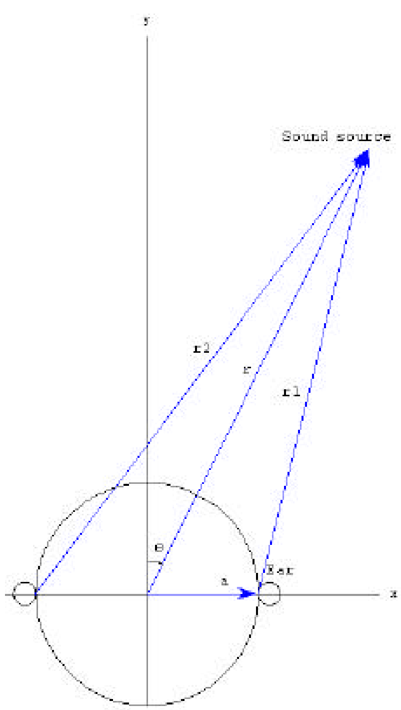
Dustin Lindley
Physics 398 EMI
Final Report

Experiments in Sound Localization

Understanding how the brain interprets the data it recieves is a very difficult task. In the field of audio processing, we have psychoaccoustics. Possibly the most important problem in psychoaccoustics is how people have the ability to locate sound sources with no visual cues. Like our eyes, our ears are receptors, placed a distance apart, so that triangulation may occur. Unlike light waves, however, sound waves are not necessarily blocked by the head, so the process of visual localization and auditory localization are not exactly analogous, but it is expected that there are some similarities, and experiment shows that this is the case. Unfortunately, the data set consists of a very small (8) group of people, but it is hoped that this set will be added to by distributing the sound files used in the experiment via the internet. In any case, there is evidence that the human brain can perceive differences in arrival times of accoustic signals of around 0.045 mS in a signal with easily recognizable transients (the voice of 3CP0). For signals with no transients, localization was much more difficult, unless the signal was "moving" in the auditory "field of view."

I. Experimental Setup and Theory

All that is needed to do experiments in sound localization is a PC equipped with a soundcard, some sort of wave file editor that can edit on the scale of a single sample, and a pair of headphones. Basically you take a mono sound, and make two copies of it, but remove some samples from the beginning of the second copy, then paste those two sounds into a stereo file. If you remove samples from the right side, the sound will appear to come from the right, because that signal will reach your right ear before your left ear, and vice versa. I used 44100 Hz wave files, so the smallest time difference that could even be done was around .022 mS, but it is possible to use higher resolutions, which would increase the accuracy of the experiment. The sounds are played back through headphones to insure that the stereo signals do not leak from left to right.



The picture on the left shows a head (use your imagination) with a sound source at some distance r from the head with an angle theta to the right of the head. The distance from the center of the head to an ear is a. With this arrangement, we find that that the time delay between the ears is

$$\text{tRight} - \text{tLeft} = \frac{\sqrt{a^2 + r^2 - 2\,a\,r\,\mathrm{Sin}[\theta]} - \sqrt{a^2 + r^2 + 2\,a\,r\,\mathrm{Sin}[\theta]}}{\text{Speed of sound}}$$

Which is not very pretty at all. However, if we assume that r >> a, then expanding in a and taking only first order terms, we find that

$$\text{tRight} - \text{tLeft} = -\frac{2\,a\,\mathrm{Sin}[\theta]}{\text{Speed of sound}}$$

which is much easier to work with. This is an approximation, but an apt one because most sounds come from much farther than half of a heads distance away. The speed of sound is approximately 340 m/s.

The sound file used for the main experiment was the voice of 3CP0, the gold-plated robot from Star Wars. The sample was chosen because it was expected that a voice, which is rich in transients, would be easier to localize than a pure sine wave or some other pure tone. In nature, most every sound that could signify some sort of danger to us or prey for us would not be a pure tone, but would be very rich in transients. It makes sense that we can localize these sorts of sounds more effectively than pure tones, which really do not exist outside of human creation.

One large stereo wave file was made with 10 copies of 3CP0 saying a short sentence. The 10 copies were delayed either to the left (L) or right (R) in pairs, starting out with 5 samples L, then 5 samples R, then it is reduced to 4 samples with either L then R or R then L, then 3, then 2, and finally 1. Other files were made to test ability to localize pure tones, hear "moving" sounds, etc.

II. Results

The results that have been found are quite interesting. In the 3CP0 experiment, it was found that the brain can discern a difference in arrival time corresponding to 2 samples, or 0.045 mS. This is a vanishingly small amount of time. All eight participants were able to accurately answer R or L for this amount of time difference. The angle that this delay corresponds to is about 5.53 degrees. On the sounds with with a delay of one sample, 3 of the 8 participants were able to correctly answer R or L. More data needs to be gathered in the 0.05 – 0.01 mS range to find the cutoff range.

For pure tone signals (sine waves), the brain has a hard time localizing at all, regardless of the delay, unless the signal is moving across the auditory field of view. This is much the same as with vision, where movement is detected before any other visual details. This also makes evolutionary sense because something that is moving and making sounds as it does so has a good chance of being something that a human would want to eat or avoid.

Just to see what would happen, a sound was created with a very large delay of 150 samples or about 3.4 mS. It was expected that the brain would interpret these two signals from each ear as separate sounds, but this was not the case. Even up to these large time differences, a single sound was heard, but all participants in this experiment had a hard time telling if the sound was coming from the left or the right.

III. Future Experiments and Summary

The surface has been scratched, but there is much to be done. It is my hope that future physics 398 students will take an interest in this topic and find more accurate and complete conclusions.

For the 3CP0 experiment, there are many improvements to be made. Increasing the resolution of the sound files to 88200 Hz (or even higher resolutions, if the soundcard can handle it) would increase the accuracy of the experiment, as would gathering data from a larger group. If the experiment continues, it is suggested that rather than putting sounds with the same delay in L R pairs, which basically gives the subject two chances to localize the sound, the sounds be mixed more or less randomly. One limitation with this experiment is that we are only asking for L or R, so a random guess has a 50 % chance of being correct.

For the pure tone signal experiments, it would be interesting to do physical experiments with actual speakers placed around the subject rather than relying on the wave files. I believe that subjects would be able to localize sounds from speakers, but that amplitude differences, rather than time differences, would be the main indicator. Perhaps a simulated human head, with microphones attached to the ears would give some insight in this case as well.

One possible application of these experiments is in making stereo recordings. Currently, stereo recordings are made by increasing the amplitude of a certain instrument in either the left or right side of the stereo recording. The experiements above prove that amplitude differences are unnecessary, because the two sides of the stereo only differ in time, not amplitude. Perhaps stereo done by time delay rather than amplitude differences would be useful for recording studios. I am of the opinion that the time delay method of stereo mixing would give much better "stereo separation" than traditional methods do. A very accurate digital delay (a fairly well known and common piece of equipment) would be used to delay one signal or the other in real time, rather than piecemeal as I have done.

The main goal is to understand how hearing works, and how the brain understands the world around it through hearing. There are two basic properties of sounds with respect to our two ears, relative amplitudes and arrival time differences, and both surely play a part in the localization of sounds. The way to understand these effects is to understand them separately, and then try to put them together to form a coherent whole.