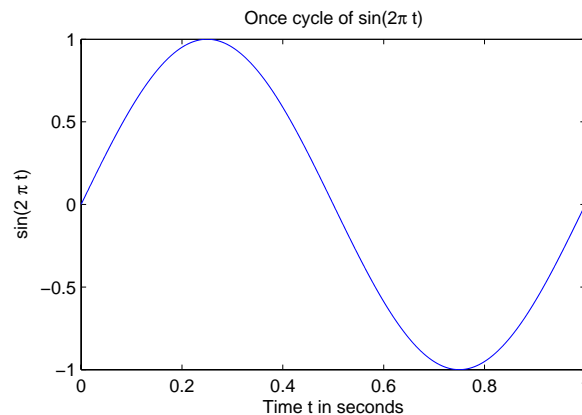


## Trigonometric functions and sound

The sounds we hear are caused by vibrations that send pressure waves through the air. Our ears respond to these pressure waves and signal the brain about their amplitude and frequency, and the brain interprets those signals as sound. In this paper, we focus on how sound is generated and imagine generating sounds using a computer with a speaker.

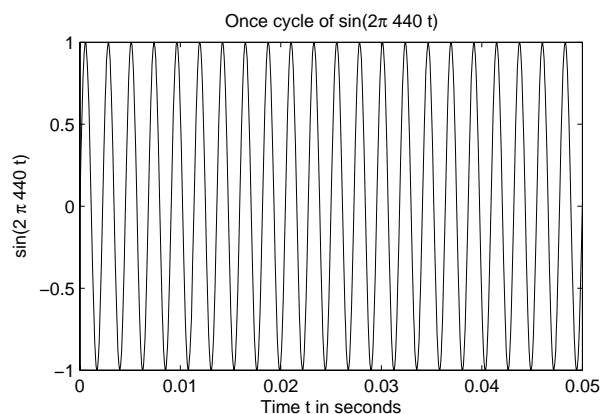
### Frequency of oscillation

We need to describe oscillations that occur many times per second. The graph of a sine function that oscillates through one cycle in a second looks like:



A function that oscillates 440 times per second will look more like this. Note that the time axis only runs to 1/20 of a second.

We say that the oscillation is 440 Hertz, or 440 cycles per second.



### Generating sound with a computer speaker

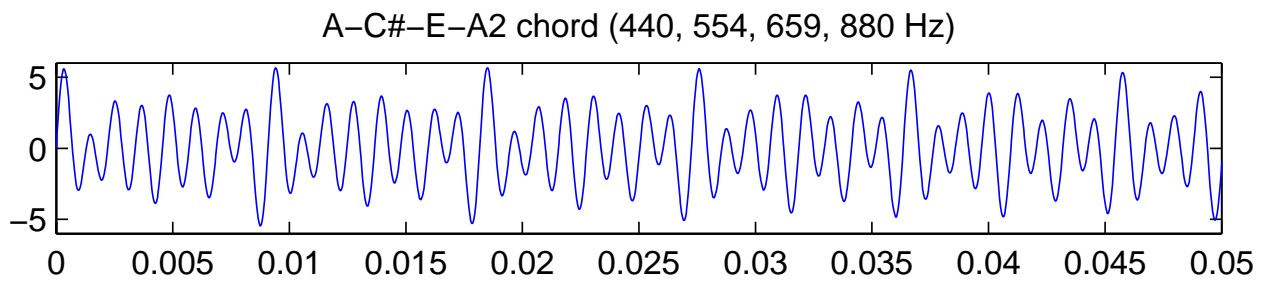
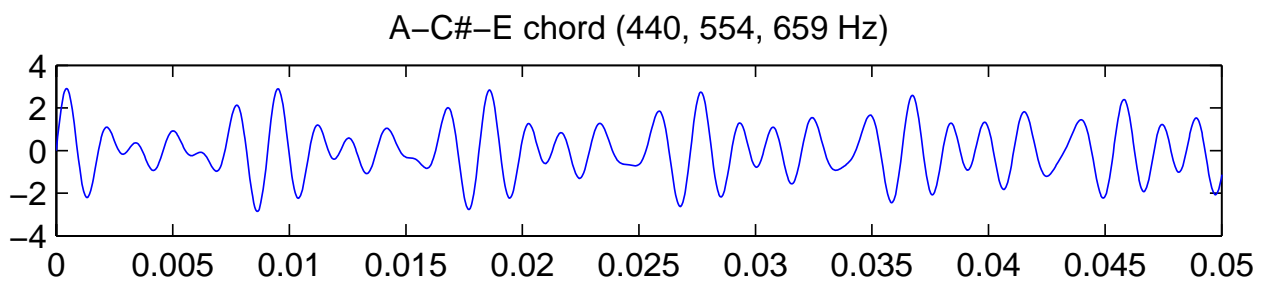
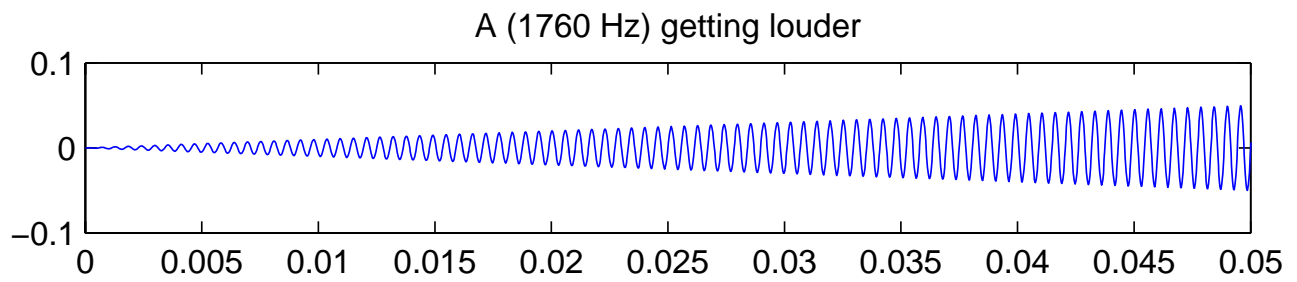
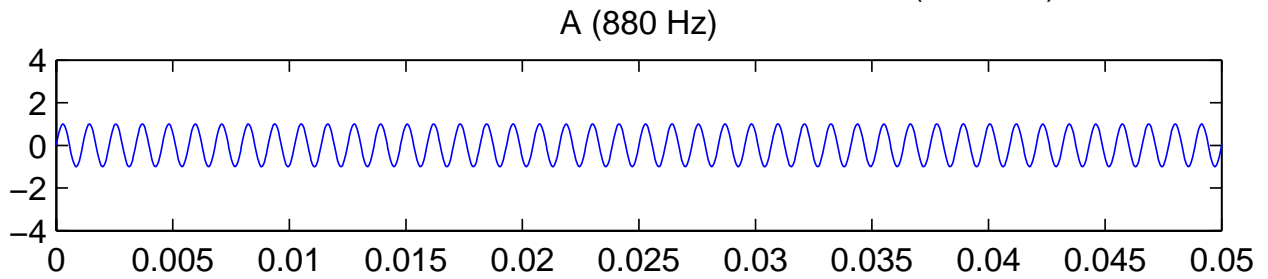
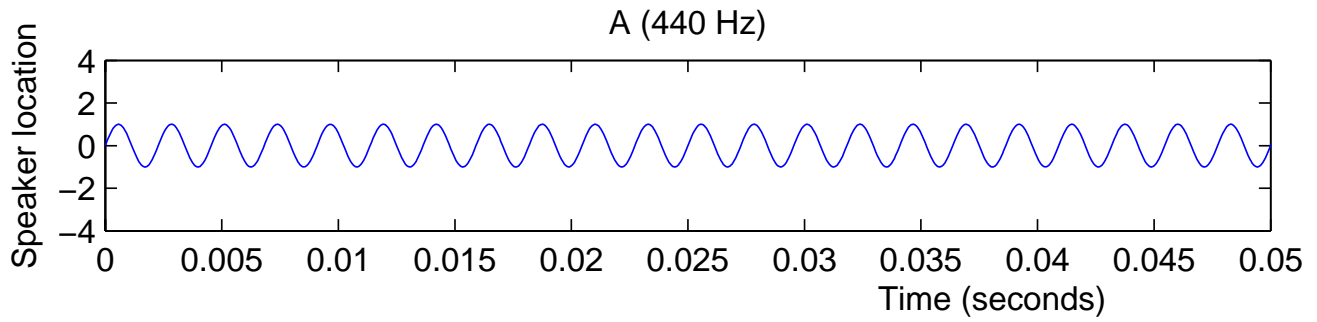
A speaker usually consists of a paper cone attached to an electromagnet. By sending an oscillating electric current through the electromagnet, the paper cone can be made to oscillate back and forth. If you make a speaker cone oscillate 440 times per second, it will sound like a pure A note. [Click here to listen.](#) If you make a speaker cone oscillate 880 times per second, it will sound like an A, but one octave higher. [Click here to listen.](#) We'll call this A2. On a later page, there are graphs of the location of the speaker cone as a function of time, each for one twentieth of a second. Raising the frequency to 1760 Hertz raises the pitch another octave to A3. Changing the amplitude of oscillation, that is, how high and how low the graph goes, or how far forward and backward the speaker cone goes, changes the volume of the sound. The middle graph on the next page shows the amplitude of an oscillation of 1760 Hertz rising from 0 to 1. [Click here to listen](#) to this increasing volume A3 four times.

### Chromatic and major scales

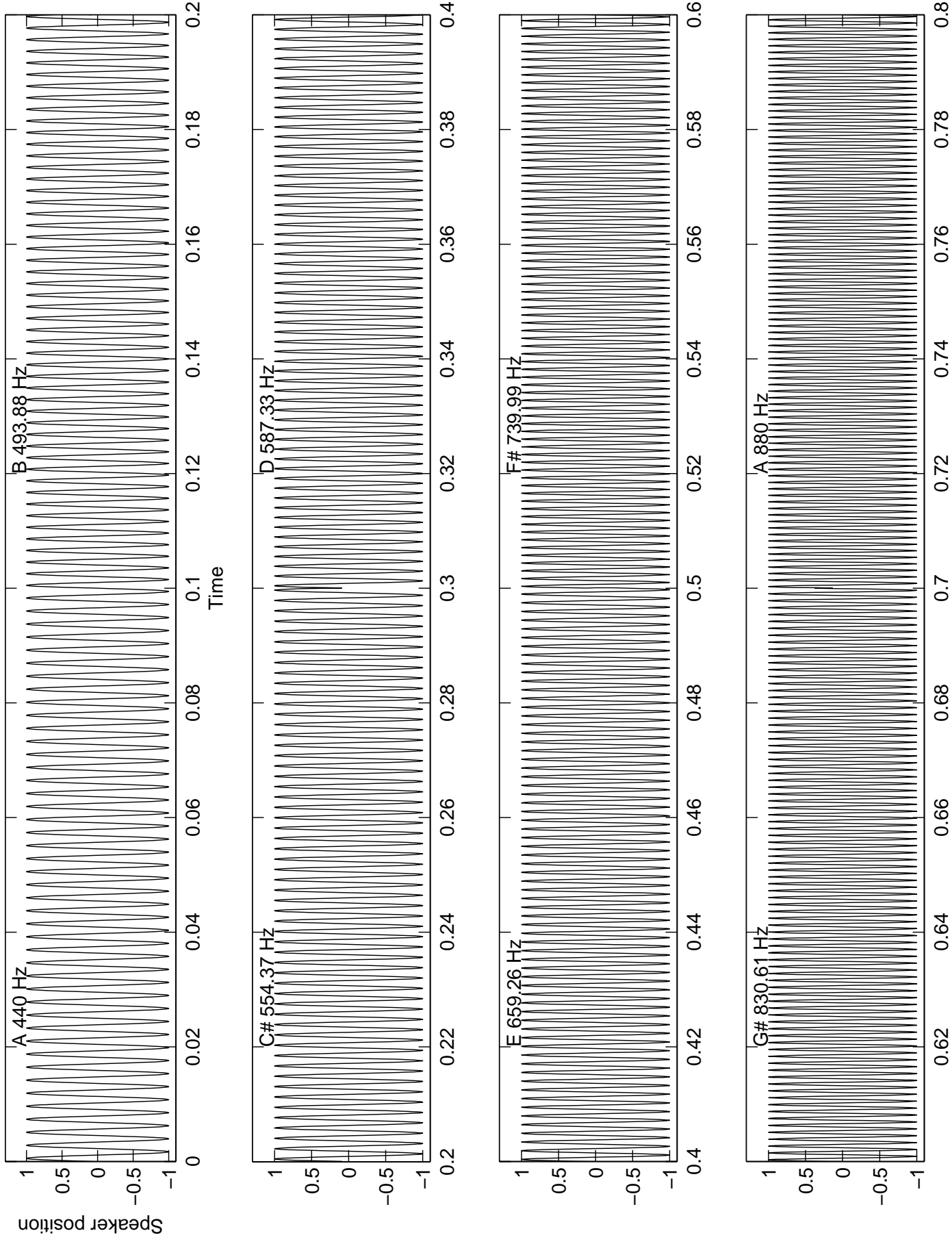
The chromatic scale increases the frequency of oscillation by 12 steps from one octave to another. Starting at A 440, the frequencies of the chromatic scale would be

$$440 \cdot 2^0, 440 \cdot 2^{\frac{1}{12}}, 440 \cdot 2^{\frac{2}{12}}, \dots, 440 \cdot 2^{\frac{12}{12}} = 880$$

Each of the notes in between has a name of its own; they are: A, A#, B, C, C#, D, D#, E, F, F#, G, G#, A2. The A major scale starts at 440 Hertz, then increases over 8 steps to A2, along the notes A (0/12), B (2/12), C# (4/12), D (5/12), E (7/12), F# (9/12), G# (11/12) and A2 (12/12). A graph of how the speaker cone moves in an A-major scale can be found on a later page. [Click here to listen](#) to the A-major scale.



A-major scale



## Chords and superposition of sounds

A **chord** is formed by playing multiple notes at once. You could play a chord with three notes by putting three speakers side by side and making each oscillate at the right frequency for a different note. Or, you can add together the functions for each frequency to make a more complicated oscillatory function and make your speaker cone oscillate according to that function. For example, if you want to play an A – C# - E chord, you can separately make three speakers oscillate according to the functions:

$$\sin(440 \cdot 2\pi t), \quad \sin(440 \cdot 2^{\frac{4}{12}} \cdot 2\pi t), \quad \sin(440 \cdot 2^{\frac{7}{12}} \cdot 2\pi t)$$

If you want one note louder or softer than the others, you can multiply the whole function by a constant to increase or decrease the volume of that note. Each speaker will make pressure waves in the air, and these pressure waves from different speakers will overlap as they move toward your ear. By the time they are at your ear, you will be unable to tell which speaker they came from; the pressure waves will have been superimposed on one another, or added to one another. Your ear is amazing at being able to respond to the different frequencies separately to perceive the three notes being played at once. This explains why you can simply add the three functions above and make the speaker cone of a single speaker oscillate following the sum:

$$\sin(440 \cdot 2\pi t) + \sin(440 \cdot 2^{\frac{4}{12}} \cdot 2\pi t) + \sin(440 \cdot 2^{\frac{7}{12}} \cdot 2\pi t)$$

Here again, if you want the different notes to have different volumes, you can multiply each sine wave by a constant.

[Click here to hear](#) an A – C# - E chord. [Click here to hear](#) an A – C# - E – A2 chord. The graphs on the preceding page show what happens when you add the three functions to make the A – C# - E chord and the A – C# - E – A2 chord.

## Frequency spectrum

The sounds we have generated so far are very simple, being sine functions or sums of a few sine functions, and they sound very much computer-generated. Real sounds are more complex, and it isn't entirely clear that sine functions have anything to do with them. However, it can be shown that any "continuous" sound (that is, a sound that is constant, or unchanging over time) can be reproduced as a sum of sine functions of different frequencies and amplitudes. That is, if a speaker is playing a continuous sound by making the speaker cone follow some function  $L(t)$  over the time interval from 0 to 1 second, then we can write  $L(t)$  as a sum of sine functions this way:

$$L(t) = \sum_{k=20}^{20,000} a_k \sin(k \cdot 2\pi t)$$

The number  $k$  is the **frequency**, and for sounds that humans can hear, we should use frequencies from 20 Hertz to about 20,000 Hertz. As we age, we can't hear sounds at 20,000 Hertz very well anymore. Click here for: [20 Hz](#), [60 Hz](#), (the sound of mechanical devices that hum because of alternating current being 60 Hz) [100 Hz](#), [10000 Hz](#), [16000 Hz](#), [20000Hz](#). Be aware that some of these may be outside the range of your speakers or your ears! The numbers  $a_k$  are called the **amplitudes** for frequency  $k$ . It is easy to find the values of  $a_k$  using integrals:

$$a_k = \int_0^1 \sin(k \cdot 2\pi t) L(t) dt.$$

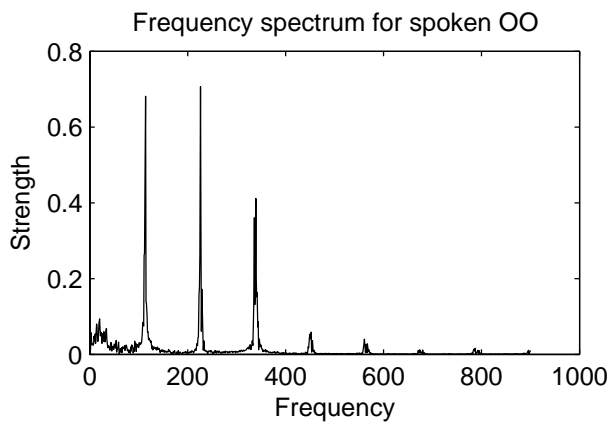
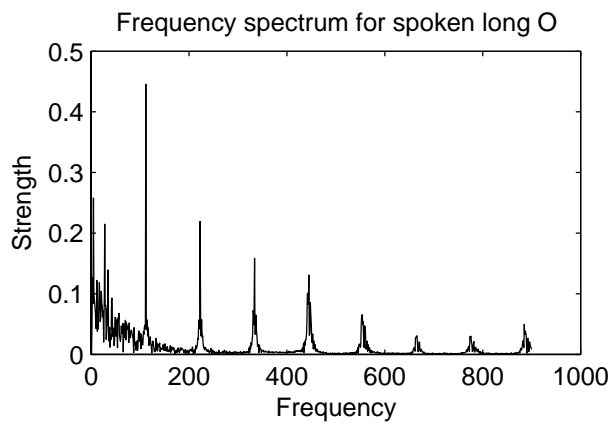
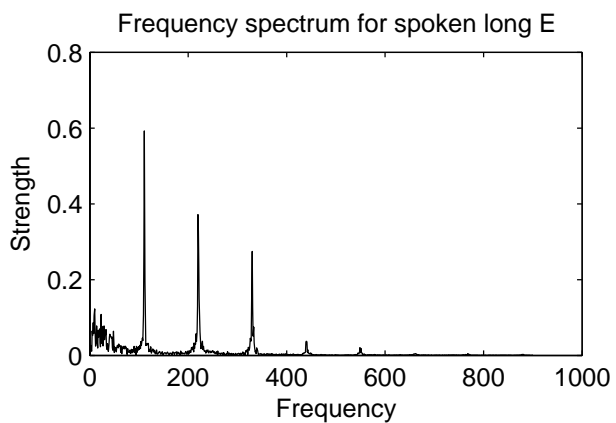
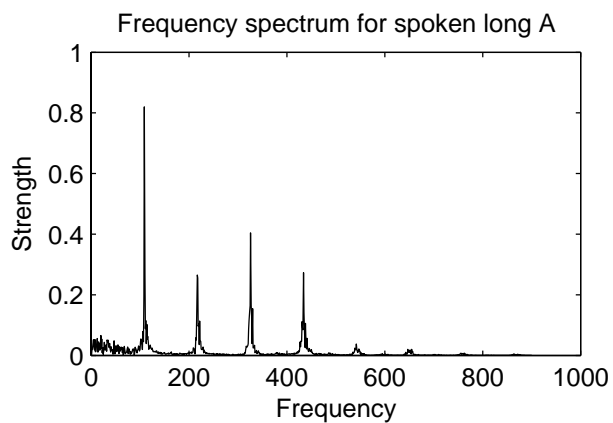
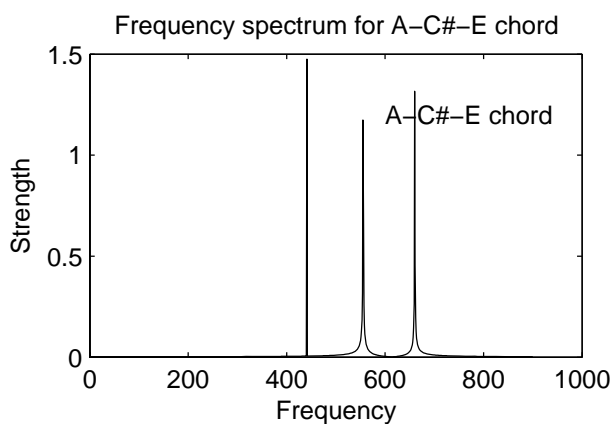
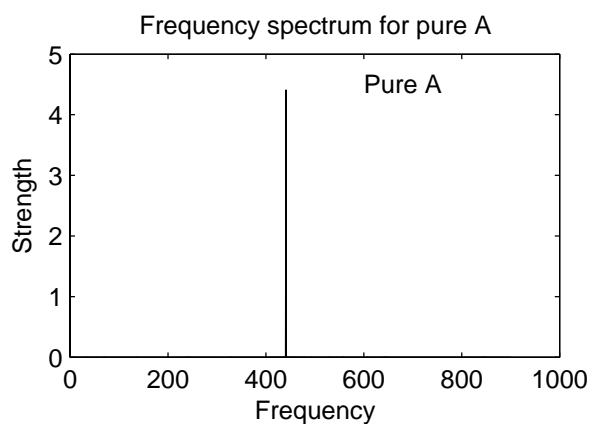
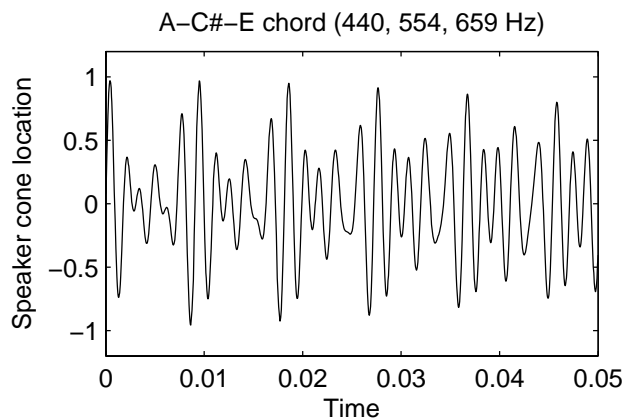
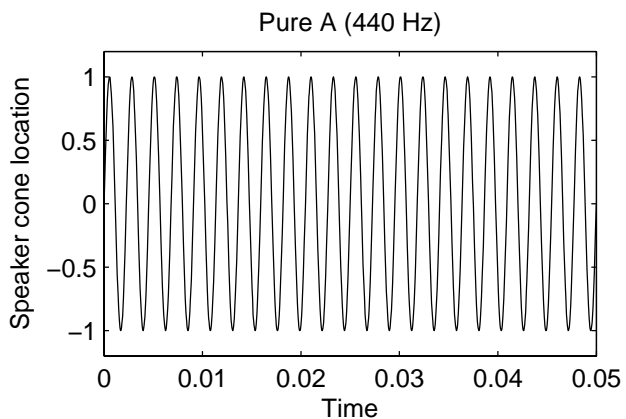
(Actually, this isn't the **whole** story – for this to be exactly correct, you either need to use cosine functions or you need to be able to shift each sine function horizontally on the  $t$  axis, but this is close

enough to the truth that you can learn from it!) The point is that you can take a sound and **think of it** in terms of the different frequencies and amplitudes that it is made up of.

The graph of  $a_k$  versus  $k$  is called the **frequency spectrum** of the sound. It shows graphically which frequencies are present in the sound. On the next page, there are graphs of the frequency spectrum for a variety of sounds that we have seen so far, and some new ones. First, below the graph of the pure 440 Hz A is the graph of the frequency spectrum. The amplitudes  $a_k$  are zero except when  $k$  is 440, the frequency of the oscillation. Next, below the graph of the A – C# - E chord is its frequency spectrum, concentrated at the three frequencies present in that chord. Some frequencies near the C# and E frequency are also present due to the numerical technique that finds the frequency spectrum.

The next four graphs are the frequency spectra of Dr. Craig Zirbel saying the continuous sounds long A, long E, long O, and OO. These were all spoken at the same pitch, so they all have frequency spikes at similar frequencies. The difference between these sounds is in the relative heights of the peaks. If you wanted to make a computer recognize and differentiate between these sounds, you could train it to pay attention to the relative heights of the different peaks.

Click here to listen to the [long A](#), [long E](#), [long O](#), and [OO](#).



## Telephones

In the past, telephone companies would lay down copper wire or fiber optic cables, or use radio, to carry telephone conversations. By varying the voltage in a copper wire, the telephone company could send audio signals over long distances. Varying the voltage corresponds directly to varying the speaker location (and thus the air pressure). However, electronic devices are capable of making much finer voltage variations, and of measuring much finer variations, than the human ear can. As a consequence, telephone companies can send multiple conversations over the same channel, essentially by converting your conversation into a frequency representation, and then shifting the frequencies, say, all by 1000 Hz, and sending other phone calls through with a shift of 2000 Hz, 3000 Hz, etc. At the other end, these are shifted back into the normal range and sent to the receiving telephone. This is called multiplexing. In order to do this, however, the frequency range available for YOUR telephone conversation has to be limited. Telephone companies would do this by chopping off (setting to zero) the frequency components at the low and high ends of the frequency spectrum. This is why your voice sounds strange over a telephone. The limited “bandwidth” is also what prevents computer modems from transferring information more quickly over copper wires. There is no inherent limitation in the wires; copper wires also supply broadband internet. The difference is that the telephone company limits the range of frequencies that can be used, which has the effect of limiting the rate at which information can be pumped through the circuit.

## Digital audio

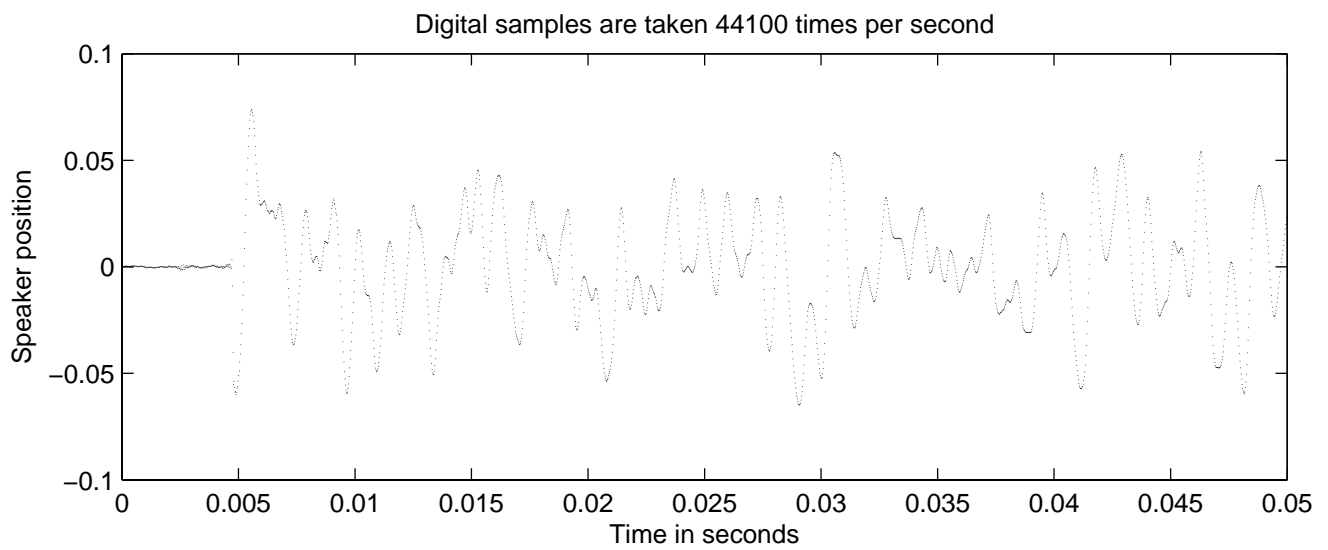
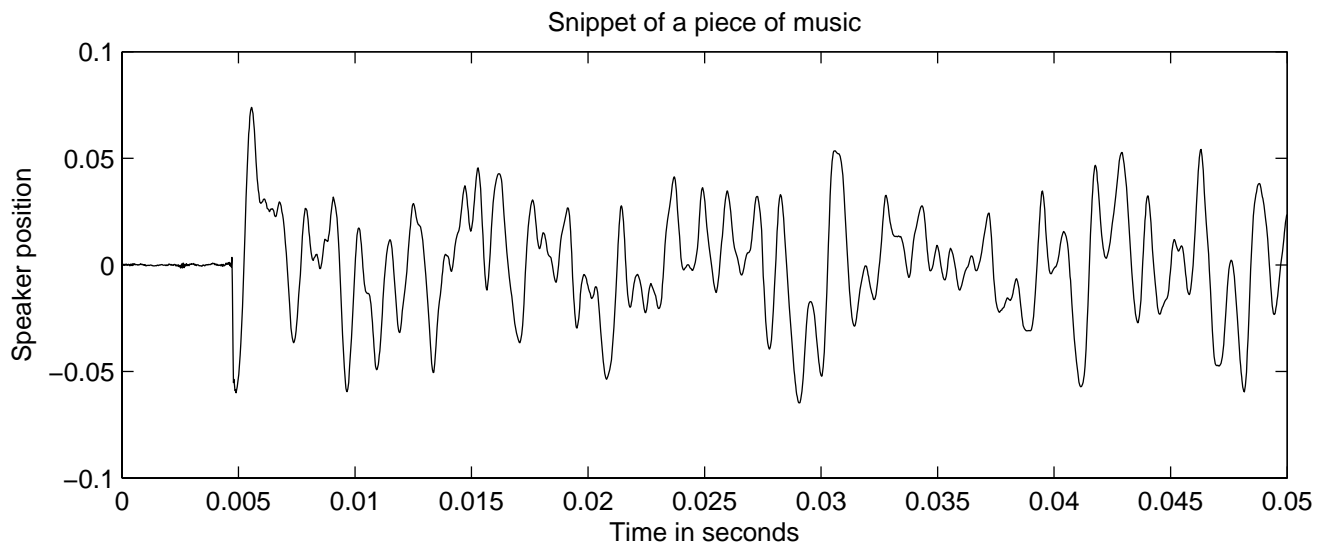
Actual audio occurs in continuous time, from instruments, voices, etc. See the next page. [Listen to the snippet](#) shown. [Listen to it several times](#).

**CD audio.** To make a compact disc, a digital device measures and records the microphone location 44100 times per second and stores the number using 16 bits ( $2^{16} = 65536$  possible values). This number is stored directly on an audio CD. A 70 minute CD would require:

$70 \text{ minutes} * 60 \text{ seconds per minute} * 44100 \text{ data values per second} * 2 \text{ bytes per value} * 2 \text{ audio channels} = 740880000 \text{ bytes} = 707 \text{ megabytes,}$

which is about what a CD can hold. Sampling at 44,100 times per second means the upper limit on the frequency that can be accurately reproduced is 22,050, which is beyond the range of the human ear.

**MP3 audio.** Since the late 1970's, the game has been to try to use fewer megabytes to store music. There are a variety of approaches one could try. The most successful ones so far take into account the fact that music can be decomposed into various frequencies, as we have discussed above. The frequencies that are outside the range of human hearing can be discarded, and so one can store the sound using fewer bytes of data. Several other observations from the field of psychoacoustics allow one to throw away even more data without (significantly) compromising sound quality. For example, in a certain snippet of music, a certain frequency may play very softly. If this is below the audible threshold for that frequency, that part of the music need not be stored. If other frequencies are being played very loud at that moment, again the ear cannot hear the quiet frequency, and there is no need to store the data that that frequency is being played. The quiet frequency is **masked** by the louder frequencies. In addition, the human ear cannot distinguish sounds with very similar frequencies, so this limits how many different frequencies need to be stored.



This handout is available at <http://www-math.bgsu.edu/z/sound>