2. Forward and backward masking show that the influence of a stimulus is *extended over time*, affecting the detection of stimuli presented after or before. This influence may reflect a persistence of neural activity after stimulus offset (and a build-up in response after onset), which can be modeled using a sliding temporal integrator or *temporal window*. Alternatively, forward masking may be a consequence of the reduction in sensitivity associated with neural adaptation.

3. We are extremely sensitive to repetitive fluctuations. We can detect amplitude modulation at rates up to 1000 Hz.

4. One pattern of modulation may interfere with the detection of another pattern of modulation, but not when the modulation frequencies are very different. We may have specific neural channels that are tuned to different rates of modulation and behave like a *modulation filterbank*.

5. The addition of frequency components in different regions of the spectrum, but with the *same pattern of modulation* as a masker, can improve our ability to detect a signal at the masker frequency. This finding may reflect a general ability to use coherent patterns of modulation across frequency to separate out simultaneous sounds.

6. Frequency modulation may be detected by the induced amplitude modulation in the excitation pattern, for modulation frequencies above about 10 Hz. At lower rates, the frequency excursions may be tracked by a (sluggish) mechanism based on phase locking.

7. Performance improves in many hearing tasks as the duration of the stimulus is increased. These improvements may result from a multiple-looks mechanism that combines several short samples of the stimulus, or from a long integration mechanism, which analyzes a continuous chunk of the stimulus. Flexible integration times may allow the auditory system to respond to rapid changes in a stimulus, and to integrate over longer durations when necessary.

8.5 READING

The ideas expressed in this chapter are developed further in:

Viemeister, N. F., and Plack, C. J. (1993). Time analysis. In W. A. Yost, A. N. Popper & R. R. Fay (Eds.), *Human psychophysics* (pp. 116–154). New York: Springer-Verlag.

Eddins, D. A., and Green, D. M. (1995). Temporal integration and temporal resolution. In B. C. J. Moore (Ed.), *Hearing* (pp. 207–242). New York: Academic Press.

9

Spatial Hearing

"Where is that sound coming from?" is a question our brains often pose to our ears. Most sounds originate from a particular place because the *source* of most sounds is a vibrating object with a limited spatial extent. There are a number of reasons why we would like to be able to locate the source of a sound. First, the location of the sound may be important information in itself. For example, did the sound of distant gunfire come from in front or behind? Second, the location of the sound may be used to orient visual attention: If someone calls your name you can turn to see who it is. Finally, sound location can be used to separate out sequences of sounds arising from different locations, and to help us attend to the sequence of sounds originating from a particular location (see Chap. 10). Location cues can help us to "hear out" the person we are talking to in a room full of competing conversations. Similarly, location cues help us to hear out different instruments in an orchestra or in a stereo musical recording, adding to the clarity of the performance.

I must admit that the visual system is about 100 times more sensitive to differences in source location than is the auditory system. However, our eyes are limited in their field of view, whereas our ears are sensitive to sounds from any direction and from sound sources that may be hidden behind other objects. Much information about approaching danger comes from sound, not light. This chapter describes

how the auditory system localizes sounds, and how it deals with the problems of sound reflection in which the direction of a sound waveform does not correspond to the location of the original source. We also consider how sound reproduction can be made more realistic by incorporating spatial cues.

9.1 USING TWO EARS

Binaural means listening with two ears, as compared to *monaural*, which means listening with one ear. The visual system features many millions of receptors, each responding to light from a particular location in the visual field. The auditory system has only two ears, but just as we get a complex sensation of color from just three different cone types in the retina, so our two ears can give us quite accurate information about sound location.

Figure 9.1 shows the coordinate system for sound direction, in which any direction relative to the head can be specified in terms of *azimuth* and *elevation*. Figure 9.2 illustrates the smallest detectable difference between the direction of sound sources in the horizontal plane (differences in *azimuth*). These thresholds are plotted as *minimum audible angles*, where 1° represents one 360th of a complete revolution around the head. For example, if the minimum audible angle is 5°, then we can just discriminate sounds played from two loudspeakers whose

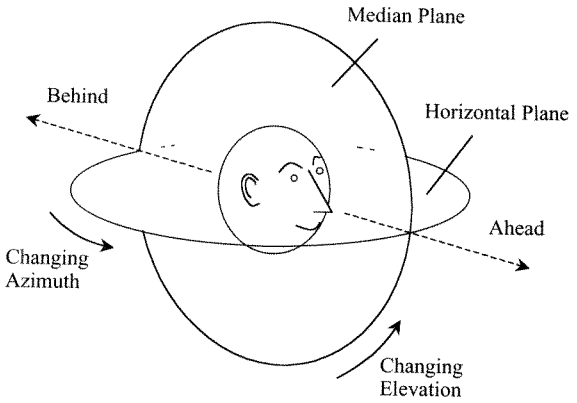

FIG. 9.1. The coordinate system for sound direction. The direction of a sound source relative to the head can be specified in terms of azimuth (the angle of direction on the horizontal plane; positive for leftward directions, negative for rightward directions) and elevation (the angle of direction on the median plane; positive for upward directions and negative for downward directions). A sound with zero degrees azimuth and zero degrees elevation comes from straight ahead. A sound with 90 degrees azimuth and 45 degrees elevation comes from the upper left, and so on. Adapted from Blauert (1997) and Moore (2003).

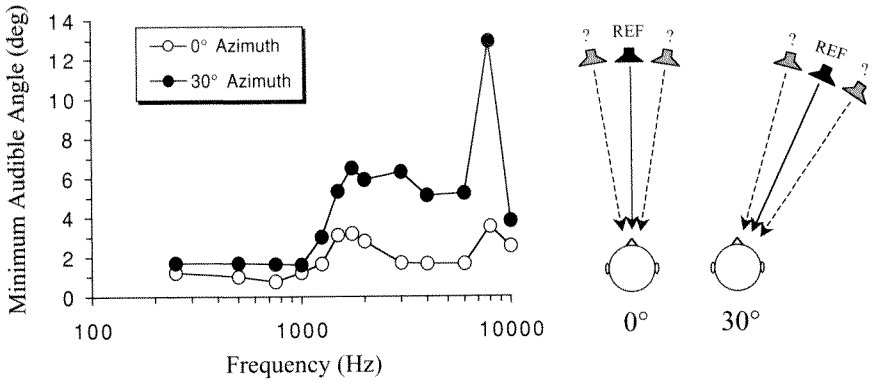

FIG. 9.2. The smallest detectable change in the direction of a pure-tone sound source in the horizontal plane, plotted as a function of frequency. The listener was presented with a sound coming from a reference location (labeled REF in the diagram to the right) and was then played a second sound from a loudspeaker slightly to the right or to the left. The listener had to indicate if the second sound was to the right or left of the reference. The *minimum audible angle* was defined as the angle between the reference and the second sound at which the listener chose the correct direction 75% of the time. The minimum audible angle was measured at two reference locations, straight ahead (0° azimuth) and to the side (30° azimuth). Data are from Mills (1958) cited by Grantham (1995).

direction differs by an angle of 5° with respect to the head. Note that our ability to discriminate pure tones originating from different directions depends on the frequency of the tones, with best performance at low frequencies. Note also that we have better spatial resolution if the sound source is straight ahead than if the sound source is to the side of the head. The minimum audible angle increases from about 1° for a sound straight ahead, to perhaps 20° or more for a sound directly to the right (−90° azimuth) or directly to the left (90° azimuth).

There are two cues to sound location that depend upon us having two ears, and involve a comparison of the sound waves arriving at the two ears. The first is the *time* difference between the arrival of the sound at the two ears, and the second is the difference in sound *level* between the two ears. We consider each of these cues in turn.

9.1.1 Time Differences

Imagine that you are listening to a sound that is coming from your right. Sound travels at a finite speed (330 meters per second in air), so that the sound waves will arrive at your right ear before they arrive at your left ear (see Fig. 9.3). A sound from the right will arrive at your right ear directly, but to reach your left ear it will have to *diffract* around your head (see Section 3.2.4). The time of arrival will depend

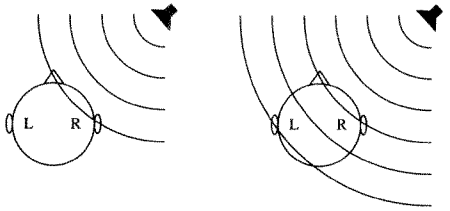

FIG. 9.3. A bird's eye view of the head with a sound source to the right. The curved lines show the peaks in the sound waveform at two consecutive instants. The sound waves arrive at the right ear before the left ear. This figure is a little misleading, because in the real world the sound would diffract *around* the head (see Section 3.2.4 and Fig. 9.4), further delaying the arrival time for the left ear.

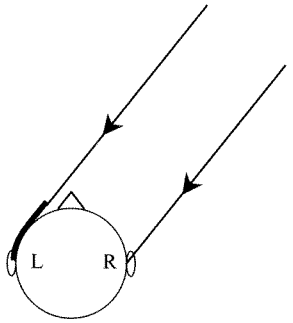

FIG. 9.4. The lines indicate the paths of sound waves arriving from a sound source far away and to the right. The thick line shows the path difference (the difference in the distance covered) between the sound waves arriving at the left ear and the sound waves arriving at the right ear. Based on Blauert (1997, Fig. 2.28).

upon the path length, which includes the distance the sound wave has to travel as it bends around your head (see Fig. 9.4). Differences in arrival time between the two ears are called *interaural time differences* (or *ITDs*). When low-frequency components are present (i.e., for most natural sounds) interaural time differences are the most important cues to sound location (Wightman & Kistler, 1992).

For a sound directly in front, directly behind, or anywhere in the vertical plane going through the center of the head (the *median* plane, see Fig. 9.1), the interaural time difference will be zero. The *maximum* interaural time difference will occur when the sound is either directly to the left or directly to the right of the head. The maximum time difference is only around 0.65 milliseconds for adult humans, which is the distance between the ears divided by the speed of sound. (The maximum

difference depends on the size of the head and is less for infants and much less for guinea pigs.) 0.65 milliseconds is a very short time, but it is much greater than the smallest *detectable* interaural time difference which is an amazing 10 microseconds, or 10 *millionths* of a second, for wideband noise in the horizontal plane (Klump & Eady, 1956). A shift between an interaural time difference of zero and an interaural time difference of 10 microseconds corresponds to a shift in sound location by about 1° in azimuth relative to straight ahead, which coincides with the smallest detectable direction difference (see Fig. 9.2). This remarkable resolution suggests that highly accurate information about the *time of occurrence* of sounds is maintained in the auditory system up to at least the stage in the ascending auditory pathways where the inputs from the two ears are combined (the superior olivary complex).

Interaural time differences may lead to ambiguity regarding location for some continuous sounds. Consider a continuous pure tone that is originating from a sound source directly to the right of the head. If the frequency of the tone is greater than about 750 Hz, the interaural time difference (0.65 milliseconds) will be greater than half a cycle of the pure tone. For a frequency a little above 750 Hz, a waveform peak at the left ear will be followed closely by a waveform peak at the right ear. Although the sound originates from the right, it may appear to the listener as if the sound waves are arriving at the *left* ear first (see Fig. 9.5). Fortunately, most natural sounds contain a wide range of frequency components, and they also contain *envelope* fluctuations. Envelope fluctuations are usually much slower than fluctuations in fine structure, so that the arrival times of envelope features can be used to resolve the ambiguity, even if the carrier frequency is above 750 Hz (see Fig. 9.5). Some measurements have used "transposed" stimuli (van de Par & Kohlrausch, 1997), in which the *envelope* of a high-frequency carrier is designed to provide similar information to the *fine structure* of a low-frequency pure tone (including matching the modulation rate to the pure-tone frequency). The smallest detectable interaural time difference for a transposed stimulus with a slowly varying envelope is similar to that of the equivalent low-frequency pure tone, even when the carrier frequency of the transposed stimulus is as high as 10,000 Hz (Bernstein & Trahiotis, 2002). This suggests that the mechanisms that process interaural time differences are similar (and equally efficient) at low and high frequencies.

Although the auditory system can discriminate stable interaural time differences with a high resolution, the system is relatively poor at tracking *changes* in the interaural time difference over time (which are associated with changes in the direction of a sound source). So, for instance, if the interaural time difference is varied sinusoidally (corresponding to a to-and-fro movement of the sound source in azimuth) then we can only track these changes if the rate of oscillation is less than about 2.4 Hz (Blauert, 1972). Because we can only follow slow changes in location, the binaural system is said to be "sluggish," especially in comparison with our ability to follow monaural variations in sound level up to modulation rates of 1000 Hz (see Section 8.2.1).

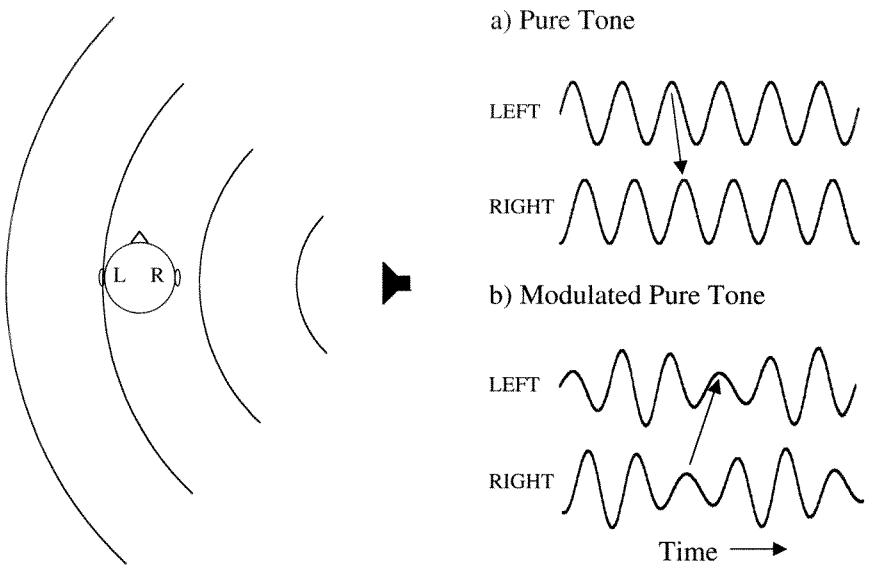

FIG. 9.5. A schematic illustration of the pressure peaks in space of a sound wave at an instant (left, ignoring diffraction effects) and the pressure variations at each ear as a function of time (right). In this illustration, the sound originates from the right, but waveform (fine-structure) peaks appear to occur at the left ear before the right ear (a). The ambiguity is resolved if the waveform is modulated, in which case it is clear that the sound waves arrive at the right ear before the left ear (b).

9.1.2 Level Differences

The other binaural cue to sound location is the difference in the *level* of a sound at the two ears. A sound from the right will be more intense at the right ear than at the left ear. *Interaural level differences* (*ILDs*) arise for two reasons. First, as described in Section 3.2.1, sound intensity decreases as the distance from the source increases. If the left ear is farther away from the source than the right ear, the level at the left ear will be less. In most circumstances, the width of the head is small compared to the distance between the head and the sound source, so the effect of distance is a very minor factor. Second, and of much more significance, the head has a "shadowing" effect on the sound, so that the head will prevent some of the energy from a sound source on the right from reaching the left ear. Low frequencies diffract more than high frequencies (see Section 3.2.4), so the low-frequency components of a sound will tend to *bend around the head* (see Fig. 9.4), minimizing the level difference. It follows that, for a sound source in a given location, the level difference between the ears will be greater for a sound containing mostly high-frequency components than for a sound containing mostly low-frequency components. For example, the

interaural level difference for pure tones played from a loudspeaker directly to the side of the head may be less than 1 dB for a 200-Hz tone, but as much as 20 dB for a 6000-Hz tone (see Moore, 2003, p. 236). The smallest *detectable* interaural level difference is about 1–2 dB (Grantham, 1984).

Interaural time differences work better for low-frequency pure tones (because of the phase ambiguity with high-frequency tones) and interaural level differences are greater for high-frequency tones (because there is less diffraction and therefore a greater head shadowing effect). It follows that these two sources of information can be combined to produce reasonable spatial resolution across a range of pure-tone frequencies (this is the "duplex" theory of Rayleigh, 1907). As we have seen, interaural time differences may provide useful information across the entire frequency range for more complex stimuli that contain envelope fluctuations. In addition, it is probable that interaural time differences are dominant in most listening situations, since most sounds in the environment contain low-frequency components.

9.1.3 Binaural Cues and Release From Masking

As well as the obvious benefits of binaural information to sound localization, we can also use binaural information to help us detect sounds in the presence of other sounds. As described in Section 9.1.2, if a high-frequency sound is coming from the right, then it may be much more intense in the right ear than in the left ear, and conversely if the sound is coming from the left. This means that if one sound is to the right and another sound is to the left, the right-side sound will be most detectable in the right ear and the left-side sound will be most detectable in the left ear. In some situations we seem to be able to selectively listen to either ear (with little interference between them) so that if sounds are separated in space, any level differences can be used to help hear them out separately.

More interesting perhaps is the use of interaural *time* differences to reduce the masking effect of one sound on another. Imagine that the same noise and the same pure tone are presented to both a listener's ears over headphones, so that the noise acts to mask the tone. In that situation, the smallest detectable level of the tone is little different to that detectable when the noise and tone are presented to one ear only. Now, keeping everything else the same, imagine that the phase of the tone is changed in one ear, so that it is "inverted": A peak becomes a trough and a trough becomes a peak. In this situation, the tone is much more detectable (see Fig. 9.6). The smallest detectable level of the tone drops by as much as 15 dB for low-frequency tones, although the effect is negligible for frequencies above 2000 Hz or so. The difference between the threshold measured when the stimuli are identical in the two ears, and the reduced threshold measured when the tone is inverted in one ear, is called a *binaural masking level difference*. Similar effects are obtained if the tone is delayed in one ear relative to the other, if the noise is delayed in one ear relative to the other, or if the tone is removed entirely from one

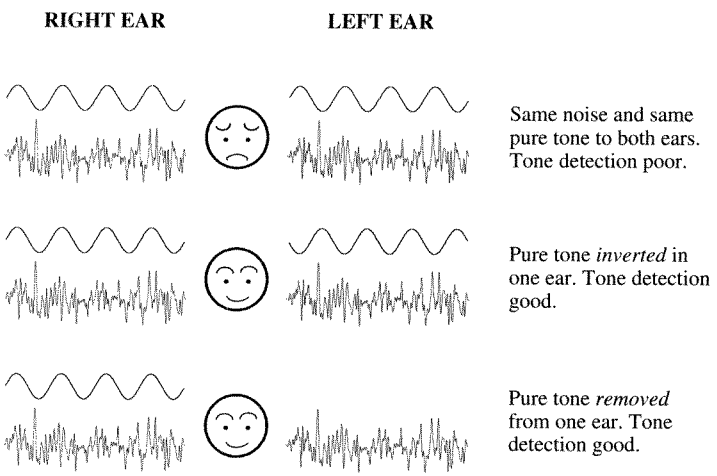

FIG. 9.6. Examples of listening situations in which a binaural masking level difference can be observed. Based on Moore (2003).

ear (see Fig. 9.6). With the exception of the last example, the level and magnitude spectrum are the same in the two ears in each case, but the small differences in interaural delay help the auditory system to separate the two sounds.

The effect also works for more complex stimuli, such as speech. If you are listening to someone speaking in an environment with competing sound sources, it can be easier to understand what they are saying with two ears than with one ear (Bronkhurst & Plomp, 1988). However, interaural time or level differences can separate two simultaneous speech sounds only when they also differ along some other dimension, such as fundamental frequency (Shackleton, Meddis, & Hewitt, 1994). When the two competing sounds are similar in all respects except location, the tendency is to *group them together* so that the sound heard is a fusion of the two (see Section 10.2.3).

It is possible that the use of interaural time differences for masking release is not directly related to their use in sound localization. Although in many cases the combination of interaural time differences in the masking release situation is similar to that which occurs when the two sounds originate from different locations in space, in some cases it is not (see Moore, 2003, p. 259–261). The largest masking release occurs when the signal is inverted in one ear with respect to the other. However, this only occurs in natural listening conditions for high frequencies (above 750 Hz, see Section 9.1.1), whereas the phase-inversion masking release effect (see middle panel of Fig. 9.6) is greatest for frequencies around 200 Hz (Blauert, 1997, p. 259). The results of masking release experiments suggest that time differences *per se* are aiding detection, not the fact that they may lead to the subjective impression that the masker and the signal come from different directions.

9.1.4 Neural Mechanisms

Jeffress suggested that interaural time differences are extracted by a *coincidence detector* that uses delay lines to compare the times of arrival at each ear (Jeffress, 1948, 1972). The Jeffress model has become the standard explanation for how the ear processes time differences. A simplified version of the Jeffress model is illustrated in Fig. 9.7. The model consists of an array of neurons, each of which responds strongly when the two inputs to the neuron are coincident. The clever bit is that each neuron receives inputs from the two ears that have been *delayed by different amounts* using neural delay lines (which might be axons of different lengths). For example, one neuron may receive an input from the right that is delayed by 100 microseconds relative to the input from the left. This neuron will respond best when the sound arrives in the right ear 100 microseconds before it arrives in the left ear (corresponding to a sound located about 10° to the right of straight ahead). For this neuron, the interaural time difference and the effect of the delay line cancel out, so that the two inputs to the neuron are coincident. Over an array of neurons sensitive to different disparities, the processing in the Jeffress model is equivalent to *cross-correlation*, in that the inputs to the two ears are compared at different relative time delays. You may have noticed the similarity between this processing and the hypothetical *auto*correlation mechanism for pitch extraction, in which a single input signal is compared with a copy of itself, delayed by various amounts (see Section 7.3.3). The arrival times at the two ears have to be specified very exactly by precise *phase locking* of neural spikes (see Section 4.4.4) to peaks in the waveform, for the mechanism to work for an interaural time difference of just 10 microseconds.

It is surmised that there is a separate neural array for each characteristic frequency. The locations of the different frequency components entering the ears can be determined independently by finding out which neuron in the array is most active at each characteristic frequency. The way in which the location information from the different frequency channels is combined is discussed in Section 10.2.3.

Is there any evidence that the Jeffress model is a physiological reality? There are certainly examples of neurons in the medial superior olive and in the inferior colliculus that are tuned to different interaural time differences. An array of these neurons *could* form the basis of the cross-correlator suggested by the Jeffress model, and, indeed, there is good evidence for such an array of delay lines and coincidence detectors in the brainstem of the barn owl (Carr & Konishi, 1990). However, the story may be very different in mammals. Recent recordings from the medial superior olive of the gerbil have cast doubt on the claim that there is a whole array of neurons at each characteristic frequency, each tuned to a different interaural time difference. Instead it is suggested that there is broad sensitivity to just *two* different interaural time differences at each characteristic frequency (McAlpine & Grothe, 2003). The binaural neurons in the left brainstem have a peak in tuning

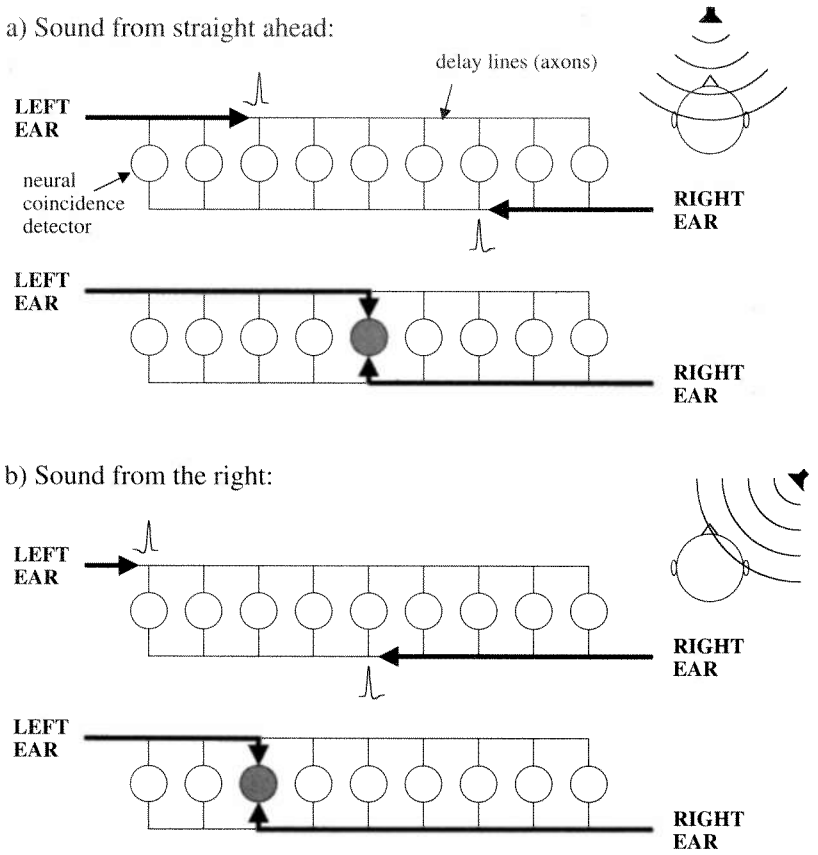

FIG. 9.7. An illustration of a popular theory of how interaural time differences are extracted by the auditory system. The circles represent neurons tuned to different interaural time differences (equivalent to different locations in the horizontal plane; neurons to the left of the array respond best to sounds from the right and *vice versa*). The thin lines represent the axons of neurons innervating the binaural neurons. Each panel shows two successive time frames. Panel (a): a sound from straight ahead arrives at the two ears at the same time, and the neural spikes from the two ears coincide to excite a neuron in the center of the array. Panel (b): a sound from the right arrives at the right ear first and the spikes from the two ears coincide at a different neuron in the array.

corresponding to an arrival at the right ear first, and the binaural neurons in the right brainstem have a peak in tuning corresponding to an arrival at the left ear first (see Fig. 9.8). Although these response peaks are at time differences outside the range that occur naturally for the gerbil, any real location may be derived from the relative firing rates of the two types of neuron. For example, if binaural

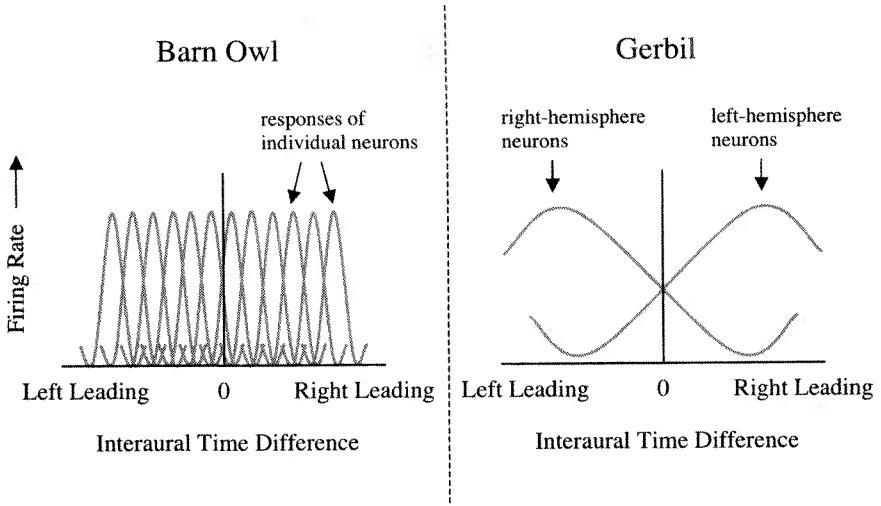

FIG. 9.8. How interaural time differences may be represented in the brainstem of birds and mammals. In the barn owl (left) at each characteristic frequency in each hemisphere of the brain there is an array of neurons tuned to different interaural time differences: Each neuron produces a maximal response when the sound arrives from a particular location. These are the array of coincidence detectors proposed by Jeffress (Fig. 9.7). In the gerbil, however, neurons in each hemisphere have a single broad tuning, responding maximally to a sound leading in the opposite ear. Location may be derived by comparing the firing rates of the neurons in the two hemispheres. Based on McAlpine and Grothe (2003, Fig. 1).

neurons in the right hemisphere fire more than those in the left hemisphere, then the sound is coming from the left. This is similar to the way in which the visual system represents color. Our ability to distinguish many thousands of different colors is based on the relative responses of just three different color sensitivities (three different cones) in the retina. The recent data suggest that the Jeffress model does *not* reflect the processing of time differences in humans.

With regard to interaural level differences, a different type of processing is involved, in which the relative levels at the two ears are compared. The interaural level difference produced by a sound source at a particular location varies with frequency (see Section 9.1.2), and this variation must be taken into account by the nervous system. Neurons that receive an excitatory input from one ear and an inhibitory input from the other ear have been identified in the lateral superior olive, in the lateral lemniscus, and in the inferior colliculus (see Møller, 2000, page 256). A neuron that receives inhibitory input from the left ear is most sensitive to sounds from the right, and *vice versa*. Some of these neurons may be responsible for extracting location by comparing the sound levels in the two ears.

9.2 ESCAPE FROM THE CONE OF CONFUSION

Interaural time differences and interaural level differences are important cues for sound location, but they do not specify precisely the direction from which the sound comes in three-dimensional space. For example, any sound on the median plane (see Fig. 9.1) will produce an interaural time difference of zero, and an interaural level difference of zero! We cannot use either of these cues to tell us whether a sound comes from directly in front, directly behind, or directly above. More generally, for each interaural time difference, there is a cone of possible sound source locations (extending from the side of the head) that will produce that time difference (see Fig. 9.9). Locations on such a "cone of confusion" may produce similar interaural level differences as well. There must be some additional cues that enable us to resolve these ambiguities and to locate sounds accurately.

9.2.1 Head Movements

Much of the ambiguity about sound location can be resolved by moving the head. If a sound source is directly in front, then turning the head to the right will decrease the level in the right ear and cause the sound to arrive at the left ear before the right ear (see Fig. 9.10). Conversely, if the sound source is directly behind, then the same head movement will decrease the level in the *left* ear and cause the sound to arrive at the *right* ear first. If the head rotation has no effect, then the sound source is either directly above or (less likely) directly below. By measuring carefully the effects of known head rotations on time and level differences, it should be possible

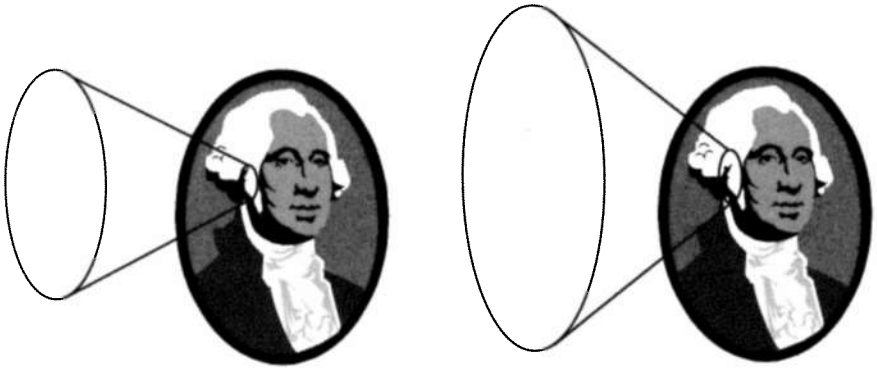

FIG. 9.9. Two cones of confusion. For a given cone, a sound source located at any point on the surface of the cone will produce the same interaural time difference. The cone on the left is for a greater time difference between the ears than the cone on the right. You should think of these cones as extending indefinitely into space.

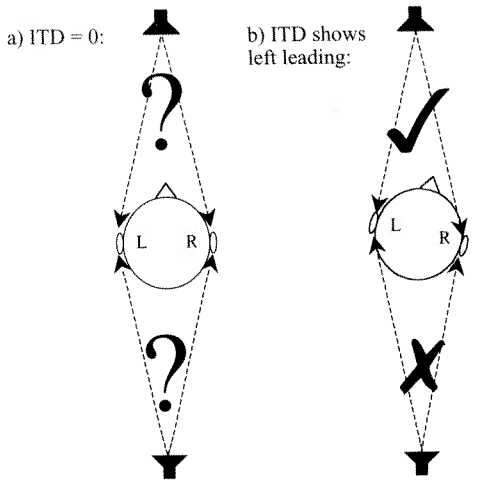

FIG. 9.10. An example of the use of head movements to resolve location ambiguities. A sound directly in front of the listener (a) produces the same interaural time difference and interaural level difference as a sound directly behind. The effect of turning the head on the interaural time difference (b) reveals the true situation.

(theoretically) for the auditory system to specify the location of any sound source, with the only ambiguity remaining that of whether the source elevation is up or down. If head rotations in the median plane are also involved (tipping or nodding of the head), then the location may be specified without any ambiguity.

It seems to be the case that listeners make use of head rotations and tipping to help them to locate a sound source (Thurlow, Mangels, & Runge, 1967). However, these movements are only useful if the sound occurs for long enough to give the listener time to respond in this way. Brief sounds (for example, the crack of a twig in a dark forest!) may be over before the listener has had a chance to make the head movement.

9.2.2 Monaural Cues

Although we have two ears, a great deal of information about sound location can be obtained by listening through just one ear. *Monaural* cues to sound location arise because the incoming sound waves are modified by the head and upper body and, especially, by the *pinna*. These modifications depend on the *direction* of the sound source.

If you look carefully at a pinna, you will see that it contains ridges and cavities. These structures modify the incoming sound by processes including *resonance* within the cavities (Blauert, 1997, p. 67). Because the cavities are small, the resonances only affect high-frequency components with short wavelengths (see Section 3.1.3). The resonances introduce a set of spectral peaks and notches in

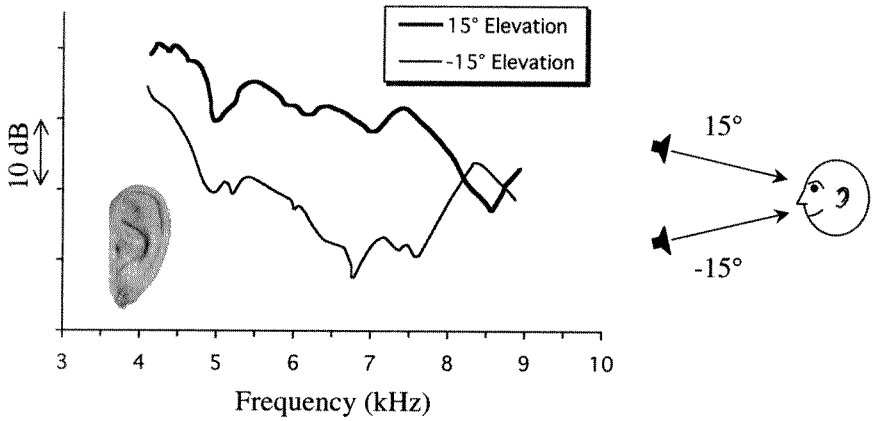

FIG. 9.11. Spectra of a broadband noise recorded from a microphone inserted in the left ear canal of a single listener. The noise was presented from a loudspeaker at an elevation of either −15° or 15° relative to the head. The spectra demonstrate the direction-specific filtering properties of the pinna. From Butler and Belendiuk (1977).

the high-frequency region of the spectrum, above 4000 Hz or so. The precise pattern of peaks and notches depends on the *angle* at which the sound waves strike the pinna. In other words, the pinna imposes a sort of directional "signature" on the spectrum of the sound that can be recognized by the auditory system and used as a cue to location. As an example, Fig. 9.11 shows recordings, made from a human ear canal, of a broadband noise presented from two elevations in the median plane. The effects of variations in elevation can be seen in the spectra.

Pinna effects are thought to be particularly important for determining the elevation of sound sources. If the cavities of the pinnae are filled in with a "soft rubber," localization performance for sound sources in the median plane declines dramatically (Gardner & Gardner, 1973). There is also a shadowing effect of the pinna for sounds behind the head, which have to diffract around the pinna to reach the ear canal. The shadowing will tend to *attenuate* high-frequency sound components coming from the rear (remember that low frequencies diffract more than high frequencies, and so can bend round the obstruction), and may help us to resolve front–back ambiguities.

9.3 JUDGING DISTANCE

Thus far we have focused on our ability to determine the *direction* of a sound source. It is also important in some situations to determine the *distance* of a sound source (is the sound of screaming coming from the next room or the next house?). Overall sound level provides a cue to distance for familiar sounds. In an open

space, every doubling in distance produces a 6 dB reduction in sound level (see Section 3.2.1). If the sound of a car is very faint, then it is more likely to be two miles away than two feet away. However, when listeners are required to estimate the distance of a familiar sound based on level cues, then the response tends to be an *underestimate* of the true distance, so that a 20 dB reduction in sound level is required to produce a *perceived* doubling in distance (see Blauert, 1997, p. 122–123). The use of the level cue depends on our experience of sounds and sound sources. If a completely alien sound is heard, we cannot know without additional information whether it is quiet because the sound source is far away or because the sound source is very weak. Generally, however, loud sounds are perceived as being close, and quiet sounds are perceived as being far away, even though these perceptions may at times be misleading.

Another cue to distance in rooms with reflective walls, or in other reverberant environments, is the ratio of direct to reverberant sound. The greater the distance of the source, the greater is the proportion of reflected sound. This cue can be used by listeners to estimate distance, even if the sound is unfamiliar. However, our limited ability to detect changes in the direct-to-reverberant ratio implies that we can only detect changes in distance greater than a *factor of two* using this cue alone. The direct-to-reverberant ratio provides only coarse information about the distance of a sound source (Zahorik, 2002b). In any natural listening environment, the auditory system will tend to *combine* the level information and the direct-to-reverberant ratio information to estimate distance. It is still the case, however, that for distances greater than a meter or so, we seem to consistently underestimate the true distance of a sound source when we rely on acoustic information alone (Zahorik, 2002a).

Finally, large changes in distance tend to change the *spectral balance* of the sound reaching the ears. The air absorbs more high-frequency energy than low-frequency energy, so the spectral balance of sounds far away is biased toward low frequencies, as compared to sounds close by. Consider, for example, the deep sound produced by a distant roll of thunder compared to the brighter sound of a nearby lightning strike. However, the change in spectral balance with distance is fairly slight. The relative attenuation is only about 3–4 dB per 100 meters at 4000 Hz (Ingard, 1953). Spectral balance is not a particularly salient cue, and is useless for small distances.

9.4 REFLECTIONS AND THE PERCEPTION OF SPACE

9.4.1 The Precedence Effect

When we listen to a sound source in an environment in which there are reflective surfaces (for example, a room), the sound waves arriving at our ears are a complex combination of the sound that comes *directly* from the sound source, and the

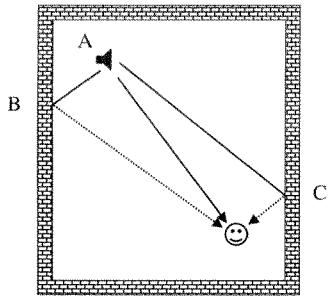

FIG. 9.12. Listening to a sound source in a reflective room. To localize the sound source correctly (A), the listener must ignore sound waves that appear to originate from the direction of the points of reflection (e.g., B and C).

sound that is reflected, perhaps many times, by nearby surfaces. The problem with reflections, in terms of localizing sound sources, is that the reflected sound provides directional information that *conflicts* with that from the direct sound. If heard in isolation, reflected sound waves would appear to come from the direction of the reflective surface, rather than from the direction of the sound source. This problem is illustrated in Fig. 9.12. How does the auditory system know that the sound waves come from the sound source (A) and not from the direction of one of the numerous sites of reflection (e.g., B and C)? The task appears difficult, yet human listeners are very good at localizing sounds in reverberant environments.

To avoid the ambiguity, the auditory system uses the principle that the direct sound will always arrive at the ears *before* the reflected sound. This is simply because the path length for the reflected sound is always longer than the path length for the direct sound. The *precedence effect* refers to our ability to localize on the basis of information from the leading sound while ignoring the information from the lagging sound. The precedence effect has been demonstrated using sounds presented over headphones and from loudspeakers (see Litovsky, Colburn, Yost, & Guzman, 1999 for a review). In the latter case, the listener may be facing two loudspeakers at different locations in the horizontal plane, but equidistant from the head. One loudspeaker acts as the source of the "direct" sound, and the other loudspeaker is used to simulate a reflection (see Fig. 9.13). If identical sounds are played through the two loudspeakers without any delay, then the sounds are "fused" perceptually, and the location of the sound source appears to be midway between the two loudspeakers. If the sound from the second loudspeaker is delayed by up to about a millisecond, then the sound source appears to move toward the first loudspeaker, because the interaural time differences between the ears suggest that the sound is arriving at the ear closest to the first loudspeaker before it arrives at the ear closest to the second loudspeaker. If, however, the sound from the second loudspeaker is delayed by between about 1 and 30 ms (these times are for complex

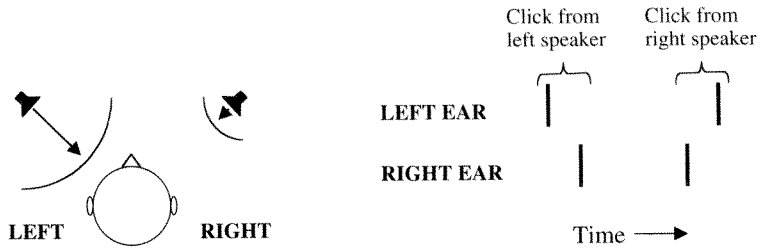

FIG. 9.13. A typical experimental configuration for observing the precedence effect. A click from the left loudspeaker arrives at the left ear before the right ear. Subsequently, a click is played from the right loudspeaker and arrives at the right ear before the left ear. If the click from the right loudspeaker is delayed by 1–5 ms, then a single sound source is heard, located in the direction of the left loudspeaker. The auditory system treats the click from the right loudspeaker as a reflection (echo) of the click from the left loudspeaker, and the location information from the second click is suppressed. Based on Litovsky et al. (1999, Fig. 3).

sounds such as speech; for clicks the upper limit is about 5 ms), then the sound appears localized almost *entirely* to the first loudspeaker. The simulated echo from the second loudspeaker is effectively ignored for the purposes of localization. For larger delays, the percept breaks down into two sound sources located at the two loudspeakers, each of which can be localized independently. In other words, the precedence effect only works for fairly short reflections (corresponding to path length differences of about ten meters). The precedence effect also works best when the sound contains large envelope fluctuations, or transients such as clicks, so that there are clear time markers to compare between the direct and reflected sounds. Stable tones with slow onsets are localized very poorly in a reverberant room (Rakerd & Hartmann, 1986).

Moore (2003, p. 256) describes how the precedence effect can be a nuisance when listening to stereo recordings over loudspeakers. The stereo location of an instrument in a recording is usually simulated by adjusting the relative levels in the left and right channels during mixing (i.e., by using a cue based on interaural *level* differences). If the guitar is more intense in the left loudspeaker than in the right loudspeaker then it sounds as if it is located towards the left. However, the music is being played out of the two loudspeakers *simultaneously*. This means that the relative time of arrival of the sounds from the two loudspeakers depends on the listener's location in the room. If the listener is too close to one loudspeaker, such that the sound from that loudspeaker arrives more than 1 ms before the sound from the other loudspeaker, the precedence effect operates and all the music appears to come from the closer loudspeaker alone! Moore suggests that if you venture more than 60 cm either side of the central position between the two loudspeakers, then the stereo image begins to break down.

Although the precedence effect shows that we can largely ignore short-delay reflected sound for the purposes of *localization*, when we are in an environment we are well aware of its reverberation characteristics. We can usually tell whether we are in the bathroom or the bedroom by the quality of the reflected sound alone. The smaller bathroom, with highly reflective walls and surfaces, usually has a higher level of reverberation at shorter delays than the bedroom. Furthermore, we see in Section 9.3 that the level of reverberation compared to the level of direct sound can be used to estimate the distance of the sound source. It follows that the information about reflected sounds is not erased, it is just not used for localization.

9.4.2 Auditory Virtual Space

The two main cues to sound location are interaural time differences and interaural level differences. By manipulating these cues in the sounds presented to each ear, it is possible to produce a "virtual" source location. As noted, most stereophonic music is mixed using interaural level differences to separate the individual instruments. Realistic interaural time differences cannot be easily implemented for sounds played over loudspeakers, because each ear receives sound from *both* loudspeakers. Realistic time and level differences can be introduced into sounds presented over *headphones*, because the input to each ear can be controlled independently. However, the sounds often appear as if they originate from within the head, either closer to the left ear or closer to the right ear. Sounds played over headphones are usually heard as *lateralized* within the head, rather than *localized* outside the head. This is because the modifications associated with the head and pinnae are not present in the sound entering the ear canal. It seems that we need to hear these modifications to get a strong sense that a sound is external.

It is possible to make stereo recordings by positioning two microphones in the ear canals of an artificial "dummy head" that includes model pinnae. In this case, the spectral modifications of the head and pinnae are included in the recordings. When these recordings are played over headphones, listeners experience a strong sense of the sound source being localized *outside* the head. *Head-related transfer functions*, which mimic the modifications of the pinna and can also include interaural time and interaural level differences, can be used to process a sound recording so that it elicits a realistic external spatial image when played over headphones. Since pinnae vary from one individual to the next, it is perhaps unsurprising to discover that head-related transfer functions work best when they are derived from the ears of the person being tested (Wenzel, Arruda, Kistler, & Wightman, 1993).

Although you may not be aware of it at all times, your brain processes the reverberation characteristics of the space you are in, and uses this information to obtain an appreciation of the dimensions and reflectivity of that space. It follows that for sounds presented over headphones or loudspeakers reverberation is also important for the simulation of a natural listening environment. Recordings made in an anechoic (non-reverberant) room sound unnatural and give a strange sensation

of the acoustic space. Such a recording may be processed to simulate the delayed reflections that might occur in a more natural space. By progressively increasing the delays of the simulated reflections, it is possible to make someone sound as if they are singing in a bathroom, in a cathedral, or in the Grand Canyon. While not all of these scenarios may be appropriate, a considered use of reverberation can benefit a recording greatly. Sophisticated "reverb" devices are considered vital tools for sound engineers and music producers.

9.5 SUMMARY

There are several different types of information about sound source location available to the auditory system, and it appears that our brains combine these different cues. Interaural time and level differences provide detailed information about direction, but this information is ambiguous and accurate localization depends on the use of other types of information, such as the effects of head movements and the location-dependent modifications of the pinna. Our appreciation of the acoustic space around us is dependent on information about source location, and the information about reflective surfaces from the characteristics of reverberation.

1. Having two ears helps us to determine the direction of a sound source. There are two such binaural cues: *interaural time differences* (a sound from the right arrives at the right ear first), and *interaural level differences* (a sound from the right is more intense in the right ear). Interaural time differences are dominant for most natural sounds.
2. Interaural time differences are most useful for the localization of low-frequency components. If the wavelength of a continuous pure tone is less than twice the distance between the ears, ambiguity arises as to whether the sound is leading to the left or right ear. However, time differences between peaks in the (slowly varying) envelope can be used at these higher frequencies.
3. Interaural level differences arise mainly because of the shadowing effect of the head. Low frequencies diffract more than high frequencies, so that the level cue is most salient at high frequencies.
4. Binaural cues can help us to separate perceptually sounds that arise from different locations. However, the use of interaural time differences for masking release may be independent from their use in localization.
5. Our remarkable sensitivity to interaural time differences (minimum threshold of 10 microseconds) implies very accurate phase-locked encoding of the time of arrival at the two ears. The information is extracted by neurons in the brainstem that receive input from both ears and are sensitive to differences between the arrival times at each ear.

6. Interaural time and level differences do not unambiguously specify the location of the sound source. The "cone of confusion" can be resolved by head movements, and by the use of monaural information based on the effects of the *pinna*, which imposes a direction-specific signature on the spectrum of sounds arriving at the ear.

7. The most salient cues to the distance of a sound source are *level* (because quiet sounds are usually from sound sources that are further away than the sources of loud sounds) and the *ratio of direct to reflected sound* (which decreases with increasing distance). The auditory system combines these cues, but tends to underestimate.

8. In a reverberant environment, the ear is able to identify the direction of the sound source by suppressing the (misleading) location information from reflections. We still perceive the reverberation, however, and this provides information regarding the dimensions and reflective properties of the walls and surfaces in the space around us.

9. Sounds presented over headphones can be made to sound external and more realistic by simulating the filtering characteristics of the pinnae. Similarly, the addition of appropriate reverberation helps produce the impression of a natural space.

9.6 READING

You may have noticed that I found Blauert's book very helpful for this chapter:

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization.* Cambridge, MA: MIT Press.

I also recommend:

Gilkey, R. H., & Anderson, T. A. (Eds.). (1997). *Binaural and spatial hearing in real and virtual environments.* New Jersey: Lawrence Erlbaum Associates.

Grantham, D. W. (1995). Spatial hearing and related phenomena. In B. C. J. Moore (Ed.), *Hearing* (pp. 297–345). New York: Academic Press.

10

The Auditory Scene

Professor Chris Darwin recently wrote: "How often do you hear a single sound by itself? Only when doing psychoacoustic experiments in a sound-proof booth!" (Darwin, 2005). Unless you are reading this in a very quiet place, the chances are that you will be able to identify sounds from several sources in the space around you. As I write these words in a study in my house, I can hear the singing of birds in the garden, the rustling of trees in the wind, and (somewhat less idyllically) the whirr of the fan on my laptop computer. Our ears receive a mixture of all the sounds in the environment at a given time: The sound waves simply add together when they meet (Fig. 10.1). As you might imagine, in a noisy environment such as a party or a busy street, the result can be very messy indeed! To make sense of all this, the auditory system requires mechanisms that can *separate out* the sound components that originate from different sound sources, and *group together* the sound components that originate from the same sound source. Bregman has termed the whole process *auditory scene analysis* (Bregman, 1990).

It is arguable that scene analysis is the hardest task accomplished by the auditory system, and artificial devices are nowhere near human performance in this respect. To explain why, I'll resort to a common analogy. Imagine that you are paddling on the shore of a lake. Three people are swimming on the lake, producing ripples that combine to form a complex pattern of tiny waves arriving at your feet. By