Chapter 14
Localization of Sound

The ability to localize the source of a sound is important to the survival of human beings and other animals. Although we regard sound localization as a common, natural ability, it is actually rather complicated. It involves a number of different physical, psychological, and physiological, processes. The processes are different depending on where the sound happens to be with respect to the your head. We begin with sound localization in the horizontal plane.

14.1 Horizontal Plane

The horizontal plane is the plane parallel to the ground that includes your two ears and your nose. (It is a fact of geometry that it takes three points to define a plane.) We regard your head as a center point and think about sound sources located at other points in this plane. Those other points can differ in distance and azimuth with respect to your head. We begin with azimuth. The azimuth is an angle typically measured in degrees from the forward direction. Therefore, a point directly in front of your nose is at 0° azimuth. Points in space that are directly to your right and left have azimuths that are +90° and −90°, respectively, as shown in Fig. 14.1.

Modern approaches to sound localization in the horizontal plane are restricted to sources that lie in a within the 180° range from your extreme left (−90°) to your extreme right (+90°). The rest of the horizontal plane is behind you, and that becomes a matter to be discussed in connection with vertical plane localization. What is significant about this restricted range of locations in the horizontal plane is that different angles lead to signals that are uniquely different in your left and right ears (unless the source is directly in front). Differences between your ears (interaural differences) are processed by part of your nervous system called the "binaural system." The binaural system consists of several processing centers in the brain stem and midbrain that receive neural spikes from a processing chain that starts with the hair cells in the cochlea. These specialized binaural centers (olivary

W.M. Hartmann, *Principles of Musical Acoustics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-1-4614-6786-1_14,

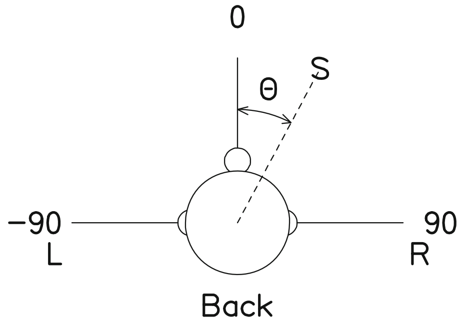

Fig. 14.1 A head and sound source (symbol S) in the horizontal plane. The source has azimuth θ with respect to the forward direction. A wavefront from the source arrives at the right ear before arriving at the left. Also, the head partially blocks the sound wave coming to the left ear causing it to be less intense at the left ear

complex, trapezoidal body, and part of the inferior colliculus) are particularly speedy and efficient. The upper levels of this binaural system are combined with visual centers that also encode location information. This combination allows the auditory and visual systems to learn from each other.

There are two main kinds of interaural differences, interaural level differences (ILDs) and interaural time differences (ITDs). We will deal with these in turn.

14.1.1 Interaural Level Differences

If a sound source is located off to your right, then it is closer to your right ear than to your left. Furthermore, a sound wave from the right has to go around your head somehow in order to arrive at your left ear. Thus, there are two effects causing the sound level to be higher in your right ear than in your left. The first is a distance effect and the second is a "head-shadow" effect. Both lead to an ILD.

It is not hard to show that unless a sound source is very close to your head the distance effect is not important. In Exercise 14.1 you will show that for a source that is 1 m away, the distance effect leads to an ILD no more than 1.4 dB. The effect is even smaller if the source is further away. But the distance effect is of obvious importance when the source is very close to your head, for instance when there is a mosquito in one of your ears.

To deal with the head-shadow effect without the distance effect we assume that the source is a few meters away. The head-shadow effect leads to ILDs that can be quite large, but the ILDs depend on frequency. Figure 14.2 shows the ILD for a source that is off to your right at an azimuth of 10°, 45°, or 90°. The vertical axis shows how much greater the level is in your right ear compared to your left. The figure shows that at low frequency there is hardly any effect at all. The ILD is essentially zero. That is because the low-frequency sound has a wavelength that is much larger than your head diameter, and the sound diffracts around your head, almost as though it were not there at all. By contrast, at high frequencies the wavelength is small, and your head represents a significant obstacle. At 4 kHz and 90° the ILD caused by head-shadow reaches 10 dB.

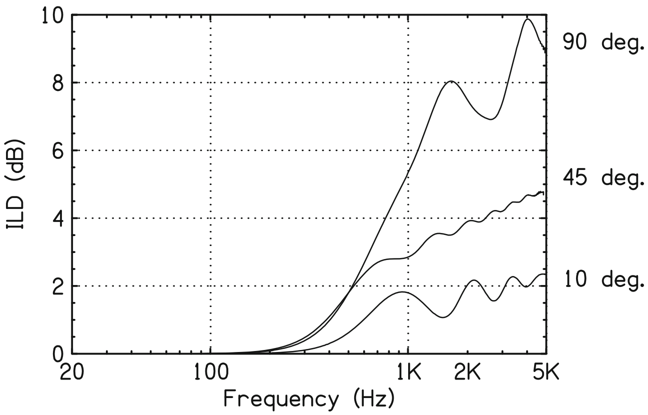

Fig. 14.2 The ILD for a sine tone is shown as a function of the frequency of the tone for three values of the azimuth, 10°, 45°, and 90°. The values were calculated from a equation for sound diffraction around a sphere having the diameter of the average human head. The oscillations in the plots reveal the complexity of the diffraction process. They can lead to some unexpected and confusing effects in controlled experiments

14.1.2 Interaural Time Differences

If a source of sound is located to your right, then every feature of the waveform arrives at your right ear before it arrives at your left ear. The difference in arrival time is known as the ITD. When you first think about it, it is difficult to believe that the ITD could have much effect. On the human scale, sound travels very fast, and the ears are quite close together. It is hard to imagine that the delay of a sound wavefront as it passes first one ear and then the other could be a large enough delay to be perceived. In fact, even for a sound at 90° azimuth (directly to your right) the delay in your left ear is less than 800 μs, equivalent to 0.8 ms. Strange as it may seem, the human binaural system is capable of dealing with interaural differences that small, *and much smaller*. It uses the differences to provide important information about the location of sounds. To understand how ITD localization works there are two things that we have to know. First, we need to know the relationship between the azimuth of the source and the ITD. Second, we need to know about the ability of the binaural system to perceive the ITD.

Source Azimuth and ITD

There is a formula, coming from a simplified theory of sound diffraction, that can be used to calculate the ITD for a source at azimuth θ:

$$\mathrm{ITD} = \frac{3\,r}{v}\sin(\theta), \tag{14.1}$$

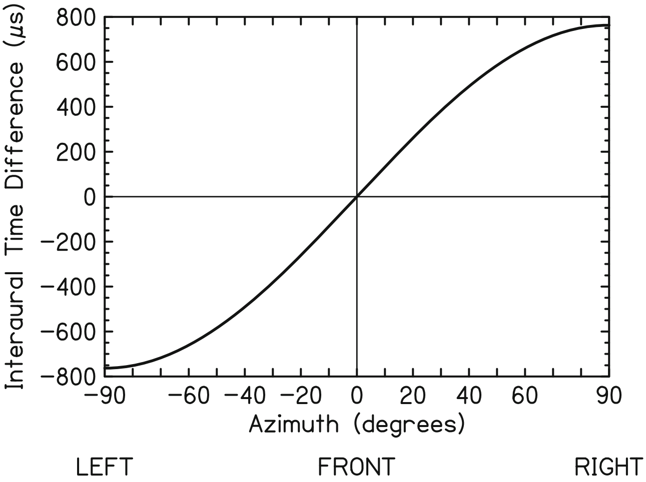

Fig. 14.3 The interaural time difference (ITD) as a function of azimuth from Eq. (14.1). The ITD is zero in the forward direction (azimuth 0°) because a source directly in front is equally distant from the two ears. For a negative value of the azimuth, the ITD is the negative of what it is for the corresponding positive azimuth value

where r is the head radius and v is the speed of sound. A typical human head radius is 8.75 cm, and the factor $3r/v$ is 763 μs. Because the sine function is never greater than 1.0, the ITD can never be greater than 763 μs. The ITD vs azimuth, according to Eq. (14.1), is shown in Fig. 14.3.

The Binaural System and the Usefulness of ITDs The ability of the human brain to make use of the ITD caused by the head is different for different azimuths. The system works best near an azimuth of zero (directly in front), where it is sensitive to an ITD as small as 20 μs, corresponding to a change of 1.5°. At 75° off to the side, the system is less successful. It is only sensitive to an angular difference of about 8°, but that is more precise than visual localization at that large azimuth.

There are important limitations in the ability of the binaural system to make use of ITD. If the frequency of a tone or noise is higher than 1,400 Hz, the waveform fine structure is too fast for the system to follow. Then the binaural system can only register time differences in the envelope (i.e., the amplitude), including the onset. Evidently, the use of ITD depends on the frequency range. To illustrate the different ways that ITD can be used, Fig. 14.4 shows a tone with an abrupt onset and a frequency glide or sweep—musically known as a “glissando.”

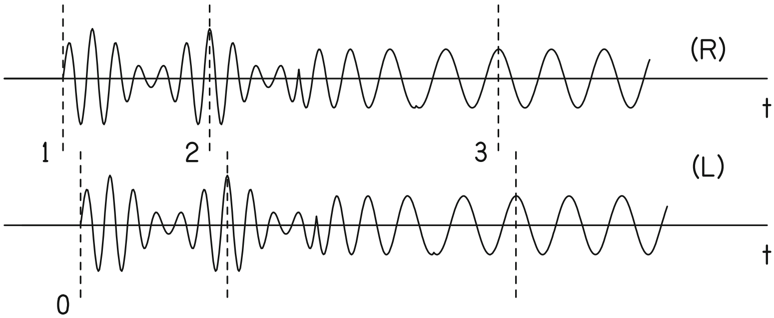

Fig. 14.4 The figure shows a high-frequency tone (1,800 Hz) that quickly sweeps to a low frequency (800 Hz). The high-frequency tone is modulated so its amplitude varies slowly. The source of the tone is 33° to the right. Therefore, from Eq. (14.1), the signal in the right ear leads the signal in the left by 417 μs. The binaural system can respond to this ITD in three ways, shown by numbers between the plots. (1) The onset occurs first in the right ear. (2) The structure in the modulation of the high-frequency waveform occurs first in the right ear. The 1,800-Hz waveform is too fast to observe the ordered structure in the waveform, but the ordered structure in the envelope (amplitude) can be seen. (3) The fine structure in the 800-Hz waveform occurs first in the right ear. Unlike the waveform fine structure at 1,800 Hz, the fine structure at 800 Hz is slow enough for the binaural system to measure

Binaural beats

Binaural beats occur when sine tones of slightly different frequency are separately sent to the left and right ears using headphones. Because the two frequencies are not physically combined there are no real beats, but there can be beats in the brain leading to a spatial effect. For instance, given a 500-Hz tone in the left ear and a 501-Hz tone in the right, most listeners hear a tone, with a pitch near 500 Hz, moving back and forth within the head once per second. If the frequencies are increased to 1,000 and 1,001 Hz, the frequencies are high enough that binaural effects are weak and the binaural beats are hard to hear.

To get a visual impression of the binaural effect, return to Fig. 6.4, which was drawn to explain real beats. Imagine that the solid curve represents the signal in your left ear and that the dashed curve represents the signal in your right. Notice that along the time line between 10 and 20 ms, the dashed curve seems to lead the solid curve. Now notice that between 80 and 90 ms the solid curve seems to lead the dashed. Your brain interprets this situation as a source first on the right and then on the left. (Recall that for visual purposes Fig. 6.4 has frequencies of 500 and 510 Hz leading to ten beats per second instead of one beat per second.)

Summary

The sections above have described horizontal (azimuthal) plane localization for steady-state sounds—the continuous parts of sine tones or complex tones with a simple spectral structure and no abrupt onset. Abrupt onsets lead to additional localization information that will be discussed in the section on the *precedence effect.* The interaural differences, ITD and ILD, in the waveform of a steady sound, or a sound with a slow onset, normally work together to localize the sound. The ITD seems to have the stronger influence, but the ITD has the limitation that it is only useful for frequencies below about 1,400 Hz. For tones above the not-particularly-high frequency of 1,400 Hz, the ILD is the main influence on localization, though some additional information is available from the ITD in the modulation of a time-varying signal, as shown in segment (2) of Fig. 14.4.

Sine tones without abrupt onsets have no modulation, as shown by segment (3) in Fig. 14.4. The usefulness of the ITD and the ILD in sine tones has a curious frequency dependence. At very low frequencies (below 200 Hz) the wavelength is so large compared to the size of the head that diffraction causes the levels in the two ears to be almost the same, and the ILD is too small to be useful. The ITD is usable at very low frequencies, though it is not particularly accurate because the period (greater than 5,000 μs) is so long that timing it is difficult. At low frequencies (400–900 Hz) the ITD process becomes highly effective and accurate. At quite high frequencies, above 2,000 Hz, the head provides a reliable shadow on the ear farther from the source, and the ILD is a useful cue. For intermediate frequencies (1,000–2,000 Hz) ITD information is weak or nonexistent. Also, head diffraction becomes a complicated function of both frequency and source azimuth causing the ILD information to be confusing. As a result, sine tone localization is worst in this intermediate frequency region.

14.2 Localization in the Vertical Median Plane

The vertical median plane is the plane that includes the points directly in front of you, directly in back of you, and directly overhead. Points in this plane are symmetrical with respect to your two ears. Therefore, sources in this plane lead to no important interaural differences. Nevertheless, you are able to localize sources in this plane. For instance, you can tell the difference between front and back. One reason that you can do this is that you can turn your head so that the source is somewhat off to the left or right. Then your ability to make left–right localization judgments coupled with your sense of motion enables you to distinguish front from back. However, if the sound is too brief for you to turn your head, or if you are unable to move your head, you can still tell front from back. That's because sounds from the front and back are differently filtered by the anatomy of your head. Your head does not have front–back symmetry (no nose on the back of your head), and this is especially true of your outer ears. The diffraction of short-wavelength sounds from these anatomical features leads to filtering of high-frequency sounds that is different for sounds coming from different directions in the median plane. The most important frequencies are above 6,000 Hz. Your brain has learned to recognize the spectral shapes that these anatomical filters impose on sound waves, and that is what enables you to localize sources in the vertical median plane.

The same kinds of spectral discriminations, in combination with the azimuthal, interaural localization processes, mediate the ability to distinguish between front and back and to perceive the elevation of sources off to the left or right. Experiments on front–back localization show large individual difference. Some people can distinguish directly in front from directly in back with noise bands near 6,000 Hz that are only a few kilohertz wide. Other people require very wide bands extending well above 10 kHz.

14.3 The Precedence Effect

The interaural differences and the spectral cues that are so important to sound localization are determined by the location of the sound source with respect to your head. Perhaps a coin is dropped on the table to your right or there is a talker to your left. The waves that come to your head directly from the source of the sound contain the useful localization cues.

A potential problem occurs when sounds are heard in a room, where the walls and other surfaces in the room lead to reflections. Because each reflection from a surface acts like a new source of sound, the problem of locating a sound in a room has been compared to finding a candle in a dark room where all the walls are entirely covered with mirrors. Sounds come in from all directions and it's not immediately evident which direction is the direction of the original source.

The way that the human brain copes with the problem of reflections is to perform a localization calculation that gives different weight to localization cues that arrive at different times. Great weight is placed on the information in the onset of the sound. This information arrives directly from the source before the reflections have a chance to get to the listener. The direct sound leads to localization cues such as ILD, ITD, and spectral cues that accurately indicate the source position. The brain gives much less weight to the localization cues that arrive later. It has learned that they give unreliable information about the source location. This weighting of localization cues, in favor of the earliest cues, is called the *precedence effect.*

There are a number of rather amazing audio illusions that demonstrate the precedence effect using electronic reverberation simulators. For instance, if one plays a recorded voice from the right loudspeaker of a left–right stereo pair and plays the reverberation of that sound from the left loudspeaker, a listener will hear the sound coming from the right, even if the sound from the left speaker is 8 dB more intense. The precedence effect operates here because the reverberant sound resembles the direct sound but arrives later.

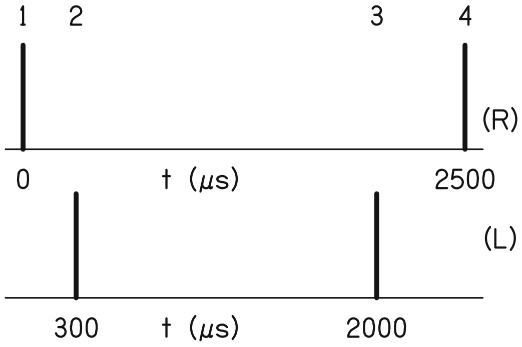

Fig. 14.5 A click pair to study the precedence effect: there are four pressure pulses, two at the right ear and two at the left. Pulses 1 and 2 form a *leading* click. Pulses 3 and 4 form a *lagging* click. However, the pulses come so close together in time that they sound like a *single click image*. (The time axis shows *microseconds*.) The leading click represents a direct sound arriving from the listener's right. The lagging click represents a reflected sound from the listener's left. Even though the reflected sound has an ITD ($-500\ \mu$s) that is larger in magnitude than the direct sound (300 μs), the direct sound wins the competition. The listener hears the sound coming from the right because of the precedence effect. The direct sound arrives first

The precedence effect is particularly easy to study with a click pair, as shown in Fig. 14.5. A click has all frequencies, high, medium, and low, and all of the anatomical localization processes are available. Experiments with click pairs have emphasized that the precedence effect relies importantly on high levels of the central nervous system. Infants do not have it—it is developed in childhood. Also, the precedence effect can be affected by context. A long series of click pairs with the leading click on the right, i.e., repeatedly presenting the click pair shown in Fig. 14.5, will bias the system. If that series is followed by a click pair with the leading click on the left, that last click will split into two images—one to the left and one to the right.

14.4 Perceived Auditory Space

Sound localization, as described earlier in this chapter, is an important perceived spatial effect, but it is not the only perceived spatial aspect of our auditory experiences. Other aspects include auditory image *externalization, distance perception,* and *compactness.*

Externalization: The burst of noise when you open a soda bottle is not only localized, it appears to be "out there" somewhere in the space where you live. It is "externalized." Similarly, an electronic burst of noise reproduced by a loudspeaker in a room leads to an externalized auditory image. If you take that same electronic burst of noise and send it to a listener wearing headphones (same noise signal into both left and right headphones), the listener will not perceive an externalized image. Instead, the listener tends to hear the noise within the head.

The distinction between externalized sounds and internalized sounds is important enough that auditory scientists use the term "lateralization," instead of the word "localization," to describe the perceived location of headphone-presented sounds heard within the head. Fortunately for the science of binaural hearing, it happens that there is a rather good correlation between localization and lateralization. Imagine a real sound source—somewhere out there—that produces interaural differences that cause the image to be localized half way between the front and the side. If the same interaural differences are produced using headphones, the image will be lateralized about half way between the center of the head and the extreme right. Therefore, much of what we know about sound localization was actually discovered in headphone experiments studying lateralization. Nevertheless, externalization is a separate perceptual phenomena, and for many years it was something of a mystery. The mystery was solved when auditory scientists began to study the detailed diffraction of sound waves by the human head, upper torso, and, especially, the external ears (pinna). These anatomical features put their own distinctive stamp on the frequency content of broadband sounds like noise bursts as a function of the relative location of the source. The different diffraction of different frequencies is a directionally dependent filter, and we listeners have got used to the features of that filtering by our own anatomy. When a sound does not exhibit the filtering of our own anatomy, our brains reject the concept that this sound is created by a real source somewhere out there. The appearance of the sound inside the head is apparently a default condition.

Perceived distance: Externalized sounds may be heard as close to us or far away. Understanding how we as listeners perceive the distance of a source is not easy. Of course, when there is a mosquito in your ear, there is a big ILD, but whenever the source is more than a few meters away from the head there are no useable interaural differences between distant sources and nearby sources. A first step in understanding distance perception is to realize that controlled experiments show that we are actually not very good at it. Sound intensity, perceived as loudness, is a cue because of the inverse square law, but when the intensity is randomized, we can't distinguish 2 m from 10 m or more. When sounds are heard in a room, the ratio of direct sound to reverberated sound is a useable distance cue. The ratio is smaller for distant sources, and this is particularly useful in comparing the distance of two different sources in the same room. In addition, very distant sounds are distinctive because the air itself particularly absorbs high frequencies. This filtering effect becomes noticeable when the path length is long.

Compactness: A compact image is perceived to be localized (or lateralized) at a point. The opposite of a compact image is a diffuse image. A diffuse image seems to be spread out in space; it may even surround us. The compact-diffuse distinction is particularly easy to study using headphones. In the early days of sound localization research, experimenters produced tones with ILDs that favored the right ear and ITDs that favored the left ear or vice versa. They wanted to see which interaural difference would win—lateralization to the right or the left? They found that conflicting interaural differences led to an image that was not compact. Sometimes it split in two, other times it became a fuzzy ball inside the head.

Compact images tend to occur when the signals in the two ears resemble one another. There may be an ILD or there may be an ITD, but if the signals are recognizably the same sort of thing on oscilloscope tracings, then a listener will tend to hear a compact image. Monophonic sound reproduction from a single loudspeaker leads to a compact image. Even in a room, a listener can accurately localize the loudspeaker because of the precedence effect. That is not a welcome fact when listening to recorded music. Stereophonic reproduction (two slightly different signals from two loudspeakers) leads to a more diffuse—more surrounding—experience because it helps to make the waves in the two ears to be more different. Similarly, it has been discovered that successful concert halls emphasize reflections from the sides which lead to significant differences in the signals received by the two ears, promoting the surround effect.

Exercises

Exercise 1, The distance effect on ILD

The distance between your ears is about 18 cm. Therefore, if a sound is 1 m away from your right ear, it can be no more than 1.18 m away from your left. (a) Use the inverse square law for sound intensities to show that the level in your right ear can be no more than 1.4 dB greater than the level in your left ear due to the distance effect alone. This calculation ignores the head-shadow effect. (b) Suppose that the source is 0.5 m away from your head. Show that the level difference is larger, namely 2.7 dB. This result generalizes. The closer the source, the greater the ILD caused by the distance effect.

Exercise 2, How big is your head?

(a) Why is it not possible to use the interaural level difference (ILD) to localize sounds with frequencies below 100 Hz? (b) If you wanted to make use of ILD at low frequency would you like your head to be larger or smaller than it is? Why?

Exercise 3, ILDs at 2,000 Hz

Use Fig. 14.2 to estimate the smallest detectable angular change for a 2,000-Hz source. Assume that the source azimuth is between 10° and 45°, and assume that a listener can detect an ILD change as small as 0.5 dB.

Exercise 4, Interaural time difference

(a) Show that Eq. (14.1) is dimensionally correct. (b) Calculate the ITD for azimuths of 30°, 45°, and 60°. Compare your answers with Fig. 14.3.

Exercise 5, Which plane?

Is localization better in the horizontal plane or the vertical plane? What would one mean by “better?”

Exercise 6, The front/back problem

How is it possible to tell whether a sound comes from in front of you or behind you? Such locations would seem to be symmetrical with respect to your ears and head.

Exercise 7, Older listeners

Older listeners have much more difficulty making front–back distinctions compared to left–right distinctions. Can you explain why?

Exercise 8, Precedence effect

What is the precedence effect? Why is it important for localizing sound in rooms?

Exercise 9, Personal experiences

Recall experiences you have had when you were confused about the true location of a source of sound. Sound localization is such a tricky business that everyone has had such experiences.

♠