[Most of this lecture was finishing up the linear systems material from lecture 17. See the lecture 17 notes.]

Introduction to sound

A few lectures from now, we will consider problems of *spatial hearing* which I loosely define as problems of using sounds to compute where objects are in 3D space. There are two kinds of spatial hearing problems in perception. One involves sounds that are *emitted* by objects in the world, which is the spatial hearing problem we are used to. The other involves sounds that are *reflected* off objects in the world. This is the spatial hearing problem used by echolocating animals e.g bats. This year we will (probably) just have time to cover the first.

Emitted sounds are produced when forces are applied to an object that make the object vibrate (oscillate). For example, when one object hits another object, kinetic energy is transformed into potential energy by an elastic compression [1]. This elastic compression results in vibrations of the object(s) which dampen out over time depending on the material of the object. These object vibrations produce tiny pressure changes in the air surrounding the object, since as the object vibrates it bumps into the air molecules next to it. These air pressure changes then propagate as waves into the surrounding air.

There are three general factors that determine a sounds. The first is the energy source which drives some object into oscillatory behavior. The second is the object that is oscillating – its shape and material properties. The third factor is the space into which or through which the sound is emitted. This space cavity can attenuate or enhance certain frequencies through resonance. The varying shapes of musical instruments are an obvious example. Voice is another, and we will return to both.

Emitted sounds are important for hearing. They inform us about events occurring around us, such as footsteps, a person talking, an approaching vehicle, etc. Here we have a major difference between hearing and vision. Nearly all the visible surfaces reflect light rather than emit light.[2] However, light sources themselves are generally not informative for vision but rather their 'role' is to illuminate other objects, that is, shining light on other objects so that the visual system can use this reflected light to estimate 3D scene properties or recognize the object by the spatial configuration of the light patterns. By contrast, emitting sound sources *are* informative. They tell us about the location of objects and their material properties.

What about reflected sounds? To what extent are they useful? Blind people seem to use reflected sounds (echos) to navigate. Blind people can walk through an environment without bumping into walls and to do so they use the echos of their footsteps and the echos of the tapping of their cane. They hear the reflections of these sounds off walls and other obstancles. People like us that have normal vision probably use reflected sound also but not nearly as much as blind people. This hasn't been studied much.

[1] Another component of the kinetic energy is transformed into non-elastic mechanical energy, namely a permanent shape change. This happens when the object cracks, chips, breaks or is dented

[2] Typically only hot objects emit light. Electronic displays/lights e.g. LEDs are obvious exceptions to this statement. Not only are they non-hot light emitters, but the light patterns they emit are often meant to be informative – perhaps the light pattern you are seeing right now.

Pressure vs. intensity

Sound is a set of air pressure waves that are measured by the ear. Air pressure is always positive. It oscillates about some mean value I_a which we call *atmospheric pressure*. The units of air pressure are atmospheres and the mean air pressure around us is approximately "one atmosphere".

Sounds are small variations in air pressure about this mean value I_a. These pressure variations can be either positive (compression) or negative (rarefaction). We will treat the pressure at a point in 3D space as a function of time,

$$P(X, Y, Z, t) \; = \; I_a + I(X, Y, Z, t)$$

where $I(X, Y, Z, t)$ is small compared to atmospheric pressure I_a. Note that we are using "big" X, Y, Z rather than little x, y, z, since we are talking about points in 3D space.

I emphasize that sounds are quite small perturbations of the atmospheric pressure. The quietest sound that we can hear is a perturbation of 10^{-9} atmospheres. The loudest sound that we can tolerate without pain is 10^{-3} atmospheres. Thus, we are sensitive to 6 orders of magnitude of pressure changes. (An *order of magnitude* is a factor of 10.)

Loudness: SPL and dB

To refer to the loudness of a sound, one can refer either to *sound pressure* $I(X, Y, Z, t)$ or to the square of this value, which I will loosely refer to as *intensity*. Sound pressure oscillates about 0, whereas intensity is of course always positive. Think of intensity as the energy energy per unit volume, namely, the work done to compress or expand a unit volume of air to produce the deviation from I_a.

When describing the loudness of a sound, we typically don't care about the instantaneous pressure or intensity, but rather the average over some time. Averaging the sound pressure $I(X, Y, Z, t)$ over time makes no sense, since the average is I_a which has nothing to do with the loudness of a sound. Instead one averages the intensity over some time T. Let I be the root mean square of the sound pressure:

$$I \equiv \sqrt{\frac{1}{T} \sum_{t=1}^{T} I(t)^2}$$

The loudness will then be defined in terms of the ratio of the RMS sound pressure to the RMS sound pressure of some standard I_0 which is very quiet sound called the "threshold of hearing". I_0 is a specific extremely soft sound level which has been determined in careful experiments in acoustically isolated rooms. It is the quietest sound below which one cannot discriminate.

The range of sound pressures that we can hear comfortably is very large. Moreover, psychophysical studies have shown that people are sensitive to ratios of RMS sound pressures rather than differences. For these reasons, one defines the loudness by the log of this ratio:

$$\text{Bels} = \log_{10} \frac{I^2}{I_0^2} = 2 \log_{10} \frac{I}{I_0}$$

It is common to use a slightly different unit, namely ten times Bels. That is, the common units for defining the loudness of a sound is the *sound pressure level* (SPL) in SPL in decibels (dB):

$$10 \log_{10} \frac{I^2}{I_0^2} \; = \; 20 \;\; \log_{10} \; |\frac{I}{I_0}|$$

Multiplying by a factor 10 is convenient because the human auditory system is limited in its ability to discriminate sounds of different loudnesses, such that we can just discriminate sounds that different from each other by about $\frac{1}{10}$ of a Bel, or 1 dB. So, we can think of 1 dB as a just noticable difference (JND).

As an example, suppose you were to double the RMS sound pressure from I to $2I$. What would be the increase in SPL (dB) ? To answer this, note

$$20 \log_{10} \frac{2I}{I_0} = 20 \log_{10} 2 + 20 \log_{10} \frac{I}{I_0}$$

So the increase in dB is $20 \log_{10} 2 \approx 6$ dB.

Here are a few examples of sound pressure levels:

Sound	dB
jet plane taking off (60 m)	120
noisy traffic	90
conversation (1 m)	60
middle of night quiet	30
recording studio	10
threshold of hearing	0

