Lecture 9: Sound Localization

Localization refers to the process of using the information about a sound which you get from your ears, to work out where the sound came from (above, below, in front, behind, left, right, ...).

It is essential to distinguish from the start, the problem of localizing a steady sound from the problem of localizing a sudden sound. The steady sound is much harder (as we experience daily!).

Steady sounds are localized mainly by two things:

- relative loudness in the two ears,
- timing in the two ears.

Relative loudness

Relative loudness is fairly straightforward. If one ear is in the direction of the sound, and the other ear is on the far side of the head from the sound, then the "sound shadow" of the head will make the sound softer on the far side than on the near side.

However, last lecture we saw that sound tends to bend around obstacles. Whether it does so or not, depends on the relative sizes of the object and the wavelength of the sound. The head will give a good shadow provided that,

$$\lambda_{\rm sound} < D_{\rm head} \,,$$

with $D_{\rm head}$ the diameter of your head, which is around 20 cm = 0.2 m. This gives,

$$\lambda_{\rm sound} < 0.2 \text{ m} \quad \Rightarrow \quad f_{\rm sound} = \frac{v_{\rm sound}}{\lambda_{\rm sound}} > \frac{340 \text{ m/s}}{0.2 \text{ m}} = 1700 \text{ Hz} \,.$$

Of course, this is not an exact statement. The further above this frequency the sound is, the more pronounced the head shadow effect will be. Therefore, it will be a pronounced effect at 5000 Hertz, a modest effect at 2000 Hertz, and nearly nonexistent below 1000 Hertz.

Naturally, the more information you have, the more accurately you can determine the sound's origin. For instance, a complex tone with strong harmonics will let your ears simultaneously determine the size of the head shadow effect at the fundamental and at each harmonic. Above several thousand Hertz, the pinnae (outer ears) also start to provide shadows, giving more forward-backward information.

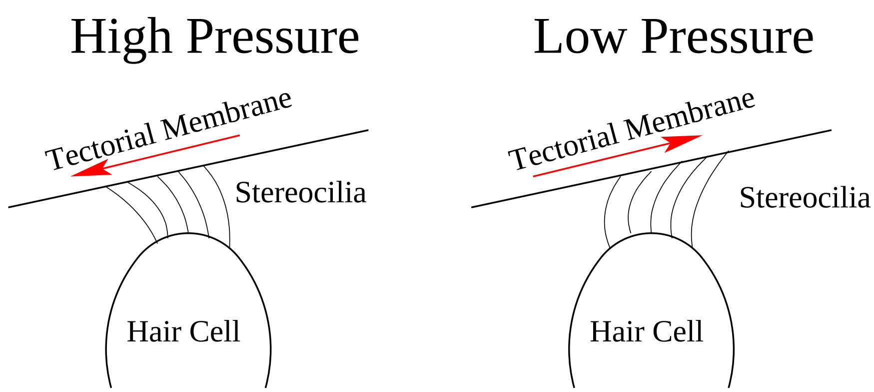
To illustrate this, try the following experiment. Get two speakers, and place them so they will be about 40° degrees from each other as seen by you. Then play the (stereo) first MP3 file provided with this lecture, and see if you can identify which speaker each tone is coming from. The tones are,

- a 2000 Hertz sine wave,
- a 4000 Hertz sine wave,
- a 6000 Hertz sine wave,
- a 2000 Hertz wave with strong harmonics.

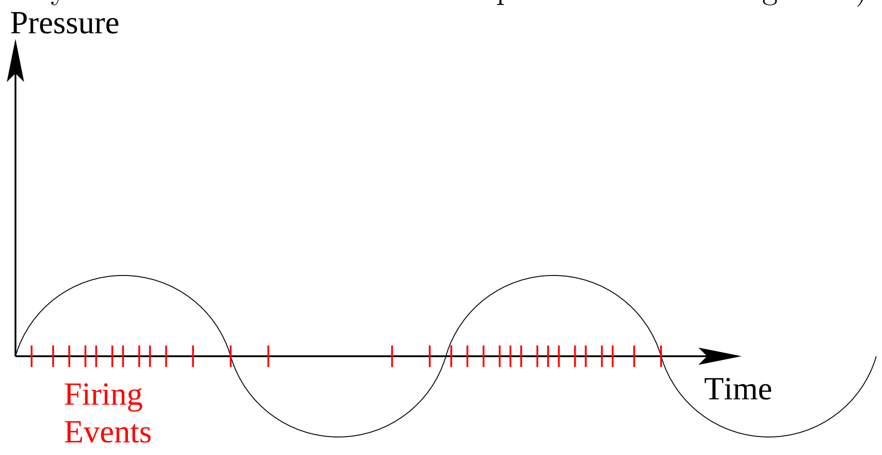
You will agree that localization this way is possible, but not very effective. (Try seeing how close together the speakers can be before you can no longer tell.)

Timing

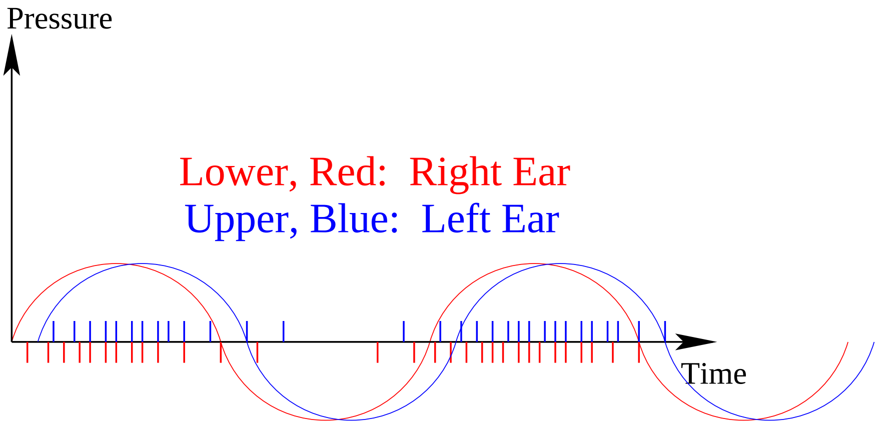
You would think that there would be no way to use timing to tell the origin of a steady tone. However, you would be wrong. To see why, one has to think about the mechanism by which the sound wave gets converted into a nerve signal by the hair cells. Recall the picture. The pressure on the eardrum turns (via the ossicles and the fluid system of the cochlea) into alternating right and left waving of the stereocilia of a hair cell:



The nerve cell "fires" when the stereocilia bend one direction. Bending the other direction actually *inhibits* their firing. (If there is no sound, the hair cells actually fire occasionally, randomly. You can tell the hair cells are bending one way because the firing becomes frequent, and the other way because it becomes more infrequent or absent altogether.)



If there is a sound in each ear, even if they are the same loudness, the high pressure peaks occur earlier in the ear closer to the source of the sound (because the pressure peak reaches that ear sooner). Then the firings of nerves in one ear will start a little sooner than the firings in the other ear:



The figure is for the case where the right ear is nearer the signal.

These signals are then carried by nerves to the brain, where their relative time of arrival is determined in a small lobe near the bottom middle of the brain, called the *Medial Superior Olive* (really it is, honest). From the difference of time of arrival of the "high pressure" signals in each ear, this organ can find the difference of time between arrivals in the two ears.

The limitation of this mechanism is that a nerve cell which has just sent a nerve signal cannot send another signal right away. The nerve requires a **refractory period** before it has recharged and can send another signal. The refractory period is about 1 millisecond (.001 seconds), and there is a further 4 milliseconds or so during which it takes a larger stimulus than usual to cause the cell to fire. This actually does not mean it is impossible to get timing information at higher frequency, but it certainly makes it harder. Therefore, localization through timing becomes ineffective for frequencies above around 1000 Hertz.

(Incidentally, the timing of nerve cell responses is a secondary way that you determine the frequency of a sound, besides determining based on the spot in the cochlea where the excitation occurs. The fact that it works well at low frequencies helps make up for the larger errors in the location method at low frequencies. Most music involves fundamental frequencies lower than 1000 Hertz, which might be related to the limit mentioned above. Of course, it also might not.)

To test how well this works, do the same experiment as before using the second sound file provided, with frequencies of 400, 800, and 1200 Hertz.

Problems

You will notice two problems with these methods.

First, there is a gap between about 1000 Hertz and 2000 Hertz where neither method is very effective. Steady sounds in this frequency range are hard to localize. In nature, this is not that big of a problem, because most real sources of sound have a lot of overtone structure (many frequencies at once). Some will be above or below this range, and the more frequencies, the more pieces of information you have to determine direction with. Surprisingly many human-made devices for getting your attention (cell phones, emergency sirens) make sounds in this inefficient range, though. Recently manufacturers are starting to get smarter about this.

Second, in many environments, much or most of the sound you get is reflected sound, from the various objects around you. That means that sound will be approaching your head from different directions; the direct line between the sound source and you, the simplest path for a reflected sound, the next simplest path, and so on. Especially inside rooms and in crowded environments, this makes for trouble.

Now let us talk about

Sudden sounds

which you will remember from the first day of lecture, are quite easy to localize. This is done almost entirely by arrival time of the sound in the two ears, and to a lesser extent by echos involving the pinna and head shadow. To convince yourself of how easy it is to determine direction, play the third sound file in the same experiment as before. See how close together the speakers can be before you cannot tell which speaker a sound comes from.

There is still the problem of reflections. However, for a sudden sound, it can be solved in an interesting way. To see this, play the last sound file and try to figure out which speaker each sound came from. The answer is–it came from *both* speakers *each* time. However, it had a delay of,

1. 10 ms on the right
2. 5 ms on the left
3. 1 ms on the right
4. 50 ms on the left

For the last case, you must have noticed that the sound came from each speaker, but one was late. For the others, you may not have.

The ear *automatically ignores* localization information for about 35 milliseconds after a sudden sound. That is because this is a normal amount of time for echos from nearby objects to reach you. This corresponds to about 10 meters of extra travel distance for the reflected sound. (Think about our ancestors living in trees, or us living in rooms.) This is done subconciously, and prevents the brain from being confused by reflections. This matches our experience–when there was a loud noise in the front of the room, the people at the back turned forward, even though most of the sound they recieved was reflected sound from the walls, ceiling, floor, and back of the room.