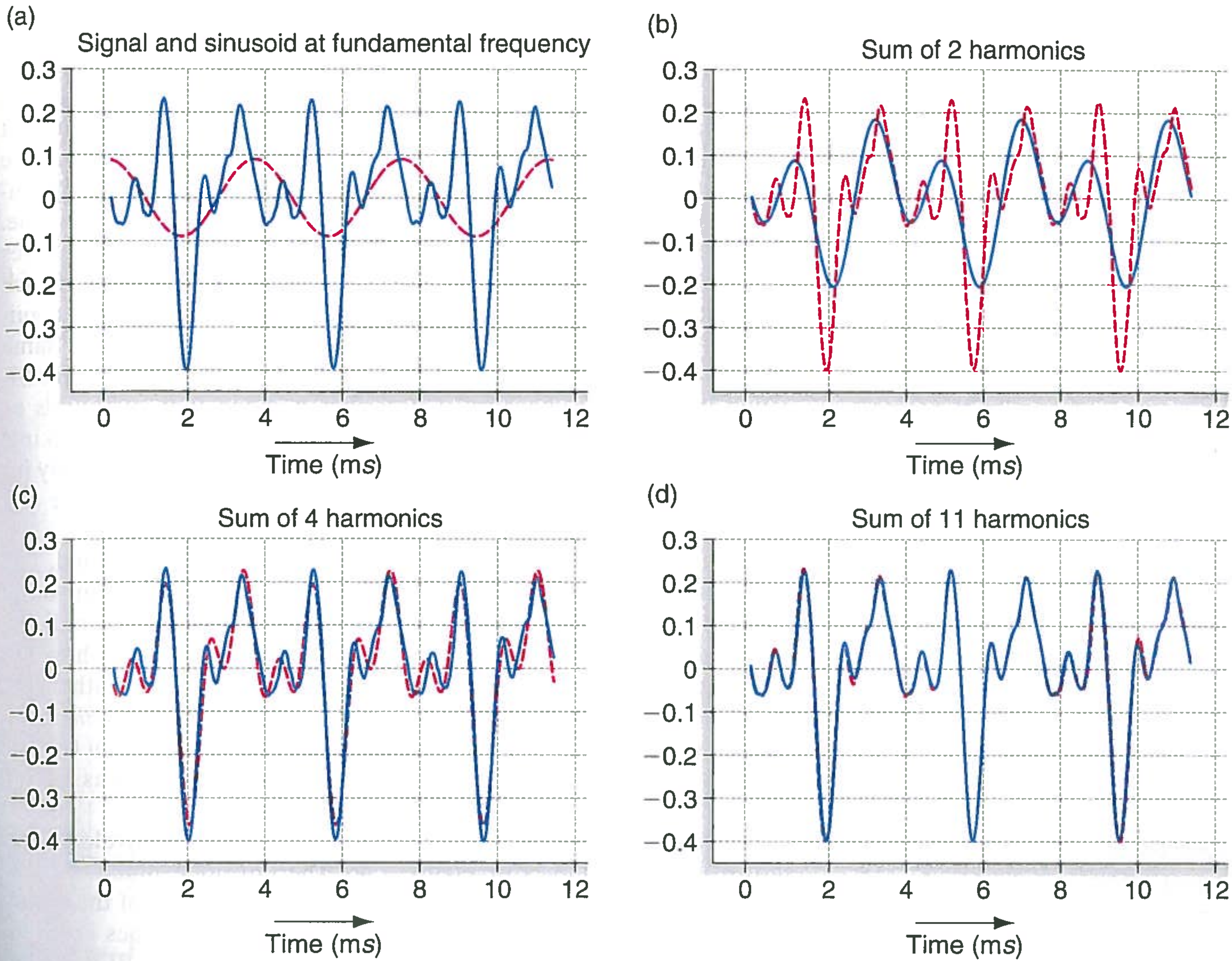


Figure 2.39 In (a), three cycles of a saxophone waveform are shown in blue, with the sinusoid at the fundamental frequency shown in red. In (b), the sum of the first 2 harmonics is shown. The sum of the first 4 harmonics in (c) is much closer to the correct signal shape, and the sum of 11 harmonics in (d) is almost an exact match.

saxophone sound. By adding enough sinusoids together, the synthesized sound becomes so much like the original one that our ears can't tell the difference.

Additive synthesis is a great way to make music because it uses simple blocks—sine generators and adders—in a simple way. Our ability to make sound by using additive synthesis is always getting better, because the number of sine generators and adders we can implement in a single device is always increasing, due to Moore's law. Despite these apparent advantages, additive synthesis is not as popular a method for re-creating audio as waveform synthesis. The reason is that the amplitudes, frequencies, and time delays needed to make really good instrument sounds change from time instant to time instant. Specifying and controlling the values of all of these mathematical quantities is hard. For this reason alone, additive synthesis is not often used in modern electronic musical synthesizers today.

Envelope of a Sound

The sound-synthesis methods described so far go a long way towards making realistic versions of instrument sounds. They leave out one important feature, however. When we look at plots of instrument sounds over time intervals longer than a few periods, we find that the signals do not have the same amplitude all the time. The changing amplitude of these signals is heard as a change of loudness over time. Different instruments have different ways in which their amplitudes change with time. Some instruments are loud at the beginning of each note and then gradually fade away, like the sound of a piano or a guitar. Other instruments have a loudness that is about the same throughout the note being played. This characteristic of a signal is called its **envelope**, because it describes a shape that surrounds or contains the entire signal. The envelope of a signal, roughly speaking, is equivalent to the volume or loudness of the signal over time. By including the envelopes of signals inside of our digital band, we can make more realistic music sounds.

Envelope: A description of a signal's general size or amplitude over many periods.

Figure 2.12 provides an example of the envelope of a signal. The left-hand side of this signal shows how the loudness of the guitar sound changes with time, where vertical deviations away from zero indicate how loud the sound is at any point in time. This plot shows us how the amplitude of the sound changes while the sound is heard. Mathematically, we define this amplitude variation as the *envelope function* $e(t)$ which changes slowly over several, even hundreds, of periods of the signal. A large value of $e(t)$ corresponds to a loud volume, whereas a small value of $e(t)$ corresponds to a soft volume.

INTERESTING FACT:

Exponential functions show up in many different problems and fields of study, from biology to economics. When food and space are plentiful, the population of a living species in a particular environment grows exponentially. If you put your money in a savings account, it also grows exponentially. Unfortunately so does your debt if you borrow money from the bank.

Different instrument sounds can have different envelopes, as shown in the examples of Figure 2.40. The envelope function for a piano or a guitar sound is initially large at the beginning of the sound when the string is struck or plucked, and it gradually goes down to zero over time. This function can be written as $\exp(t) = 10^{-at}$, where the value of a controls how fast the sound fades away. In Figure 2.40(a) and (b), we used the values of $a = 1$ and $a = 2$, respectively. The envelope of a trumpet or trombone sound typically looks fairly constant while the note is being played. The envelope of a violin or cello sound could start with a gradual increase to a particular point and then decrease steadily back down to zero. The envelope of a clarinet sound might start out large and then get softer as the clarinetist plays. All of these examples are approximate envelopes for these instruments, but we could measure more accurate ones from the actual sounds if we wanted to.

In our digital-band design, we would like to re-create the sounds of real instruments by applying the envelope function $e(t)$ to the periodic function $p(t)$. This process is relatively straightforward if we recall that $e(t)$ is nothing more than an amplitude that changes with time. All we have to do is scale the periodic function by the envelope function. A very good approximation to a musical signal is given as

$$s(t) = e(t) \times p(t) \tag{2.14}$$

where $e(t)$ is the time-varying description of the signal's amplitude and $p(t)$ contains the underlying periodic structure of the signal. In this

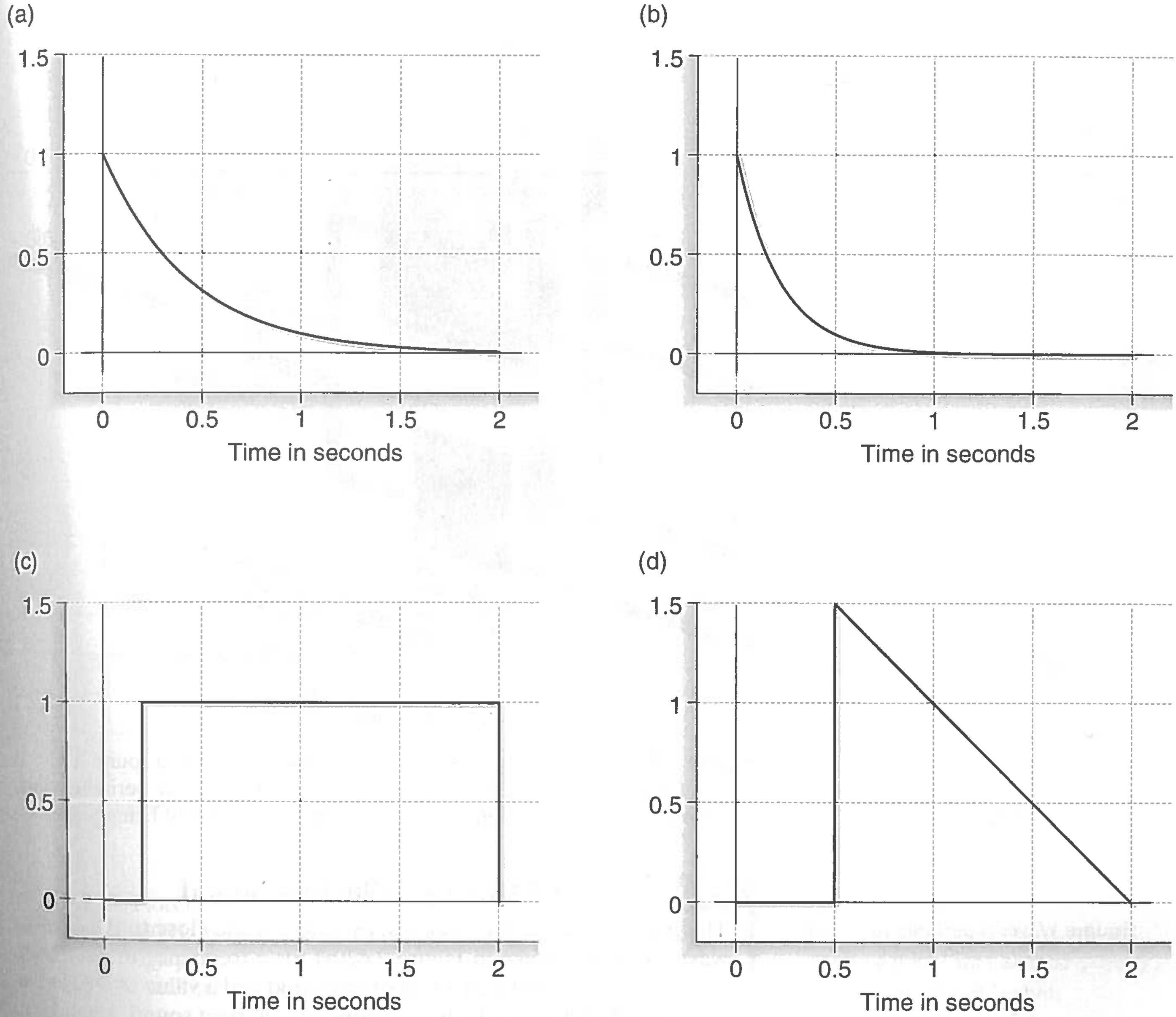


Figure 2.40 Four envelope signals. In (a) and (b), the exponentially decaying envelopes are described by 10^{-t} and $10 - 2t$, respectively. These envelopes are good models for stringed instruments. In (c), the envelope of a trumpet is modeled as a constant amplitude for 1.8 s. In (d), a clarinet envelope is modeled as a constant decrease in amplitude for 1.5 s.

case, the product of the two signals $e(t)$ and $p(t)$ is taken at each time instant. For example, if at time $t = 1.5$ s the values of $e(t)$ and $p(t)$ are $e(1.5) = 0.4$ and $p(1.5) = -0.3$, then the value of $s(t)$ at $t = 1.5$ s is

$$\begin{aligned} s(1.5) &= e(1.5) \times p(1.5) \\ &= 0.4 \times (-0.3) \\ &= -0.12 \end{aligned}$$

Figure 2.41 shows a block diagram for the creation of a realistic note sound, using both the envelope and the periodic signal of a instrument sound. Now that we understand this process, we can create realistic sounds for several musical instruments.

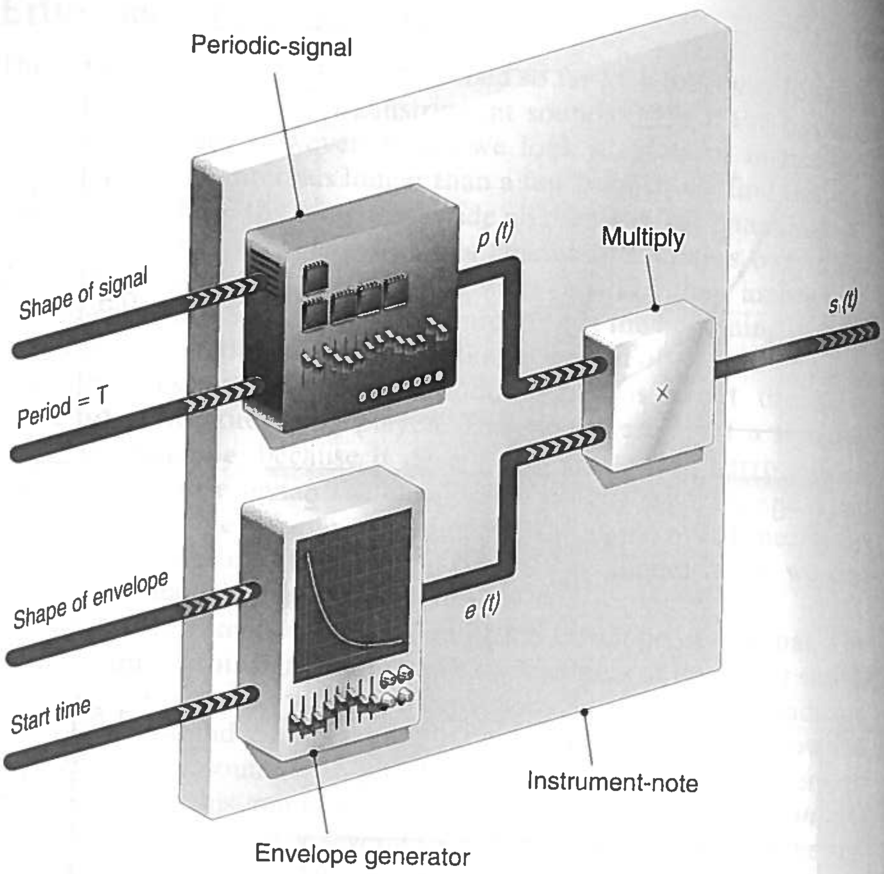


Figure 2.41 Block diagram for creating a realistic instrument sound. The value of T depends on the note being played. The shape of the periodic signal and the shape of the envelope also depend on the instrument being played.

Square Wave: A periodic function that is 1 for half its period and −1 for the other half.

EXAMPLE 2.8 Making a Clarinet Sound

The periodic function of a clarinet sound is quite close to the **square-wave** function, shown in Figure 2.42(a) for a frequency of 440 Hz. It has a value of 1 for the first half of its period and a value of −1 for the second half of its period. The envelope of a clarinet sound depends on how the clarinetist plays the instrument and, more specifically, how much air the clarinetist blows through the instrument. Suppose that a particular clarinetist creates a sound with an envelope given by

$$e(t) = \begin{cases} 0, & t < 0.5\,\text{s} \\ 2 - t, & 0.5\,\text{s} < t < 2\,\text{s} \\ 0, & t > 2\,\text{s} \end{cases}$$

The value of $e(t)$ is plotted in Figure 2.40(d). The combined signal $s(t)$ is computed using Equation (2.14) and is plotted in Figure 2.42(c). Over two seconds, the function $p(t)$ repeats 880 times, so we cannot see the individual periods of $p(t)$ in this plot. If we were to zoom in on any small region of the plot, $p(t)$ would look very much like Figure 2.42(a) with a different amplitude, because $e(t)$ changes very little over a few periods. To get a better sense of what is going on, Figure 2.42(b) and (d) show the same set of plots for a $p(t)$ signal with a fundamental frequency of 2 Hz. We could not hear such a low frequency, but the plots show how the envelope changes the size of the periodic signal.

INTERESTING FACT:

Our perception of the sound of an instrument is affected by its initial attack, or the way the note first begins. Instruments that are excited using blown air, such as flutes, clarinets, saxophones, trumpets, and the like, generally have more gradual attacks than pianos or guitars. For this reason, $e(t)$ increases gradually at first for these instruments.

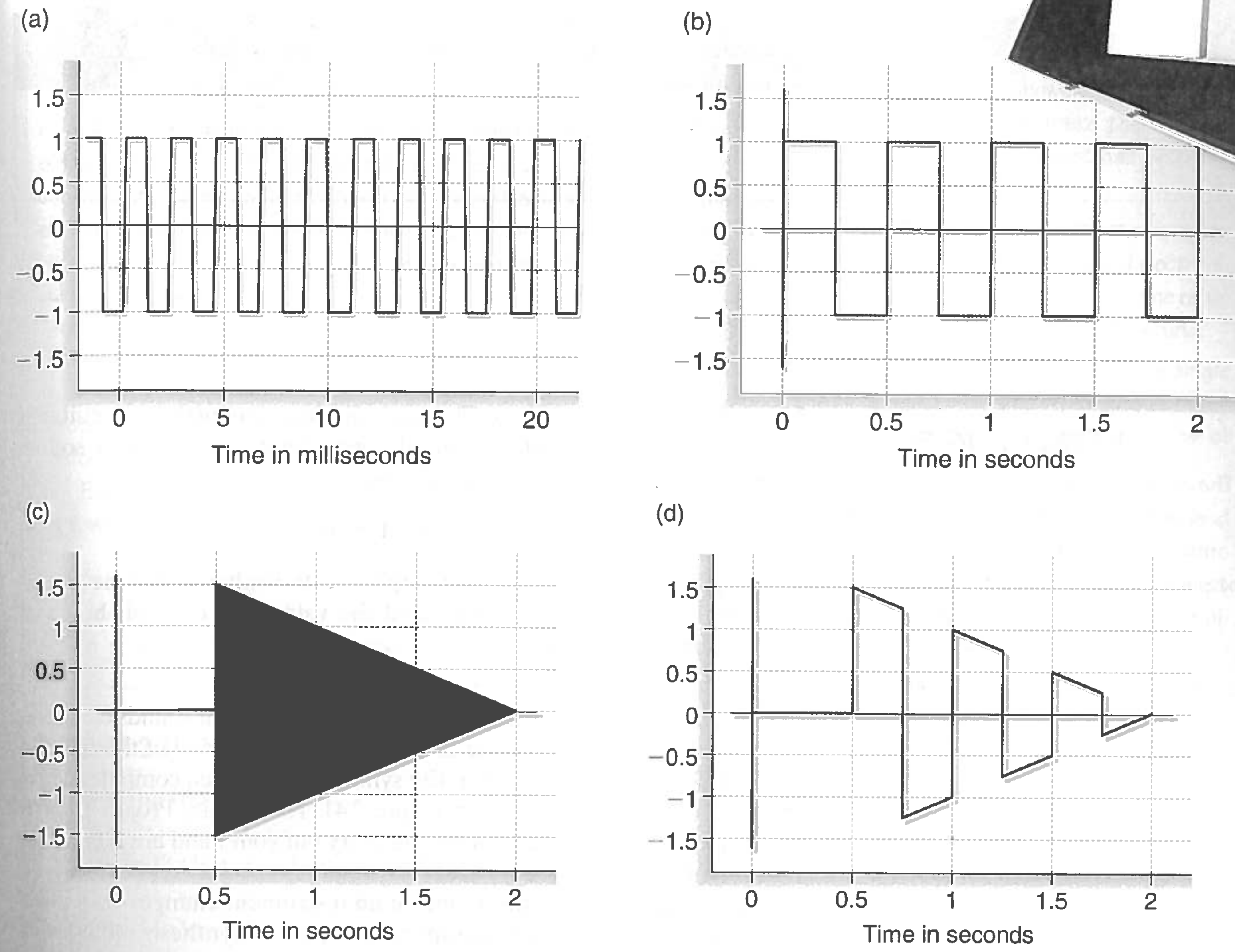


Figure 2.42 Plots of clarinet note with periodic function and envelope. In (a), the square wave at a fundamental frequency of 440 Hz is shown for 25 ms. In (c), the square wave is multiplied by the envelope function from Figure 2.40(d) and displayed for 2 s. In (b), a square wave with a fundamental frequency of 2 Hz is shown, and in (d) the same wave is shown multiplied by the same envelope function.

EXAMPLE 2.9 Making a Guitar Sound

When you listen to someone pluck a guitar string, you can hear the sharp attack of the sound followed by a slow decay to silence. This envelope is very well modeled by the **exponential function**

$$e(t) = B^{-at}$$

for $t \geq 0$, where B can be any constant value. The positive constant a describes the **rate of decay** of the function over time. This type of envelope is shown in Figure 2.40(a) and (b) for $B = 1.0$ and $a = 1$ and 2.

A larger value of a will cause a faster decay, making the guitar sound more like a ukelele or banjo. A smaller value of a will cause a slower decay, making a sound more like that of a bass guitar. Exponential functions show up in many real-world problems in engineering, physics, and mathematics, and they can be easily computed with

Exponential Function: A waveform that grows or decays at a constant rate.

Rate of decay: A number that describes how fast an exponential function decreases over time.

Infinity Project Experiment: SketchWave with Envelope Functions

We can modify our original SketchWave experiment by adding a sketch-pad to draw the corresponding envelope of the sound. Using this system, try to re-create the waveforms corresponding to the clarinet and guitar sounds of this section. How close do the resulting sounds come to their physical counterparts?

INTERESTING FACT:
The small changes in a instrument's signal shape over time are due to many factors. For example, when a saxophone is played, slight differences in the lip pressure and air pressure being exerted at the mouthpiece cause its sound signal to change significantly.

a calculator or computer. The periodic function $p(t)$ for a guitar is very nearly sinusoidal, so a good approximation to a guitar sound, using three parameters, is as follows:

$$s(t) = e(t) \times p(t) = A \times 10^{-at} \times \cos(2\pi f t) \qquad (2.15)$$

The parameter A is the initial amplitude, the value of f is the fundamental frequency of the note, and the value of a controls how fast the sound fades away.

We can use Equation (2.14) to synthesize instrument sounds electronically. All that is needed is some device to input one period of the periodic signal $p(t)$ and the shape of $e(t)$. The synthesizer can then compute the resulting product $s(t)$, as shown in Figure 2.41. The Infinity Project Experiment at the top of this page allows you to try out your hand in synthesizing instrument sounds this way.

In practice, the way the sound of an instrument changes with pitch or frequency is much more complicated than the synthesis methods in this section will allow us to describe. Most signals from instruments have shapes that change a little for each note and within each note while it is playing. Even so, the signals we have created sound quite similar to those of instruments, and they are an improvement over the sinusoids we were using before. More importantly, we now have ways to tailor the sound of our digital band to make more interesting-sounding music. We can now make reasonable approximations of many instrument sounds. We can then create these sounds from different notes and instruments, using the note-generating blocks defined by Figure 2.41. Adding together all of the outputs of these blocks produces the music that we want our digital band to play.

Infinity Project Experiment: Echo Generator

This experiment allows you to see the effect of echo on both recorded music and your own voice. Experiment with different echo amplitudes and delays. What happens when you make the echo amplitude larger than one? Does this sound seem natural?

INTERESTING APPLICATION

Sound Effects

Making great-sounding music doesn't end with the signals that each instrument produces. The room in which music is played alters the music's sound in a subtle, but important, way. A rock guitarist trying to create a new sound might turn up the loudness setting on his electric guitar until it starts distorting. We use the generic term "sound effects" to describe all of the changes that we can make to a sound after it has been recorded or while it is being played. We create sound effects by mathematically manipulating the sound signal before it is played.

Examples of some sound effects are described as follows:

—*Reverberation*, or "reverb," as it is sometimes called, is caused by reflected sound that we hear when we are in a room or concert hall. Reverberation is what makes us sound so good when we sing in the shower. It also is used by professional recording artists to make their tracks sound as if they are played in a concert hall or other room with good acoustics..

—*Echo* is caused by reflected sound that returns after traveling a large distance. Echo can be simulated electronically using simple digital devices. It can provide some really cool effects to a singer's voice or an instrument's sound.

—*Flanging* is a sound effect that is created by a single echo whose delay changes with time. Flanging is often used by electric guitar players to make the sound of their instruments "fatter" or "more alive."

Audio engineers constantly are coming up with new and interesting ways to make great sound effects. Who knows—you might experience some great sound effect the next time you go see a movie, listen to a CD, or play music with your friends!

EXERCISES 2.4

Mastering the Concepts

1. What are the two most popular methods for synthesizing sound? How are they similar? How are they different?
2. How is time scaling used in waveform synthesis? What do you need to know in order to change the frequency of a waveform?
3. Is it possible to add different sinusoids together and still make a periodic signal? How can this idea be used to make audio signals?
4. Write down the expression for synthesizing sounds by using additive synthesis. Can you describe what this equation is doing, in simple words?
5. What is the envelope of a sound? Why is it important in making signals?
6. Give three different examples of envelope functions and how they are used to make signals.

Try This

7. Figure 2.43 shows the plots of two signals $s_1(t)$ and $s_2(t)$. What is the period of $s_1(t)$? If $s_2(t)$ has a period of 2 s, plot it for $0 \le t \le 6$ s.
8. Suppose we want to convert the frequency of a periodic signal corresponding to a D above middle C (293.7 Hz) to that of the B below middle C (246.9 Hz). What time-scaling factor is required?
9. Suppose a periodic signal with period $T = 0.001$ s is to be converted to a signal with a frequency of $f = 600$ Hz. What time-scaling factor is required?

10. Suppose the envelope of a particular sound starts out loud with a value of 1.0 at time $t = 0$ and then decreases linearly to 0 over 4 s. Plot the envelope of the sound. Also, assuming that the period of the periodic signal for this sound is much smaller than 1 s, sketch the overall shape of the signal over 4 s.
11. After getting home from school, you turn on the radio. Somebody else left the radio volume on high, so it blasts a loud sound for 2 s before you turn it down to half its original volume. You then listen to another 60 s of music before turning it off to go do something else. Draw the envelope of the sound of the radio, where $t = 0$ indicates the time that you turned on the radio.

In the Laboratory

12. Say any word in your normal voice that has an "eee" sound in it, such as "weave." Capture your sound, using the equipment in the laboratory, and measure the period of the "eee" sound. How does the period of your sound compare to that of other students in the class?

Back of the Envelope

13. Can the envelope of a sound be periodic as well? Give an example of a signal with a periodic envelope.
14. Suppose you are standing on an highway, and a single truck appears on the horizon traveling toward you. The sound of its large diesel engine gets louder and louder as it comes toward you, and after 30 s, it races by you. Thirty seconds later, it disappears over the horizon again. Sketch the envelope of the truck sound that you heard, labeling the time axis carefully. Take an educated guess as to the shape of the envelope over the first 30 s (you need to know more about the physics of sound and the way the truck is traveling to get more specific), but sketch carefully the envelope of the latter 30 s.

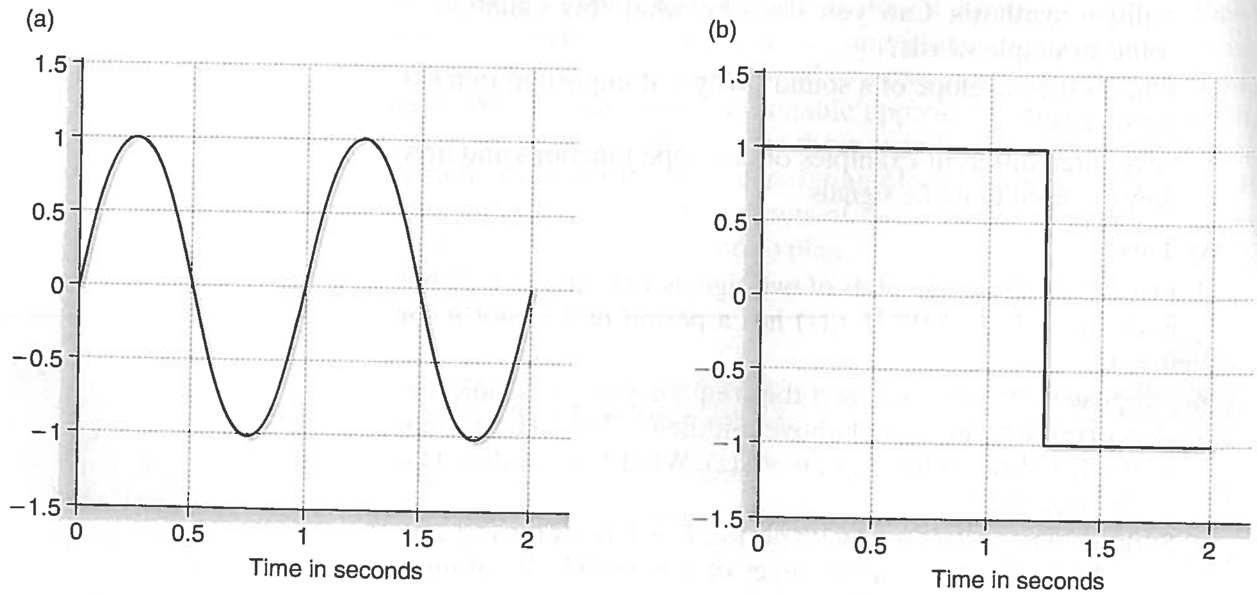


Figure 2.43 Plots of two signals.

Big Ideas

Math and Science Concepts Learned

Any device or system for making music creates a complex sound that can be characterized by signals. A signal is a pattern of variation over time that contains information. Signals are everywhere. They can be used to describe physical phenomena, such as sound waves in the air, the temperature over a single day, and voltages inside of our bodies. We also communicate using radio signals that are transmitted through the air and electrical signals that are transmitted through wires.

Sound is a physical wave or disturbance that travels through air by moving air molecules back and forth. Sound doesn't travel as fast as light does; it takes sound about 3 s to travel 1 km, or 5 s to travel 1 mile. Sound can be converted into an electrical signal by using a microphone. We can also create sound by making an electronic signal and then converting it to sound waves, using a loudspeaker.

Signals are functions that can be plotted easily. Once described, signals can also be manipulated in simple ways. Some of the ways a signal can be manipulated include scaling the amplitude, shifting the time, and scaling the time. These manipulations are especially useful for making music.

When a musical instrument plays a single note, its corresponding sound signal has an inherent structure. The signal that it produces is periodic—repeating over and over—and this period is tied to the pitch that we hear. The numerical equivalent of pitch is the fundamental frequency of the sound. The fundamental frequency f and the period T are inversely related; that is, $f = 1/T$.

Simple melodies are made up of notes played one after the other. Sheet music describes the pitches and durations of the notes being played. Engineers have come up with a computer language—MIDI—to describe such information so that it can be easily stored, transmitted, and used electronically.

One of the simplest and most important periodic functions is the sinusoidal function. This function is generated using the sine and cosine functions from trigonometry by making the angle of these functions dependent on time. The units of angle can be given either in degrees or radians. A tuning fork makes a nearly sinusoidal sound. We can make simple melodies with sinusoids by changing the frequency of the sinusoidal function according to the frequencies in the melody. If we want to play more than one note at the same time, we simply add the signals corresponding to the individual notes being played. This simple technique produces recognizable music to our ears. We can also reverse engineer the score from music created in this way, using a computational tool called the spectrogram. A spectrum analyzer computes and plots the amplitudes and frequencies of the sinusoids within a short segment of a signal.

To make better music, we can employ more complicated sound-synthesis methods to create more interesting instrument sounds. These methods create the periodic shape of a instrument signal in different ways. Waveform synthesis uses a single period of a chosen periodic signal and time scales it so that its fundamental frequency matches that of the desired note. Additive synthesis uses the spectral content of the sound to create a version of the sound signal by summing sinusoidal signals together. The frequencies of these sinusoidal signals are multiples of the fundamental frequency and can be controlled by the fundamental-frequency value. Additive synthesis can be used to approximate any signal, and we have a choice as to how many sinusoidal functions to use in this approximation. Each of these periodic signals can then be combined with the envelope function, which describes the slowly varying amplitude of a sound from beginning to end.

Sound effects are simple manipulations of sound signals to make them sound more interesting or more realistic. Some common sound effects include reverberation, echo and flanging.

Important Equations

There are three basic ways to change a sound signal. The changes in mathematical form are directly related to our perception of the loudness of the signal, the time the signal starts, and the pitch of the signal.

Scaling the amplitude of a signal $s(t)$:

$$x(t) = A \times s(t)$$

Shifting the time of a signal $s(t)$:

$$y(t) = s(t + d)$$

Scaling the time axis of a signal $s(t)$:

$$z(t) = s(ct)$$

Most of the sound we hear when we listen to music or speech is made from periodic functions. Periodic functions also describe mechanical vibrations, planetary motion, visible light, and man-made communications signals. The most basic periodic functions are sinusoidal signals.

Property of a periodic signal where T = period of the signal in seconds:

$$p(t) = p(t + T)$$

Relationship between frequency and period, where f = frequency of the signal in Hz:

$$f = 1/T \quad \text{or} \quad T = 1/f$$

Relationship between sine and cosine:

$$\cos(\theta) = \sin(\theta + 90°)$$

Definition of a sinusoidal signal, where A = amplitude:

$$s(t) = A\cos\left(\frac{2\pi}{T}t\right) = A\cos(2\pi f t)$$

Using our methods of changing signals, we can use sinusoidal signals and simple periodic signal to make realistic music sounds.

Waveform synthesis (changing the period of a periodic signal):

$$p_{\text{new}} = p_A\left(\frac{T_A}{T_{\text{new}}}t\right) = p_A\left(\frac{f_{\text{new}}}{f_A}t\right)$$

Using envelope and periodic signals to synthesize sounds:

$$s(t) = e(t) \times p(t)$$

Simple extensions of these equations can be used to model the sound effects discussed briefly at the end of the chapter. One of the most basic effects is adding an echo to a sound signal. An echo signal can be created by scaling the amplitude of a signal and also shifting the signal's time variable. The scale factor, A, should be chosen to be less than 1, so the echo will not be as loud as the original signal. The time shift, d, should be less than 0 so the echo will happen after the original signal.

If $d = -T_{\text{delay}}$, then the equation for an echo of the signal $s(t)$ is

$$s_{\text{echo}}(t) = A \times s(t - T_{\text{delay}})$$

The total signal is the sum of the original signal and the echo which occurs after a delay of T_{delay}.

$$s_{\text{tot}}(t) = s(t) + s_{\text{echo}}(t) = s(t) + A \times s(t - T_{\text{delay}})$$

A signal with a double echo would be:

$$s_{\text{tot}}(t) = s(t) + A \times s(t - T_{\text{delay}}) + A^2 \times s(t - 2 \times T_{\text{delay}})$$

Another interesting sound effect can be created when the envelope function $e(t)$ is replaced by a sinusoidal function, and the periodic function $p(t)$ is replaced by any sound signal $s(t)$. This is called modulation. The most interesting sound effects are heard when f_m, the frequency of the sinusoidal modulation function, is no more than a few hundred Hertz, but for communications systems much higher frequencies are used.

Modulation:

$$s_{\text{mod}}(t) = \cos(2\pi f_m t) \times s(t)$$