CHAPTER 10

Auditory Spatial Layout

Frederic L. Wightman
Rick Jenison

I. INTRODUCTION

Everyday sights and sounds are typically described with reference to the environmental object that produced them and not to the physical pattern of stimulation at the sensory receptor. Thus, we say that we see a house rather than an array of points and edges and that we hear a bell rather than a complex of inharmonic partials. This object-oriented view of perception has come to be known as *object perception*. In the case of vision the physical features of environmental objects map directly to patterns of stimulation on the retina. Quite naturally, then, the study of visual object perception concentrates on revealing the details of further processing of the peripheral representation, on such issues as size and shape invariance under various transformations of the retinal image. In contrast, hearing offers no direct peripheral representation of environmental objects. All auditory sensory information is packaged in a pair of acoustical pressure waveforms, one at each ear. While there is obvious structure in these waveforms, that structure (temporal and spectral patterns) bears no simple relationship to the structure of the environmental objects that produced them. The properties of auditory objects and their layout in space must be derived completely from higher level processing of the peripheral input. Thus, many of the issues

Perception of Space and Motion

central to the study of auditory object perception are different from those involved in visual object perception.

The definition of what constitutes an auditory object is an issue of some controversy and considerable importance. Many acoustical waveforms evoke a mental reference to the source of the waveform. These are clearly auditory objects. We hear a church bell, for example, or ice tinkling in a glass. We hear the objects themselves and are generally unaware of the spectral and temporal structure of those waveforms. However, reference to an identifiable physical object may not be a necessary condition for auditory "objectness." As we mention later, waveforms made up of sequences of pure tones can also contain what most would agree are primitive auditory objects, even though no known physical object could have produced the sounds.

That the study of auditory object perception is immature is reflected in the fact that there are few empirical data on the important issues. Thus, while we can be precise here in our descriptions of the physical features of auditory stimuli and somewhat certain about the details of the peripheral encoding of those features, discussion of the higher level processing that subserves auditory object formation and segregation must be speculative. In the context of our discussion of the spatial layout of auditory objects, for example, we can and do review the substantial body of evidence on the factors that determine the apparent spatial positions of single, static sound sources. However, since there are relatively few data on the perception of moving sources and virtually no data on perception of the spatial relations among auditory objects, our treatment of these important issues is limited to an analysis of the potential sources of information and does not attempt to address in detail the questions related to how those sources of information may be utilized.

The chapter begins with a discussion of the peculiarities of acoustical stimuli and how they are received by the human auditory system. A distinction is made, following Gibson (1966), between the ambient sound field and the effective stimulus to differentiate the perceptual distinctions among various simple classes of sound sources (ambient field) from the known perceptual consequences of the linear transformations of the sound wave from source to receiver (effective stimulus). Next we deal briefly with the definition of an auditory object, specifically the question of how the various components of a sound stream become segregated into distinct auditory objects. The remainder of the chapter focuses on issues related to the spatial layout of auditory objects. Stationary objects are considered first. Since much of the material relevant to this subject has recently been reviewed elsewhere (e.g., Middlebrooks & Green, 1991; Wightman & Kistler, 1993), the section concentrates on topics not covered in those previous reports. The sources of information related to the apparent distance of an auditory

object is one such topic. The spatial layout of moving auditory objects is discussed next, and in this context we offer a detailed treatment of the acoustics of moving sound sources. A distinction between source movement and observer movement is made to draw attention to the possible role of proprioceptive feedback in the perception of auditory spatial layout. The chapter concludes with a brief treatment of experimental evidence on the importance of input from other senses (vision, primarily) in establishing auditory spatial layout.

II. ACOUSTICAL INFORMATION: THE AMBIENT SOUND FIELD AND THE EFFECTIVE STIMULUS

As we use the term here, *information* is an abstract construct that serves as the bridge between an organism and its environment. It has a structure that is not related to the characteristics of either the transmitting medium or the receptor surface. For example, the "squareness" of a visual object is specified by information (e.g., relationships among visual patterns) that is not defined in terms of the physics of light or the anatomy and physiology of the retina. In the case of auditory objects, the mechanical events that produce them have lawful acoustical consequences in the sound patterns that are represented to the peripheral auditory system. If those patterns map in a one-to-one or many-to-one fashion onto the object properties, then they constitute information that potentially specifies those properties. In principle, then, for any physical property of an environmental object to be recoverable by an organism there must be information available to the perceiver that specifies that property.

The specific property of auditory objects that is of interest here is spatial layout. The information about auditory spatial layout is acoustically conveyed, and thus the stimulus that must be decoded by the perceiver to determine spatial layout is a sound wave. There is information about spatial layout contributed both by the specific type of sound wave that is generated and by the transformations that sound waves undergo in their passage from the source to our ears. This section of the chapter provides an overview of the broad classes of simple sound sources and the characteristics of the waves they produce (the ambient field), and also in this section there is a detailed discussion of the source-to-receiver transformations that convey information about the spatial layout of the sound sources (the effective stimulus).

A. The Ambient Sound Field

Waves in general are important means by which information about a physical event is conveyed to a perceiver. Discussion of wave generation and

propagation is beyond the scope of this chapter since both are extraordinarily complex topics, especially in the case of naturally occurring physical events and natural environments. Simplifying assumptions are not only useful but mandatory for our purposes here. In the case of sound-producing events, a convenient assumption is that the sound is produced by a so-called *point* source, or acoustic monopole, and that the propagation equations are linear. Any small object vibrating in a mass of fluid (air) has all the attributes of an acoustic monopole, provided the dimensions of the object are small relative to the sound wavelengths produced and the sound field of interest is several object lengths away. The sound field produced by a monopole is omnidirectional, that is, the same in any direction equidistant from the source.

The sound fields produced by two or more simultaneously active monopoles can be assumed to combine linearly. Thus, an acoustic *dipole*, a very common type of sound source in nature, can be described as the superposition of two spatially separated monopole sources that are 180° out of phase. In contrast with monopole sources, which are omnidirectional, dipole sources have both magnitude and orientation. The structure of the dipole field can best be understood by considering the dipole in terms of its canceling monopoles. The field has an angular dependence with no sound at all produced at 90° to the dipole axis where the sound fields of the constituent monopoles exactly cancel.

The intensity of a sound wave (proportional to pressure squared per unit area) diminishes as the wave travels away from the source. Several factors are responsible for this. One that applies to all sound waves, including those proposed by monopoles and dipoles, is atmospheric absorption. Absorption is the result of nonadiabatic propagation caused by temperature differentials between compressions and rarefactions in the propagating wave and in air depends on temperature, humidity, and wavelength. The attenuation coefficient in air at 20°C with 50% humidity is approximately $1 \times 10^{-10} f^2/\mathrm{m}$, where f is frequency in Hz. For a monopole source, intensity also decreases with the inverse square of the distance from the source because the total acoustical power is spread out over the surface area of a sphere, the radius of which is the distance from the source. When considering both geometrical spreading and absorption, the intensity (I) of a monopolar source as a function of distance can be written

$$I(r) = \frac{P}{4\pi r^2} e^{-\alpha r},$$

where r is the distance from the sound source, P is the total power produced by the source, and α is the attenuation coefficient. Sometimes the term *attenuation length*, $1/\alpha$, is used to describe the distance over which the inten-

sity decreases to $1/e$. At short distances the decrease in intensity with distance is dominated by spherical spreading, whereas at distances well beyond the attenuation length, absorption is dominant.

The intensity of the sound field produced by a dipole decreases somewhat differently with distance. For a dipole field it is simplest to discuss the decrease in pressure (proportional to the square root of intensity). The equation governing the pressure decrease is complicated, but its essential elements are a magnitude and a direction component. The magnitude part has two terms: one decreasing with the inverse square of distance, and the other linearly. The inverse square dependence dominates the field near the source, and the linear component dominates at large distances.

The characteristics of sound radiation, whether modeled as a monopole or as a dipole, may contribute significant information to aid source identification and to determine spatial layout. As described above, monopoles radiate sound evenly in all directions, but dipoles have a figure-eight directivity pattern. While the compression and rarefaction components cancel in a plane perpendicular to the dipole axis, a pressure gradient does exist in the field near the source that may be useful for tracking a sound source. An example of a dipole source that we are particularly interested in tracking is a flying insect near our ear. There are also more complex sources in nature that can be modeled as the sum of several constituent dipoles.

B. The Effective Stimulus

For our purposes here the effective stimulus is defined in terms of the acoustical pressure waveforms produced by an ambient sound field as they exist just before transduction at the listener's eardrums. For simplicity we assume that the ambient field is produced by one or more acoustical monopoles. The relationship between the ambient field and the effective stimulus is defined by a series of linear transformations of the acoustical waveform that incorporate a number of potential sources of information about the spatial layout of sound sources in the environment. In this section of the chapter we identify the relevant transformations and describe the spatial information that each incorporates. In a later section we examine in detail the evidence on whether the information is perceptually relevant.

The acoustics of the local environment that includes the source and the listener contribute several potentially important sources of information about spatial layout. For example, because of the long wavelengths and slow propagation velocity of sound, the reflections and diffractions of an emitted sound wave off the walls, floor, ceiling, and contents of a typical room enrich the ambient sound field considerably. There is information about the size of the room in the timing of the reflections, information about the wall coverings and contents in the pattern of reverberation, and information

about the distance between source and listener in the ratio of direct to reflected sound. If long distances are involved, such as in large rooms or in open spaces, the high-frequency content of the effective stimulus is reduced by atmospheric absorption. There is ample evidence that all of these effects are detectable by a normal-hearing listener.

The listener's shoulders, head, and outer ear structures (especially the pinnae) are significant components of the local acoustical environment and as such contribute additional information relevant to auditory spatial layout. The pattern of reflections and diffractions of an incident sound wave off these structures is heavily dependent on the direction from which the sound arrives, and thus, the information contributed by these effects relates primarily to the direction of auditory objects. The pinnae, in particular, are highly directional, modifying incident sound waves in ways that are specific to each different angle of incidence. As in the case of room effects, there is ample evidence of the detectability of pinna effects.

The fact that we have two ears separated by an acoustically opaque head suggests that information about auditory spatial layout may come from three sources: the effective stimulus at the left ear, the effective stimulus at the right ear, and the difference. These are clearly not independent sources of information. However, there are reasons to believe that all are important. Information from the difference signal, for example, is uniquely independent of the characteristics of the source, and because of the insensitivity of the auditory system to the absolute timing of events, this is the only source of information on the direction-dependent difference in the time-of-arrival of an acoustic waveform. Because of the approximate lateral symmetry of the head, interaural difference information is ambiguous. Interaural time difference, for example, is the same for sources in the front and sources in comparable positions (on the same side of the head, and at the same angles relative to the interaural axis) in the rear. Information from each of the individual ears can potentially resolve these ambiguities.

The information relevant to auditory spatial layout that is contained in the effective stimuli at the two ears can be described as either temporal or spectral patterns. At a formal mathematical level the two descriptions are isomorphic, so one might think the choice is arbitrary. However, when higher level processing of the information is considered, the distinction becomes important because temporal and spectral processing mechanisms in the auditory system are thought to be so different. For this reason, we discuss temporal and spectral separately. Because of the auditory system's relative insensitivity to monaural phase (the phase spectrum of a stimulus at one ear), our discussion of temporal information concentrates on interaural time differences and the temporal patterns of room reflections. Interaural phase, defined as the difference between the phase spectra of the left and

right ear stimuli, is relevant only when considering single-frequency components of a stimulus. Our discussion of the spectral information in effective auditory stimuli focuses on the direction-dependent changes in the magnitude components of the complex source-to-eardrum transformation.

III. AUDITORY OBJECTS

It seems obvious that before any discussion of the rules that govern the spatial layout of auditory objects, we should know what an auditory object is. Unfortunately, there is little consensus on what might constitute a satisfactory definition of an auditory object nor on what alternative terms might be better. One alternative that has been proposed is *sound event* (Blauert, 1983), but this term seems to refer more directly to a disturbance of the ambient sound field than to any aspect of the perception of that disturbance. Another alternative is *sound stream* (Bregman, 1990), but this term does not convey the obviously close association between everyday auditory stimuli and the environmental objects that produced them. The term *auditory object* is borrowed from the field of visual perception in which the features of environmental objects map directly to features of the effective stimulus, a pattern of light on the retina. Its use in auditory perception is less satisfying since there is no straightforward mapping of object features to stimulus features. Nevertheless, the fact that auditory percepts in daily life are so naturally and immediately associated with the objects that produced the sounds is undeniable and gives currency, if not clarity, to the term *auditory object*.

The effective stimulus at each ear consists of a one-dimensional acoustical pressure waveform. This waveform contains the superposition of the acoustic outputs from all of the objects in the listener's environment. A complete understanding of what constitutes an auditory object would therefore include specification of the rules, whereby the various components of the single-pressure waveform are segregated into discrete auditory objects. These rules are the object of considerable current interest in the auditory research community (e.g., Bregman, 1990; Handel, 1989), and it is not our purpose to summarize them here. Rather, we focus on the contributions to this segregation process offered by spatial separation. For the purposes of our discussion, it may be helpful to distinguish between two kinds of auditory objects: *concrete* and *abstract*. Concrete auditory objects are formed by sounds emitted by real objects in the environment. Although experimental data are scarce, segregation of concrete objects seems to be primarily determined by spatial and temporal rules. Abstract auditory objects do not often correspond to real environmental objects. They consist typically of more primitive sound elements and are formed by simpler frequency and tempo-

ral relations. There has been considerable research on the rules governing the formation of abstract auditory objects (e.g., Bregman, 1990). We concentrate here exclusively on concrete auditory objects.

IV. SPATIAL LAYOUT OF STATIONARY AUDITORY OBJECTS

Much of the experimental literature on auditory spatial layout concerns the accuracy with which the spatial position of a sound-producing object is indicated to a listener, that is, the degree of correspondence between the actual position of the object and its apparent position. It is our view that experiments that focus on accuracy can fail to consider other important features of the auditory percept. For example, consider experiments on monaural listening. The results generally show that the apparent positions of auditory objects are strongly biased toward the interaural axis and the side of the functioning ear. However, those same results are often reported as indicating that monaural localization accuracy is near normal on the side of the functioning ear and progressively poorer off the interaural axis on that side. The emphasis on accuracy obscures the fact that in monaural listening all of the sounds appear to emanate from one place. For reasons such as this, we prefer to ignore the accuracy component of spatial layout altogether, and we discuss only the factors that govern the apparent spatial positions of auditory objects.

The apparent spatial position of an auditory object is defined by its apparent direction and its apparent distance relative to the listener. The potential sources of information for apparent direction and the stimulus features that appear to govern apparent direction have extensively and recently been discussed elsewhere (Middlebrooks & Green, 1991; Wightman & Kistler, 1993). Therefore, the material on apparent direction is only summarized here. Much less attention has been paid to apparent distance, and although data are scarce, they are covered in some detail in this chapter.

A. Acoustical Sources of Information about Static Spatial Layout

The spatial position of each sound-producing object in a listener's environment is specified by several acoustical sources of information that for brevity we call *cues*. Many of the cues are a result of the interactions of the sound waves with the listener's head and pinnae. These interactions are conveniently summarized by a linear transformation, the so-called *head-related transfer function* (HRTF), which represents the changes in the amplitude and phase of the sound wave from the sounding object's position to the listener's eardrum. Mathematically, HRTFs are usually specified in terms of the sound wave's spectrum. Thus, if $X(j\omega)$ is the source spectrum (j is the

complex operator and ω is angular frequency) and $Y(j\omega)$ is the spectrum of the waveform at the eardrum, then the HRTF, $H(j\omega)$, is given by

$$H(j\omega) = \frac{Y(j\omega)}{X(j\omega)}. \tag{1}$$

More generally, since the HRTF varies with source direction and distance and thus is different at each ear, we must write two equations for $H(j\omega)$: one for the left ear and one for the right ear. Each depends on source azimuth (θ), elevation (ϕ), and distance (d) relative to the listener:

$$H_l(\theta, \phi, d, j\omega) = \frac{Y_l(\theta, \phi, d, j\omega)}{X(j\omega)} \tag{2}$$

and

$$H_r(\theta, \phi, d, j\omega) = \frac{Y_r(\theta, \phi, d, j\omega)}{X(j\omega)}. \tag{3}$$

All of the information about sound source position are represented in the pair of HRTFs shown above. These HRTFs vary in complicated ways with changes in source position, so simplifying assumptions must be made to appreciate the essential elements. Two convenient assumptions are that the acoustical space enclosing the source and listener is anechoic and that the listener's head is spherical with pinna-less ears at opposite ends of a diameter of the sphere. The anechoic assumption allows the main effect of distance to be modeled as a simple attenuation of 6 dB for every doubling of distance from the source. The spherical head assumption leads to a greatly simplified account of the effects of diffraction of the sound wave around the head. Figure 1 illustrates the latter point. When ignoring the details for a moment (the spherical model is described in detail in Kuhn, 1977), we see that at each ear variations in source azimuth (or elevation, not shown in the figure) can be expected to produce mainly variations in effective stimulus intensity, a result of the *head shadow* effect when the source is on the opposite side of the head from the ear under consideration. The head shadow effect can be expected to be much larger at high frequencies than at low frequencies. This is because at low frequencies sound wavelengths would be long with respect to the dimensions of the head, and thus the sound waves would travel around the head without attenuation. The covariation of stimulus intensity with azimuth (and elevation) that occurs at each ear individually can be viewed as a potential *monaural cue* to sound source position. Figure 1 also illustrates the potential *binaural cues* to sound source position that are offered by interaural differences (defined by the ratio of the two HRTFs). Note that for all source azimuths other than 0° and 180°, the acoustical path from

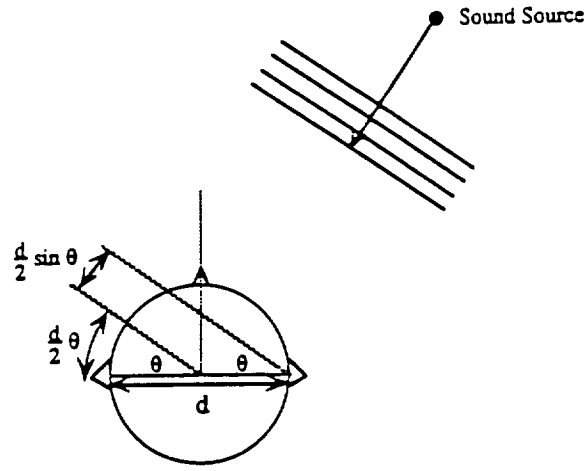

FIGURE 1 Schematic top-down representation of a listener and a sound source. The source is assumed to be sufficiently far from the listener such that the acoustical wavefronts are planar, and the listener is assumed to have a spherical head with ears at opposite ends of a diameter.

source to ear has a different length for the two ears. This path-length difference produces a small difference in the time of arrival of the sound wave at the two ears. The interaural time difference (ITD) varies systematically with source azimuth and is largest for azimuths of $+90°$ and $-90°$. In addition, because of the head shadow effect mentioned earlier, there will be an interaural level difference (ILD) that varies with azimuth in roughly the same way as ITD and is large at high frequencies and small at low frequencies.

The utility of monaural cues is compromised by the fact that some or all features of the sound source waveform must be known for the cue to be unambiguous. In the simple spherical head case described above, while stimulus intensity at a given ear varies systematically with source azimuth, a listener with access only to the effective stimulus at that ear would have no way of knowing whether a weak stimulus was produced by a source on the opposite side of the head or by a weak source. In more general terms, note that (from Equation 3) the effective stimulus at one ear, say the right ear, is defined by the product of the source spectrum and the HRTF:

$$Y_r(\theta,\phi,d,j\omega) = X(j\omega)H_r(\theta,\phi,d,j\omega). \tag{4}$$

Thus, even if a listener had perfect memory for the HRTF at each and every possible source position, a given effective stimulus could unambiguously indicate a specific source position only if the source spectrum were known.

Binaural cues to source position are derived from the ratio of the transduced representations of the two effective stimuli. Thus, the utility of these cues does not require knowledge of the source spectrum since that term

appears in both numerator and denominator and hence cancels. Nevertheless, to the extent that the spherical head model is accurate, binaural cues are also ambiguous. Note, as shown in Figure 1, that the difference in acoustical path length from the source to the two ears, which gives rise to the ITD, is the same for sources in front and in the rear. A source at an azimuth of 30°, for example, would produce the same ITD as a source at 150° azimuth. The same could be said for ILDs and for sources at complementary positions above and below the horizontal plane. In fact, the spherical head model predicts conical surfaces projecting outward from the ears along which ITD and ILD are constant and thus along which cues that are based on ITD and ILD would be ambiguous. These are the so-called *cones of confusion*. We should mention here that cone-of-confusion ambiguities could be resolved by head movements, as Wallach (1940) pointed out in his now-classic treatise on the issue. If a listener knew both the direction of movement of the head and the direction of change of the ITD or ILD cue, the direction of the sound source could be derived without ambiguity.

Detailed measurements of human HRTFs (Middlebrooks & Green, 1990; Middlebrooks, Makous, & Green, 1989; Pralong & Carlile, 1994; Shaw, 1974; Wightman & Kistler, 1989a) provide a complete catalog of the potential acoustical cues to apparent sound position and highlight the limitations of the spherical head model. The most prominent features of HRTFs not anticipated by the spherical head model are the directional filtering characteristics of the pinnae and the large listener-to-listener differences in HRTFs. The multiple ridges and cavities of the pinna produce resonant peaks and antiresonant notches in the magnitude response of the HRTF. The frequencies at which these peaks and notches appear are dependent on sound source direction and thus could serve as potential spatial position cues, provided some a priori information about the sou ce was available. Figure 2 shows an example of how the frequency of a given notch in the HRTF changes with sound source elevation. HRTFs from two listeners are shown in this figure to illustrate individual differences. Note that while the general characteristics of the notches are the same from listener to listener, the frequencies at which the notches appear are highly listener dependent.

The spherical head model provides a reasonably accurate prediction of the ITDs derived from actual HRTF measurements. Figure 3 shows ITDs from the horizontal plane HRTFs of a representative listener estimated by Wightman and Kistler (1989a). Also plotted in the figure are the ITDs predicted by

$$\mathrm{ITD} = \frac{d}{2c}(\theta + \sin\theta), \tag{5}$$

where θ is the azimuth angle as in Figure 1, c is the velocity of the sound wave (cm/s), and d is the interaural distance (cm), chosen for this example to fit the HRTF data shown. While this equation is usually cited as repre-

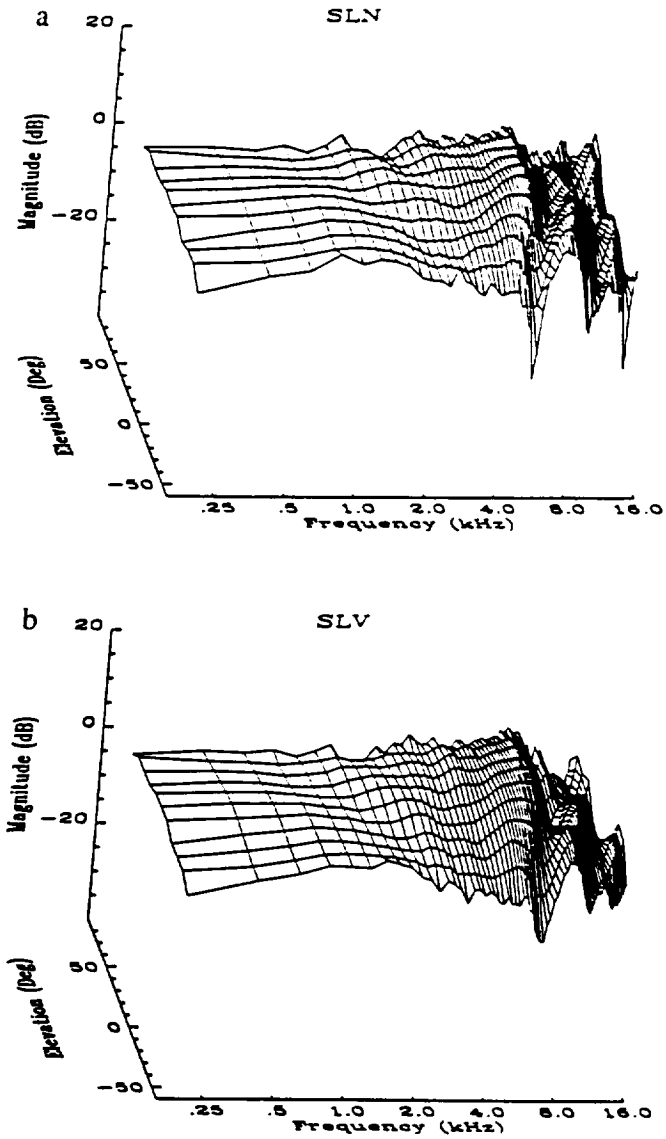

FIGURE 2 Directional transfer functions from two listeners produced by a source at 90° azimuth. Directional transfer functions (DTFs) are head-related transfer functions (HRTFs) divided by the root-mean-square average of the HRTFs from all spatial positions measured. Thus, the DTFs represent the deviation in dB from the average response of the ear. (Adapted with permission from Wightman and Kistler, 1993.)

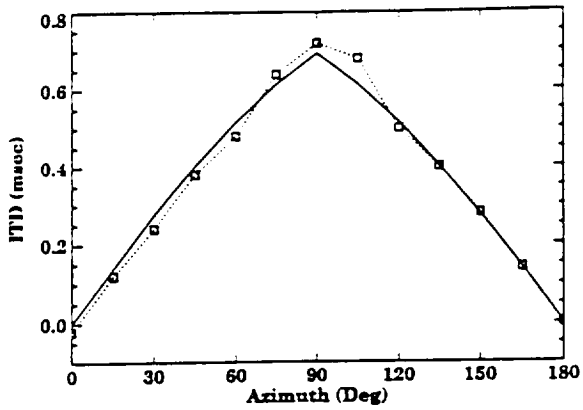

FIGURE 3 Interaural time differences (ITDs), produced by a source at 0° elevation, predicted by the spherical head model (solid line) and measured from a typical listener by using a wideband correlation technique. (Reproduced with permission from Wightman and Kistler, 1993.)

senting the predictions of the spherical head model (e.g., Green, 1976; Woodworth, 1938), it is really just a first-order approximation (Kuhn, 1977). Nevertheless, as Figure 3 shows, it provides an accurate representation of horizontal plane ITDs. Figure 4 (from Wightman & Kistler, 1993) shows a more complete set of ITD data from the same listener. This figure also shows the contours of constant ITD, which for the spherical head model would be circular. Clearly, the spherical head model provides a good first-order approximation to measured ITDs. Just as clearly, ITD is an ambiguous cue to sound source direction since any given ITD signals not one but a whole locus of potential directions.

Interaural level differences derived from HRTF measurements are complicated functions of frequency at each and every source direction, a situation caused at least in part by pinna filtering effects. Figure 5 shows ILD functions derived from a single listener's HRTF measurements at a source elevation of 0 and azimuths of 0° and 90°. Note that even for a source on the median plane (0° azimuth), where ILDs would result only from interaural asymmetries, ILDs are large enough (greater than 0.5 dB, the ILD threshold) to be considered potential sources of information about source position. For a source at 90° ILDs are generally much larger, especially at high frequencies as would be expected because of head shadowing.

The elaborate frequency dependence of ILDs complicates our discussion of them as potential cues to sound source position. We can discuss the interaural level cue either as an *interaural spectral difference*, referring to the entire pattern of ILDs across frequency, or as ILD averaged across one or more frequency bands. Figure 6 illustrates the latter approach. In the upper

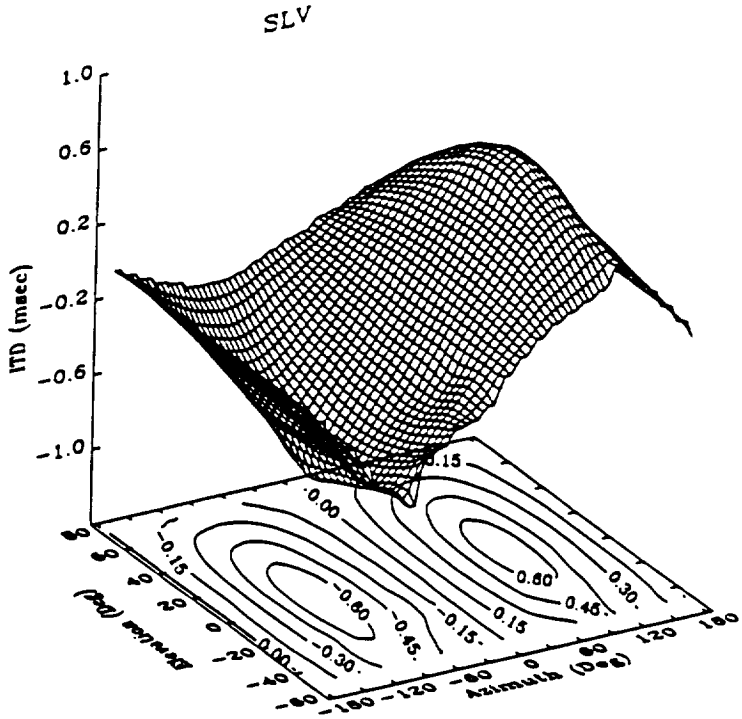

FIGURE 4 Interaural time differences (ITDs) from head-related transfer function (HRTF) measurements from a typical listener plotted as a function of the azimuth and elevation of the sound source. Note the contours of constant ITD below the surface plot. (Adapted with permission from Wightman and Kistler, 1993.)

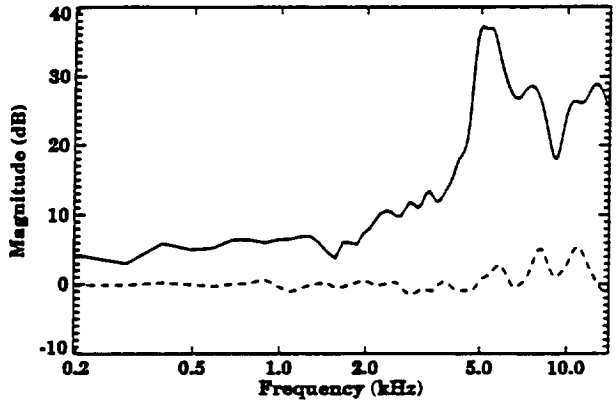

FIGURE 5 Interaural level difference (ILD) as a function of frequency from a typical listener, produced by a source at 0° elevation and 0° azimuth (dashed line) or 90° azimuth (solid line).

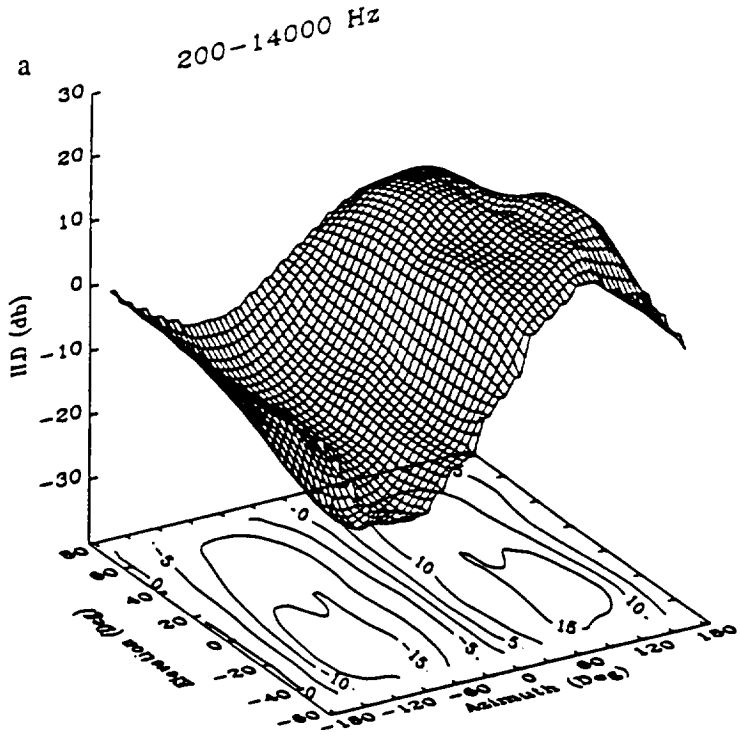

FIGURE 6 Interaural level difference (ILDs) from a typical listener in different frequency regions. Figure 6a shows ILDs across the entire frequency spectrum, and Figures 6b and 6c show ILD in two high-frequency critical bands. (Adapted with permission from Wightman and Kistler, 1993.)

panel we show one extreme, ILD averaged across the entire frequency spectrum. The bottom panels illustrate the other extreme, ILDs in two high-frequency critical bands. Note that the general pattern of ILD as a function of sound source direction is the same regardless of the bandwidth over which ILD is considered or the center frequency of the band. Note also that the general pattern of ILDs is the same as the pattern of ITDs, showing a similar kind of cone-of-confusion ambiguity. Thus, unless a listener could analyze the idiosyncratic details of ILD patterns in narrow bands, ILD information could not be used to disambiguate errors resulting from dependence on ITDs, and vice versa. As mentioned above, information provided by head movements can, in theory, offer such disambiguation.

The acoustical sources of information about the distance of a sound-producing object are not well understood. Nor have they been well documented by systematic measurements. In an anechoic environment, the two most obvious stimulus features that depend on distance are overall level and

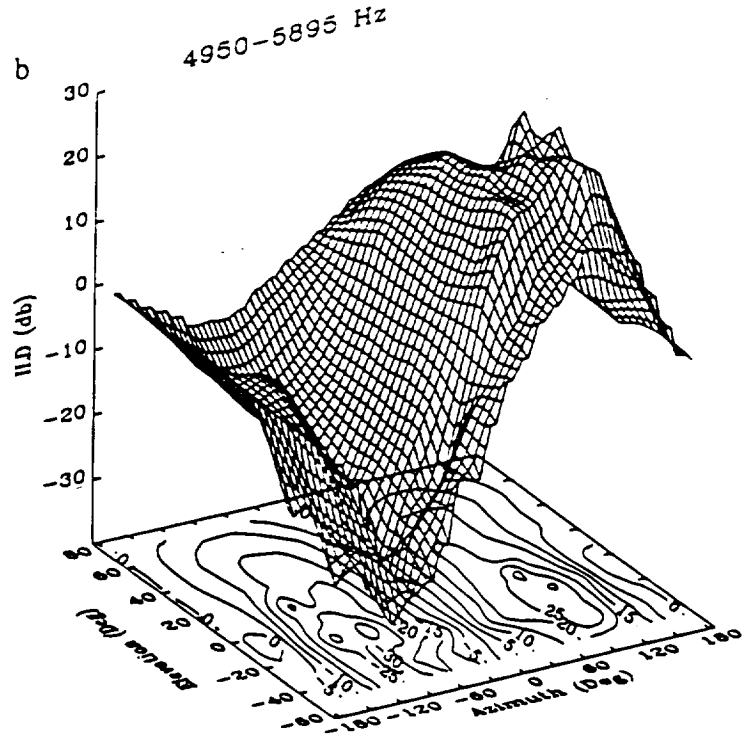

FIGURE 6 *Continued*

spectral content. Overall level decreases by 6 dB for every doubling of the distance between the source and the listener (the inverse square law), and atmospheric absorption gradually attenuates the high-frequency components of a sound as the distance between source and listener is increased (about 2 dB/100 ft at 6 kHz and 4 dB/100 ft at 10 kHz). The utility of both of these monaural cues, of course, depends on knowledge of source characteristics. However, the requirement for a priori knowledge about the source can be eliminated if the perceiver is allowed two or more "looks" at the stimulus from different vantage points. For example, Lambert (1974) pointed out that just two looks at stimulus intensity, as might be obtained if the perceiver's head were rotated, would provide sufficient information for a determination of source distance, without the need for knowledge of source characteristics.

There are two potential binaural distance cues: ITD and ILD; both vary slightly with the distance between source and listener (Coleman, 1963). In the case of ITD, for a source at 90° azimuth, there can be as much as a 150 μs difference in the ITD produced by a near source and a far source. A near source produces a larger ITD than a far source. This change in ITD with

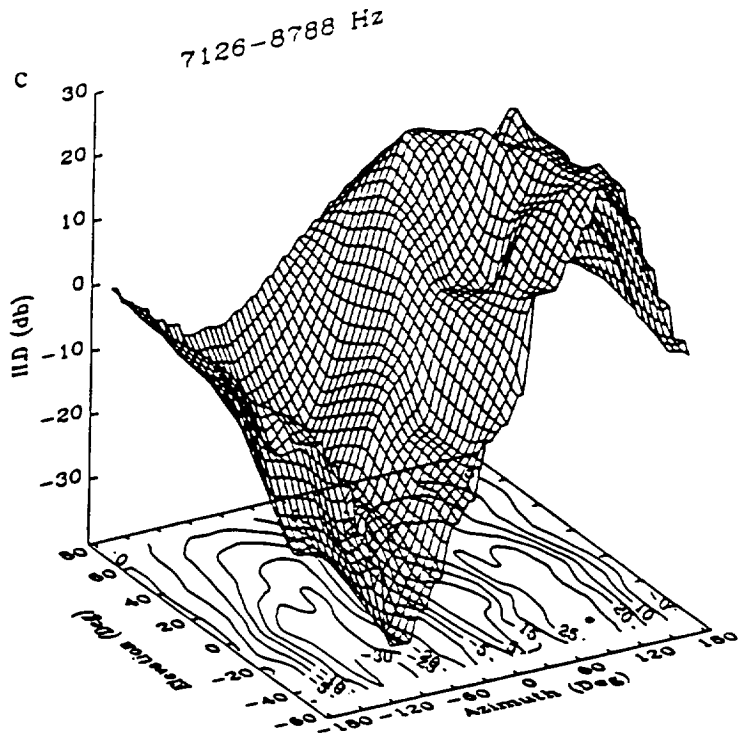

FIGURE 6 *Continued*

distance occurs because with a source close to the head the extra distance around the head is greater than if the source were far from the head. Distance affects ILDs in a comparable way, although in this case the effect is highly frequency dependent. At low frequencies the distance effect is greatest. For a 300-Hz tone at 90° azimuth, for example, the ILD for a source far from the head (several wavelengths) is about 0.5 dB, but for a source at 44 cm it is over 10 dB. The effects at higher frequencies and at source azimuths off the interaural axis are considerably smaller.

In a nonanechoic environment, which of course includes nearly all everyday listening situations, there is an additional distance cue provided by the mix of the direct sound wave from source to listener with the reflections of that sound wave off the surfaces of the listening room. When the sound source is close to the head the direct sound dominates since because of the extra distance traveled and absorption at the surfaces, the level of the reflected sound is always lower. However, as the source-to-listener distance increases, the direct sound level decreases, and the ratio of direct to reflected sound level decreases. Given a specific enclosure, then, this ratio is perfectly

correlated with source-to-listener distance. Moreover, even though it is a monaural cue, its validity does not depend on a priori knowledge of stimulus characteristics.

B. Acoustical Determinants of Apparent Spatial Position

Our purpose in this section is to review what is currently known about how the acoustical information about the spatial position of stationary sources is actually used. Most of the experiments in this area have considered apparent source direction and apparent distance separately, and for convenience we maintain this separation here. Several comprehensive reviews of this area have appeared recently (Middlebrooks & Green, 1991; Wightman & Kistler, 1993), so the material is only summarized here.

In the vast majority of experiments on the apparent spatial position of stationary auditory objects, only apparent direction (azimuth and elevation) has been considered. Until recently, the dominant theoretical position, epitomized by the duplex theory (Strutt, 1907), was that ITD provided the dominant source of information about apparent direction at low frequencies and that ILD was dominant at high frequencies. The duplex theory derived from the facts that the auditory system was much less sensitive to ITDs at high frequencies than at low frequencies (Joris & Yin, 1992; Yin & Chan, 1988) and from the fact that ILDs are much larger at high frequencies than at low frequencies (see Figure 5). Information provided by pinna filtering was not considered in the duplex theory.

Few empirical data on apparent source direction contradict the duplex theory. However, there are many natural circumstances that reveal the limitations of the theory and that argue for a situation-dependent weighting of the various sources of information about apparent sound direction. Localization of narrowband sounds is one such circumstance. Most narrowband sounds offer conflicting cues to apparent direction, so it is not surprising that they are not often localized accurately. The extreme case of a narrowband sound is sinusoid. Sinusoids offer doubly ambiguous ITD cues. A 1000-Hz sinusoid, for example, could provide a 400-μs ITD leading to the right ear while at the same time indicating a 600-μs ITD leading to the left ear. As Figure 4 shows, each ITD signals a whole range of potential source directions. It should not be surprising that unless a sinusoid has a broadband transient associated with onset or offset, its apparent position is unclear (Hartmann, 1983). Other narrowband sounds are somewhat less ambiguous but still inaccurately localized. The apparent azimuth of a high-frequency noise band is given by ILD, as suggested by the duplex theory (Middlebrooks, 1992). However, the apparent elevation seems to be determined by a learned association between spatial position and the spectral peaks and valleys produced by pinna filtering (Middlebrooks, 1992). The resultant

apparent direction is often far removed from the actual source direction and well off the contour of directions indicated by ILD alone. In this case and others (e.g., monaural localization, as described by Butler, Humanski, & Musicant, 1990), the learned association between spatial position and pinna filtering details appears to be a favored source of information about apparent sound direction. In general, the data suggest that in the absence of unambiguous (i.e., wideband) ITD, the information provided by pinna filtering appears to dominate.

If a wideband source contains both low and high frequencies, apparent direction seems to be governed primarily by ITD (Wightman & Kistler, 1992). In the Wightman and Kistler experiments, free-field noise sources were synthesized by using algorithms that were based on listeners' own HRTFs. The *virtual sources* were then presented by means of headphones, affording complete control over the acoustical stimulus. When the ITD information was manipulated to signal one direction and all other cues were left to signal another direction, the listeners' judgments of apparent direction always followed the ITD cue. Thus, even in the presence of opposing ILDs of as much as 20 dB, ITD was dominant. The dominance of ITD occurred for all listeners so long as the stimuli contained energy below about 1500 Hz. When the low frequencies were filtered out, ITD was effectively ignored and judgments of apparent position followed the ILDs and pinna filtering cues.

The importance of the ITD cue is further emphasized by the fact that listeners' make frequent front–back confusions in certain conditions (Oldfield & Parker, 1984a, 1984b; Stevens & Newman, 1936; Wenzel, Arruda, Kistler, & Wightman, 1993; Wightman & Kistler, 1989b). Recall that if apparent direction were governed by ITD, front–back confusions would be expected given the spherical symmetry of the head (Figure 4). While the rate of front–back confusions in everyday life is unknown, with laboratory stimuli and especially virtual source stimuli, front–back confusion rates can be as great as 25% (Oldfield & Parker, 1984a, 1984b; Wightman & Kistler, 1989b). Contours of constant ITD from actual measurements are smooth and regular, as predicted by the symmetry argument, though slightly different for different listeners (Wightman & Kistler, 1993). Contours of constant ILD, on the other hand, are quite irregular and variable from one frequency band to another (Figure 6). We suggest that the fact that listeners make consistent and frequent front–back confusions argues at least indirectly for the dominance of ITD cues and the lesser importance of ILD and pinna filtering cues.

The relative salience of the various acoustical cues to the spatial layout of auditory objects also depends on the "realism" of the cues. In experiments with virtual sources similar to those described above in which ITD was in conflict with other cues (Wightman & Kistler, 1992), we have produced

stimuli in which cues in one frequency region conflict with cues in another frequency region. In one condition, for example, the ILD and spectral cues were the same throughout the frequency range (200 Hz–14000 Hz) and signaled, or "pointed to," one of five possible directions on the horizontal plane. The ITD cue in each of four bands (roughly 1.5 octaves wide) pointed to a different direction. Thus, the ITD cue could be said to be "inconsistent" across the frequency range, and the other cues could be said to be "consistent." In other conditions, the ITD cue was consistent and the other cues were inconsistent, and in still other conditions, the frequency range was divided somewhat differently. The results were unambiguous. Listeners' judgments always followed the consistent cue. Even if the ITD cue was inconsistent in a single high-frequency band (above 5 kHz) listeners appeared to ignore ITD and put maximum weight on the ILD and spectral cues that were consistent across the spectrum. Not only does this result suggest that high-frequency ITD cues are encoded as well as low-frequency ITD cues, but it also suggests that cues that are realistic are given greater weight than unrealistic cues. With real sources and real listening environments, it is highly unlikely that either the ITD or the other cues could be inconsistent across the frequency spectrum.

The fidelity of the ITD, ILD, and spectral cues to spatial position is compromised in most natural listening situations by the presence of echoes. These echoes, which to a first approximation are filtered copies of the sound wave, are produced when a sound wave bounces off objects or surfaces in the environment and because of the extra distance they have to travel they reach the listener slightly later than the original or direct sound wave. Typically, the intensities of the echoes are considerably weaker than the intensity of the direct sound, both because of the additional path length and because most objects and surfaces absorb some of the sound energy, particularly at high frequencies. Nevertheless, when the echoes combine with the direct sound, the acoustical cues that signal the spatial position of the sound source are disrupted. With echoes the effective stimulus at each ear consists of the superposition of sounds from a number of different directions. Thus, both the monaural and binaural cues are distorted.

It might be expected that the presence of echoes would seriously impair a listener's ability to determine the spatial layout of sound sources. In fact, in all but the most extreme cases, the echoes are hardly noticed, and localization performance is not impaired (Begault, 1992; Hartmann, 1983). The substantial body of empirical data on this phenomenon can be summarized in the hypothesis that listeners attend only to the first few milliseconds of a stimulus, the time before echoes arrive, to determine the spatial position of a source. The spatial information arriving later, which would be corrupted by echoes, is somehow suppressed. This is the well-known *precedence effect* (Clifton & Freyman, 1989; Wallach, Newman, & Rosenzweig, 1949; Zurek,

1980). Although many of the characteristics of the phenomenon and most of the underlying mechanisms are not well understood, it is clear that the precedence effect is of central importance to the determination of auditory spatial layout in natural listening situations.

Compared with our well-developed understanding of how various sources of acoustical information are combined to determine the apparent direction of auditory objects, relatively little is known about how listeners might form a judgment of apparent distance. Available evidence suggests that perception of auditory distance is not well developed in humans. Apparent distance is typically very different than real distance (e.g., Gardner, 1968; Mershon & King, 1975), and only relative distance can be determined with any accuracy (Cochran, Throop, & Simpson, 1968; Holt & Thurlow, 1969). While there are suggestions in the literature that the distances of familiar sounds are judged more accurately (Coleman, 1962; McGregor, Horn, & Todd, 1985), the classic demonstration by Gardner (1968) shows that in an anechoic room with levels equalized, even the apparent distance of speech is not accurately reported. The most reliable finding seems to be that sounds presented with reverberation are judged to be more distance than the same sounds presented without reverberation (e.g., Mershon & King, 1975).

From several different perspectives inaccuracies in judging the distance of an auditory object are not surprising. First, the primary acoustical correlates of distance, level, and spectrum are unambiguous only if the characteristics of the source are known. Second, in everyday life the absolute distance of an auditory object carries little significance. Direction is clearly much more important, it serves to orient our gaze. Of course, if an auditory object is moving, and especially if that movement is toward the listener, distance carries considerable significance. Experiments on estimation of distance of a moving auditory object typically ask listeners to judge the time at which the object will reach to listener's position, this is called *time-to-contact*. The available data on listeners' judgments of auditory time-to-contact is reviewed in a later section of this chapter.

V. SPATIAL LAYOUT OF DYNAMIC AUDITORY OBJECTS

In everyday life an individual's auditory world is constantly in motion. The orientations of sound-producing objects with respect to a listener's head and ears are ever changing, either because the objects themselves are moving or because the listener's head is moving. In either case, the result is a constantly changing pattern of directional cues at the ears and, if conditions are right, the introduction of additional cues to movement such as the Doppler shift. This section of the chapter describes those additional movement cues in some detail, and we then discuss the available psychophysical data on listeners' processing of dynamic spatial information.

A. Additional Acoustical Information from Moving Sounds

Moving sounds can be described by using the mathematics of kinematics (Jenison & Lufti, 1992). *Kinematics* is the branch of mechanics that describes pure motion that uses the variables of displacement, time, velocity, and acceleration. Doppler shifts, changes in ITD (described earlier) and intensity, can be shown to have dependencies that are based on kinematics. In addition to ITD, Doppler shift, and time-varying intensity, the first differentials of these observed variables may directly be sensed as well. Figure 7 shows the geometry of the sound source moving relative to an observer. φ_t is the angle of the incident wavefront at any time t and is dependent on the distance D_t to a point p on the median plane. θ_0 is the angle at the anticipated closest point of approach (CPA), and β is the angle of the source trajectory relative to the median plane. Angle β is equivalent in magnitude to $\theta_0 + \pi/2$. R_t is the distance from the sound source to the observer.

Movement of either the sound source or the observer changes the relative wavelength of the sound waves. This change is known as the *Doppler shift*. The well-known lawful dependence of the Doppler shift on velocity of the sound source relative to an observer is

$$\omega = \frac{\omega_0}{(1 - M \cos \varphi_t)},$$

where ω_0 is the intrinsic frequency, ω is the shifted frequency, M is the Mach number defined as velocity divided by the speed of sound, and φ_t is the angle of trajectory relative to the observer (see Figure 7). The frequency shift depends only on the velocity component directed toward the observer. This result holds true regardless of the time history of the trajectory. The

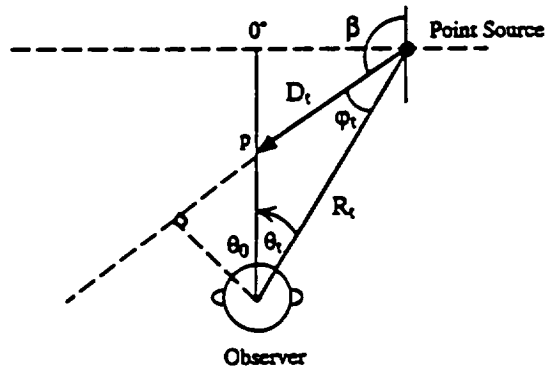

FIGURE 7 Schematic diagram showing angular relations between a listener and a sound source that is moving along a straight path (represented by the arrow).

Doppler-shifted frequency at a given time and position is affected only by the source's velocity and frequency at the instant the wave is generated. Furthermore, the source need not be traveling at a constant velocity or in a straight line for it to apply. When the sound source is far from the observer and approaching (φ_t is small, thus $\cos[\varphi_t]$ is near 1), the angle φ_t changes very little, hence little change in the frequency shift. However, the magnitude of the shift will be at its maximum. Since the sound source is approaching the observer, the shift is toward a higher frequency. As the sound source approaches the observer, φ_t increases rapidly, resulting in a rapid decrease in frequency. As the sound source passes and recedes, there is a corresponding decrease in frequency relative to the intrinsic frequency of the sound source. This of course is the experience we have all had listening to a passing train whistle that decreases in pitch as it passes by and recedes into the distance.

These observed variables, ITD, time-varying intensity, and Doppler, along with their first-order differentials with respect to time, all have characteristic spectrotemporal patterns. Zakarauskas and Cynader (1991) analyzed intensity patterns for actual moving sound sources along various trajectories and derived mathematical expressions for the observed variables that are related to the inverse-square distance relationship. Jenison (1994) extended these analyses to include Doppler and ITD patterns. The simplest trajectory is that of the rectilinear approach with constant velocity as shown in Figure 8. For illustration, the starting point for the moving sound source in these examples is located some distance R_s directly on the median line as shown in the Figure 8.

The characteristic patterns for the three sound source trajectory angles (β) of 90°, 120°, and 150° are shown in Figure 9. For the purpose of this

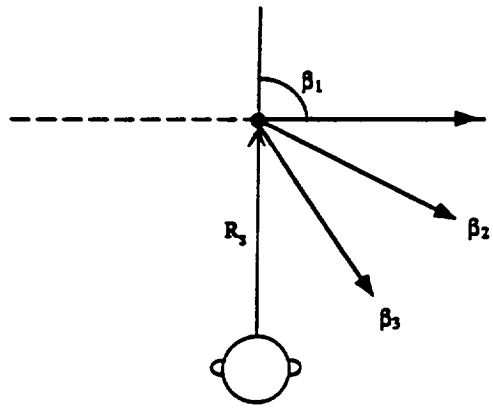

FIGURE 8 Schematic diagram showing three example trajectories for a moving sound source.

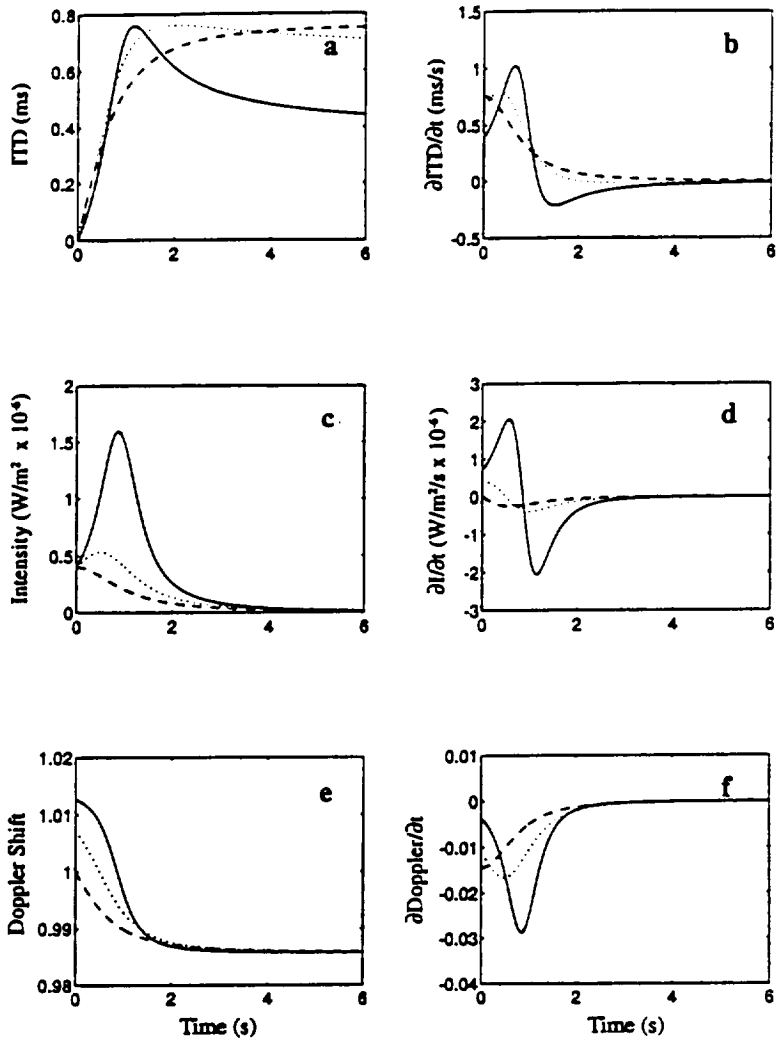

FIGURE 9 Results of kinematic analysis of the interaural time difference (ITD); (a), intensity (b), and Doppler shift (c) cues produced by a moving sound source. The rates of change of those cues are shown in (b), (d), and (f).

example, we have assumed a source of moderate intensity, a velocity of 5 m/s, and a starting distance from the observer of 5 m. Note that all of the ITD functions begin at 0 delay because of the midline starting point. The intensity functions will also start at the same intensity for a given distance from the observer. In the case of the Doppler shift, the shift is toward a

higher frequency when the sound is approaching the observer and toward a lower frequency when receding. So for β_1 equal to 90°, the frequency shift will start at unity and decline. For the cases of β_2 and β_3, where the source is initially approaching, passes through a CPA and then recedes, the frequency shift will initially be greater than unity and then decline.

Jenison (1994) has shown that acoustical kinematics sufficiently convey velocity (trajectory and speed) information regarding the moving sound source directly from the observed Doppler shift together with time-varying ITD. Although the theoretical analyses show that sufficient information is available to the observer regarding higher order variables such as the velocity and time-to-contact of the moving sound source, it remains to be known whether the human observer has sufficient sensory mechanisms to detect this information, particularly under conditions of uncertainty.

Most of the empirical research on perception of moving sound sources has focused, either directly or indirectly, on the question of whether dynamic spatial changes are processed with some kind of specialized *movement detectors*. There is considerable neurophysiological evidence that differential information lawfully related to motion is directly detected by the visual system (Maunsell & VanEssen, 1983). Recent evidence suggests that there are also direction-sensitive neurons spatially segregated in auditory cortex (Stumpf, Toronchuk & Cynader, 1992). Other findings suggest that neural processing of auditory motion involves mechanisms distinct from those involved in processing stationary sound location (Spitzer & Semple, 1991, 1993; Stumpf, Toronchuk, & Cynader, 1992). Thus, while converging physiological evidence supports the existence of motion sensitive neurons, the psychophysical evidence for specialized motion detectors is inconclusive. The two lines of research that have addressed this question involve measurements of the *minimum audible movement angle* (MAMA) and measurements of auditory motion aftereffects.

The MAMA experiments are variations of the classical *minimum audible angle* (MAA) experiments conducted with stationary sources. They are both detection or discrimination experiments that measure the threshold for discriminating small changes in spatial parameters. In the case of MAAs, what is measured is the smallest spatial separation of two static sources that can reliably be detected. The MAMA represents the smallest amount of spatial displacement or movement of a single source that can reliably be detected. Although both experiments can inform us about the processing capabilities of the auditory system, it is important to note that since they involve discrimination or detection paradigms, the extent to which the results can be generalized to questions about apparent spatial position may be quite limited. In other words, that listeners can discriminate between two sources at slightly different spatial positions does not necessarily imply that the apparent positions of the sources were different. Similarly, discrimination

between a moving source and a static source does not necessarily imply that movement itself was perceived.

While the investigators involved in the MAMA research may quibble over details, most would probably agree that the results do not support the existence of specialized motion detectors in the auditory system. Measured MAMAs, when expressed in terms of the total angle traversed at threshold, are roughly the same as or slightly larger than the MAAs measured with stationary sources, or about 2° (Grantham, 1986; Harris & Sergeant, 1971; Perrott & Musicant, 1977; Perrott & Tucker, 1988). A simple explanation of the basic MAMA results is that the listener takes an acoustic "snapshot" of the position of the source at the beginning and end of its trajectory (Grantham, 1986) and discriminates on the basis of static positional changes. Not all the available data support this view, but the exceptions are relatively minor (Perrott & Marlborough, 1989).

Gibson (1966) took issue with the notion of a series of perceptual snapshots, which requires fusion or composition to account for the perception of a single moving object. By redefining information for motion perception, Gibson eliminated the need for a concept such as fusion. Since motion information is available to the observer, even through discrete looks, the additional step of reconstruction to a continuous event is simply not necessary. To Gibson, the mechanics of the mediating sensory system were not germane to the perception of motion. To have "dynamic event perception," in contrast to the less elegant "motion perception plus inference," it must be shown that even though dynamic properties, such as mass and inertia, are not present in the optic (or acoustic) array, they are specified by the kinematics. That is, the information regarding the physical motion of an object is conveyed through the kinematics, whether discrete or continuous.

Research on motion aftereffects provides indirect evidence on the question of the existence of specialized motion detectors. The idea is that exposure to an adapting stimulus that is moving in one direction fatigues the neural elements that respond to movement in that direction. The aftereffect, a perception of movement in the opposite direction, is presumed to reflect the spontaneous activity of the neural elements sensitive to movement in the opposite direction. Movement aftereffects are common in vision, one variation of which is called the *waterfall illusion* (Sekular & Pantle, 1967).

Grantham (1989, 1992) has reported reliable though weak evidence for motion aftereffects in audition. After prolonged exposure to a free-field adapting stimulus that was moving in the horizontal plane, listeners' judgments of the direction of movement of a subsequently presented probe stimulus were slightly biased in a direction opposite to that of the adapting stimulus. While the effects were disappointingly small, the results were nevertheless suggestive.

Some of the research on perception of moving sound sources has been

less concerned with the existence of specialized motion detectors and more broadly focused. For example, several studies have attempted to quantify the relative salience of the various sources of acoustical information that signal source movement. These experiments ask listeners to indicate the time at which a moving source is closest to them (time to interception) or the time at which they would make contact with the source (acoustic tau). In a theoretical study, Shaw, McGowan, and Turvey (1991) analyzed the acoustic intensity field produced by collinear relative movement between a sound source and an observer and showed the acoustic tau to be related to the inverse of the relative change in average intensity. Jenison (1994) extended the analysis to the more general case, including *time-to-interception*, showing that time-averaged intensity and time-varying ITD and their corresponding first-order derivatives are sufficient for conveying both collision and interception information.

Empirical studies of auditory time-to-contact or time-to-interception include research reported by Rosenblum, Carello, and Pastore (1987) in which listeners heard sound sources over headphones. Three stimulus variables were manipulated: interaural time difference, overall level, and Doppler shift. Each was presented both in isolation and in competition so that each indicated a different point of closest approach or interception. The results suggested that while any of the three stimulus parameters could accurately indicate point of closest approach, overall level was the dominant cue. The authors argue that overall level should be dominant since it is the only cue of the three that is, in all environmental circumstances, unequivocal. Todd (1981) investigated how well subjects could discriminate time-to-contact for visual stimuli by simulating two simultaneously approaching objects on a computer display. Subjects were asked to judge which object would arrive first. We have recently launched analogous experiments that examine subjects' ability to discriminate the arrival of two sound sources. Sounds were synthesized according to the simple kinematics of a moving sound composed of three harmonics by using ITD, average intensity, and Doppler shift. A sound arriving to the left of the listener was mixed with a sound arriving differentially in time to the right of the observer. Subjects were asked to choose which sound would arrive sooner. Figure 10 shows preliminary results from 24 subjects. In Todd's experiment, relative time-to-contact was 75% correctly discriminated when the difference in time-to-contact was about 50 ms. In contrast, the relative auditory time-to-contact in our preliminary studies was 75% correctly discriminated when the difference was about 300 ms. Schiff and Oldak (1990) examined observers' accuracy in using visual and acoustical estimates of time-to-arrival from film and sound-recorded approaching vehicles. Their data indicate that sighted subjects were significantly more accurate in estimating time-to-arrival with sight than sound, however, visually impaired subjects performed as well as

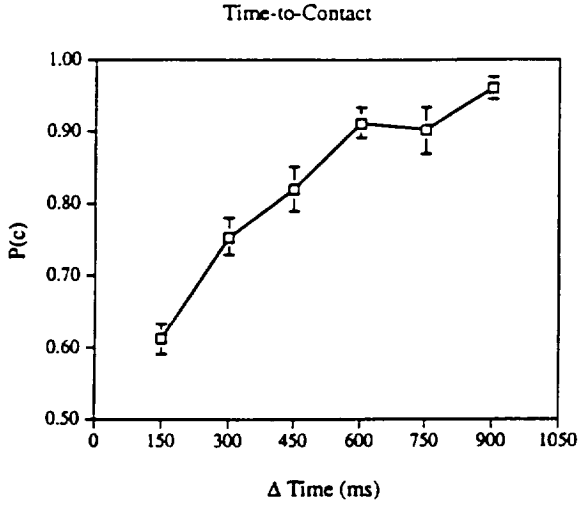

FIGURE 10 Average psychometric function from 24 listeners in the time-to-contact experiment. Percentage correct discriminations between two sounds arriving at different times is plotted as a function of the arrival time difference.

or better than the sighted subjects with only the acoustic channel. Although the evidence is only suggestive at this point, human observers have the capacity to efficiently estimate relative time-to-contact regardless of how the information is conveyed as long as the temporal window for estimation is within several seconds. This restricted window should not be surprising given the pattern of the observables described above. Significant changes in ITD, intensity, and Doppler occur only in a spatial region (hence the temporal region as well) about the CPA. This relationship holds for subtended angle in the visual domain as well.

Head movements provide a somewhat different kind of dynamic auditory stimulus from movement of the sound source. Because head movements typically involve changes only in the direction of the sound source with respect to the head there is very little Doppler shift and very little change in overall level. However, interaural parameters change more rapidly with head movements than with typical source movement. In addition, head movements provide additional information to the perceiver by means of proprioceptive feedback from the neck musculature. Although there has been speculation about the role of head movements for decades, there have been few empirical studies of their role (Pollack & Rose, 1967; Simpson & Stanton, 1973; Thurlow & Runge, 1967). Only recently has empirical research begun to provide firm evidence of the importance of head movements for perception of the spatial layout of auditory objects.

Given a stationary auditory object in the environment there is a change in the angular relation of the object and a listener's head that accompanies normal head movement. This change in relative orientation produces a systematic and predictable change in the pattern of spatial cues (ITD, ILD, and spectral cues) produced by the object at the listener's ears. If these normal changes in the spatial cues are disrupted, the apparent position of the auditory object is often disturbed. Young (1931) reported one of the first demonstrations of this phenomenon. In this experiment, sounds were routed to the ears through rubber tubes attached to fixed ear trumpets. With this arrangement the normal coupling between a listener's head movements and changes in the acoustical stimulus at the ears was eliminated. Listeners reported all sounds as originating behind the head, outside of the listeners' visual fields, regardless of the actual position of the sound source. Similar front–back confusions are reported in the modern studies of virtual sound sources that are synthesized and presented to listeners by means of headphones (Wightman & Kistler, 1989b).

As mentioned above, front–back confusions are not entirely unexpected given the rough spherical symmetry of the head and the salience of ITD cues. The idea that in everyday life a listener's head movements might provide the information needed to avoid them is usually attributed to Wallach (1940). Wallach showed that if a listener could monitor the direction of change in ITD that accompanied a head movement, the front–back ambiguity could be avoided. For example, suppose a sound is presented at an azimuth of 45° and an elevation of 0° (on the horizontal plane, roughly 45° to the right of the median plane). A front–back confusion would be represented by an apparent azimuth report of roughly 135°. If the listener's head moved to the right, the ITD produced by the source initially at 45° would decrease because the angle of the source relative to the head would approach 0°, the point of minimum ITD. However, if the source were actually at 135° azimuth, the ITD would have increased. Thus, the direction of change in ITD unambiguously indicates whether the source was in the front or in the rear.

In spite of the simplicity and face validity of Wallach's (1940) arguments, conclusive evidence that head movements are used to resolve front–back confusions has not appeared. One obvious reason for this is that experiments that control both head movements and the associated auditory stimulus dynamics have been technically too demanding until recently. Advanced technology now allows synthesis of virtual sources in such a way that the effects of head movements can directly be studied. Using magnetic head trackers and real-time convolution devices such as the Convolvotron (Foster, Wenzel, & Taylor, 1991), one can monitor a listener's head position continually during an experiment and adjust the synthesis algorithms dynamically (20–40 times per second) to simulate a stationary source. As the

listener's head moves, the device compensates for changes in the relative positions of the stationary virtual source and the head by using different left–right pairs of HRTF-based filters for each updated head position. The movement compensation is smooth and the resultant percept of an external sound source in a stationary position is compellingly realistic (Wenzel, 1992).

We have recently begun some research on the role of head movements that takes advantage of the new technology and attempts to clarify some of the issues raised by the earlier work (Wightman, Kistler, & Andersen, 1994). The essential elements of the paradigm were as described in earlier work (Wightman & Kistler, 1989b). Listeners localized virtual sources (2.5 s wideband noise bursts) in two conditions. In one, the virtual stimuli were presented over headphones with no head tracking, and the listeners were asked not to move their heads during the test. In the other, a magnetic head tracker was used to sense head position, and the virtual synthesis algorithms were modified in real time according to the head tracker's reports. In the second condition, listeners were encouraged to move their heads during stimulus presentation if they felt it would facilitate localization. Apparent position judgments were made verbally after each stimulus presentation. Preliminary results from a single listener are shown in Figure 11. Note that in the head stationary condition this listener made frequent front–back confusions, as evidenced by the off-diagonal responses in the front–back panel. In the head-movement condition, however, the front–back confusions were nearly eliminated. The listeners' gave no indication of other differences between the two conditions, either in their apparent position judgments or in their subjective reports. Thus, in contrast with suggestions in the literature, apparent source distance was the same with and without head movements (cf. Simpson & Stanton, 1973), and the images were equally well externalized in the two conditions (cf. Durlach et al., 1992). We conclude on the basis of these results that the primary role of head movements is resolution of confusions about the spatial layout of auditory objects.

VI. THE ROLE OF AUDITORY–VISUAL INTERACTIONS IN THE SPATIAL LAYOUT OF AUDITORY OBJECTS

The sensory environment of most individuals includes both visual and auditory objects, and in many cases sound-producing objects can be seen as well as heard. Thus, while it is useful and informative to consider audition alone when discussing the spatial layout of auditory objects, it is important to be mindful of the potential role played by vision. Indeed, some auditory–visual interactions are quite powerful and their consequences well documented.

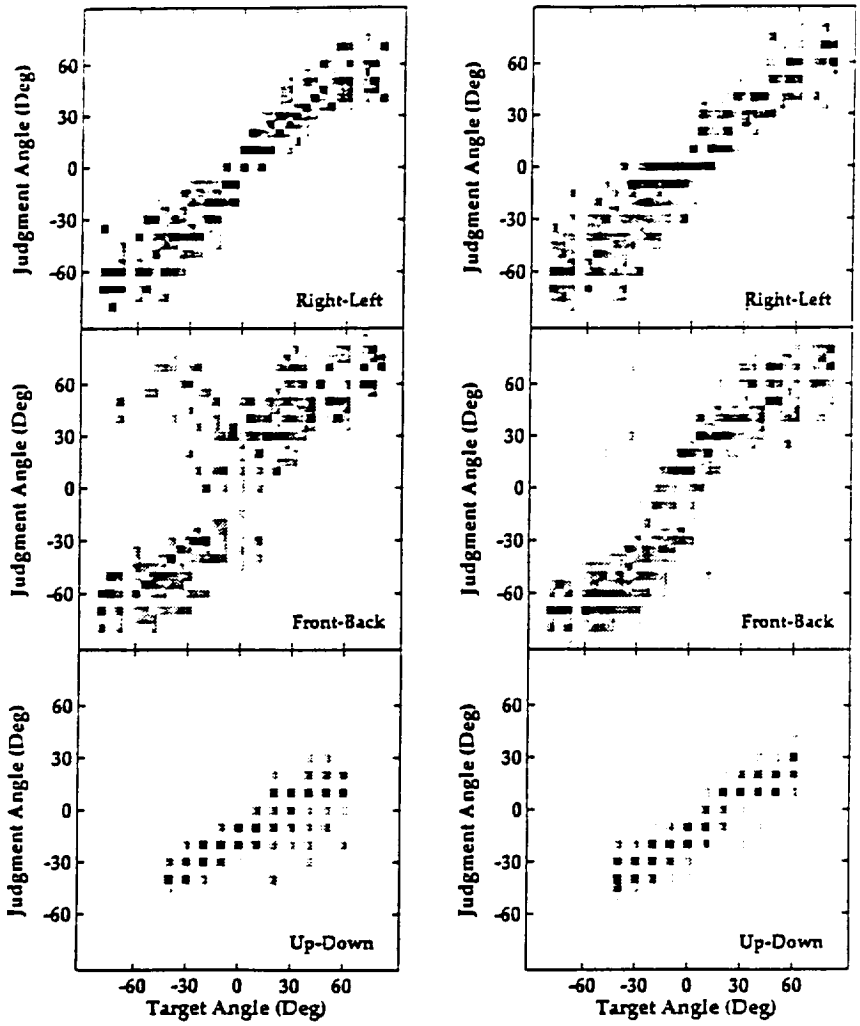

FIGURE 11 Apparent source position judgments from a single listener in an experiment in which the listener heard virtual sources presented over headphones. In one condition (left panels) the listener was required to hold his or her head still, and in the other condition (right panels) the listener was encouraged to move his or her head and the virtual stimuli were modified in real time according to the listener's head position to simulate a stationary external source. Each judgment of apparent azimuth and elevation is represented in three panels that reflect the extent (expressed as an angle from −90° to +90°) to which the judged position is on the right or left (top), in the front or back (middle), and above or below the horizontal plane (bottom). The darkness of each symbol represents the number of judgments that fell in the local area of the symbol.

The so-called *ventriloquism effect* is perhaps the best known of the auditory–visual interactions (e.g., Pick, Warren, & Hay, 1969). The typical manifestation of the effect is a strong biasing of the apparent position of an auditory object in the direction of a simultaneously present visual object. Evidence of the potency of this effect is familiar to anyone who has watched the image of someone speaking at the movies or on television. While the sound of the voice clearly seems to originate at the mouth of the person speaking, the actual source of the sound, a loudspeaker, is usually displaced far to one side. Clearly one's perception of the spatial layout of auditory objects will be heavily influenced by whether or not the source of the sound is visible.

Additional evidence for auditory–visual interactions comes from research on visual facilitation (e.g., Warren, 1970). Visual facilitation refers to the fact that the variance of localization judgments is lower when listeners hear the test stimulus in a lighted room than when they hear it in the dark. The source of sound is invisible in either case, and whether the listener makes the response in the light or the dark is irrelevant to the outcome. It is as if the listener is able to establish a frame of reference within which to place the auditory objects, and the presence of the frame of reference facilitates localization. Some investigators argue that eye movements, even in the absence of visual input, are the basis of the facilitation effect (Jones & Kabanoff, 1975), but the issue is far from being resolved. What is especially interesting about the visual facilitation effect is that it occurs only in adults. Children as old as 12 years do not show the effect (Warren, 1970).

VII. CONCLUSION

The study of auditory object perception in general and the spatial layout of auditory objects in particular is in its infancy. In the case of the spatial layout of single stationary sound sources in anechoic space much is known about the sources of information and how that information is processed. The salience of ITD cues, the importance of monaural spectral cues derived from pinna filtering, the role of head movements, and so forth, have been thoroughly documented in studies of single stationary sources. Relatively few investigators have ventured beyond the relative security of this constraint so that experiments involving nonanechoic listening conditions and moving sources are scarce, and studies of multiple sources are virtually nonexistent. The potential sources of information are reasonably well understood, but how that information might be used in the auditory system is completely unknown.

The state of affairs in hearing contrasts sharply with the relative maturity of the study of visual spatial layout, in which research on such complex topics as optic flow has been in progress for decades. One reason for the

slower progress on the hearing side may be that the experiments are technically more demanding. For example, it is easier to present an arbitrary visual pattern to a retina than an arbitrary sound waveform to an eardrum. Technology is changing this situation rapidly, so we can expect significant advances in our understanding of auditory object perception in the near future.

Acknowledgments

The authors are indebted to Doris Kistler and many others in the Hearing Development Research Laboratory for their assistance in preparing this chapter and in conducting the research that led to it. Preparation of the manuscript was supported in part by research grants from the National Institutes of Health (National Institute of Deafness and Other Communicative Disorders), National Aeronautics and Space Administration, and Office of Naval Research.

References

Begault, D. (1992). Perceptual effects of synthetic reverberation on three-dimensional audio systems. *Journal of the Audio Engineering Society, 40,* 895–904.

Bregman, A. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Butler, R., Humanski, R., & Musicant, A. (1990). Binaural and monaural localization of sound in two-dimensional space. *Perception, 19,* 241–256.

Blauert, J. (1983). *Spatial hearing: The psychophysics of human sound localization.* Cambridge, MA: MIT Press.

Clifton, R., & Freyman, R. (1989). Effect of click rate and delay on breakdown of the precedence effect. *Perception & Psychophysics, 46,* 139–145.

Cochran, P., Throop, J., & Simpson, W. E. (1968). Estimation of distance of a source of sound. *American Journal of Psychology, 81,* 198–207.

Coleman, P. D. (1962). Failure to localize the source distance of an unfamiliar sound. *Journal of the Acoustical Society of America, 34,* 345–346.

Coleman, P. D. (1963). An analysis of cues to auditory depth perception in free space. *Psychological Bulletin, 60,* 302–315.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., & Wenzel, E. M. (1992). On the externalization of auditory images. *Presence, 1*(2), 251–257.

Foster, S. H., Wenzel, E. M., & Taylor, R. M. (1991, October). Real time synthesis of complex acoustic environments. Paper presented at the *IEEE Workshop on Applications of Signal Processing to Audio & Acoustics,* New Paltz, NY.

Gardner, M. B. (1968). Proximity image effect in sound localization. *Journal of the Acoustical Society of America, 43,* 163.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Grantham, D. W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *Journal of the Acoustical Society of America, 79,* 1939–1949.

Grantham, D. W. (1989). Motion aftereffects with horizontally moving sound sources in the free field. *Perception & Psychophysics, 45*(2), 129–136.

Grantham, D. W. (1992). Adaptation to auditory motion in the horizontal plane: Effect of prior exposure to motion on motion detectability. *Perception & Psychophysics, 52*(2), 144–150.

Green, D. M. (1976). *An introduction to hearing.* New York: Wiley.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: MIT Press.

Harris, J. D., & Sergeant, R. L. (1971). Monaural/binaural minimum audible angle for a moving sound source. *Journal of Speech and Hearing Research, 14*, 618–629.

Hartmann, W. M. (1983). Localization of sound in rooms. *Journal of the Acoustical Society of America, 74*, 1380–1391.

Holt, R. E., & Thurlow, W. R. (1969). Subject orientation and judgment of distance of a sound source. *Journal of the Acoustical Society of America, 6*(2), 1584.

Jenison, R. L. (1994). *On acoustic information for auditory motion*. Manuscript submitted for publication.

Jenison, R. L., & Lutfi, R. A. (1992). Kinematic synthesis of auditory motion. *Journal of the Acoustical Society of America, 92*, 2458.

Jones, B., & Kabanoff, B. (1975). Eye movements in auditory space perception. *Perception & Psychophysics, 17*, 241–245.

Joris, P. X., & Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *Journal of the Acoustical Society of America, 91*, 215–232.

Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. *Journal of the Acoustical Society of America, 62*, 157–167.

Lambert, R. (1974). Dynamic theory of sound-source localization. *Journal of the Acoustical Society of America, 56*, 165–171.

Maunsell, J. H. R., & van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area (MT) of Macaque monkey: I. Binocular interactions and the sensitivity to binocular disparity. *Journal of Neurophysiology, 49*, 1148–1167.

McGregor, P., Horn, A. G., & Todd, M. A. (1985). Are familiar sounds ranged more accurately? *Perceptual and Motor Skills, 61*, 1082.

Mershon, D. H., & King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics, 18*(6), 409–415.

Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. *Journal of the Acoustical Society of America, 92*(5), 2607–2624.

Middlebrooks, J. C., & Green, D. M. (1990). Directional dependence of interaural envelope delays. *Journal of the Acoustical Society of America, 87*(50), 2149–2162.

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. In M. Rozenzweig and L. Porter (Eds.), *Annual review of psychology* (Vol. 42, pp. 135–159). Palo Alto, CA: Annual Reviews Inc.

Middlebrooks, J. C., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *Journal of the Acoustical Society of America, 86*(1), 89–108.

Oldfield, S. R., & Parker, S. P. A. (1984a). Acuity of sound localization: A topography of auditory space: I. Normal hearing conditions. *Perception, 13*, 581–600.

Oldfield, S. R., & Parker, S. P. A. (1984b). Acuity of sound localization: A topography of auditory space: II. Pinna cues absent. *Perception, 13*, 601–617.

Perrott, D. R., & Marlborough, K. (1989). Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source. *Journal of the Acoustical Society of America, 85*, 1773–1775.

Perrott, D. R., & Musicant, A. D. (1977). Minimum auditory movement angle: Binaural localization of moving sound sources. *Journal of the Acoustical Society of America, 62*, 1463–1466.

Perrott, D. R., & Tucker, J. (1988). Minimum audible movement angle as a function of signal frequency and the velocity of the source. *Journal of the Acoustical Society of America, 83*, 1522–1527.

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics, 6*, 203–205.

Pollack, I., & Rose, M. (1967). Effects of head movements on the localization of sounds in the equatorial plane. *Perception & Psychophysics, 2,* 591–596.

Pralong, D., & Carlile, S. (1994). Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature in-ear recording system. *Journal of the Acoustical Society of America, 95,* 3435–3444.

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception, 16,* 175–186.

Schiff, W., & Oldak, R. (1990). Accuracy of judging time to arrival: Effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 303–316.

Sekular, R., & Pantle, A. (1967). A model for after-effects of seen movement. *Vision Research, 7,* 427–439.

Shaw, B. K., McGowan, R. S., & Turvey, M. T. (1991). An acoustic variable specifying time-to-contact. *Ecological Psychology, 3,* 253–261.

Shaw, E. A. G. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *Journal of the Acoustical Society of America, 56*(6), 1848–1861.

Simpson, W., & Stanton, L. (1973). Head movement does not facilitate perception of the distance of a source of sound. *American Journal of Psychology, 86,* 151–160.

Spitzer, M. W., & Semple, M. N. (1991). Interaural phase coding in auditory midbrain: Influence of dynamic stimulus features. *Science, 254,* 721–724.

Spitzer, M. W., & Semple, M. N. (1993). Responses of inferior colliculus neurons to time-varying interaural phase disparity: Effects of shifting the locus of virtual motion. *Journal of Neurophysiology, 69,* 1245–1263.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. *American Journal of Psychology, 48,* 297–306.

Strutt, J. W. (1907). On our perception of sound direction. *Philosophical Magazine, 13,* 214–232.

Stumpf, E., Toronchuk, J. M., & Cynader, M. S. (1992). Neurons in cat primary auditory cortex sensitive to correlates of auditory motion in three-dimensional space. *Experimental Brain Research, 88,* 158–168.

Thurlow, W. R., & Runge, P. S. (1967). Effect of induced head movements on localization of direction of sounds. *Journal of the Acoustical Society of America, 42*(2), 480–488.

Todd, J. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 795–810.

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology, 27*(4), 339–368.

Wallach, H., Newman, E., & Rosenzweig, M. (1949). The precedence effect in sound localization. *The American Journal of Psychology, 62,* 315–336.

Warren, D. H. (1970). Intermodality interactions in spatial localization. In W. Reitman (Ed.), *Cognitive psychology* (pp. 114–133). New York: Academic Press.

Wenzel, E. M. (1992). Localization in virtual acoustic displays. *Presence, 1*(1), 80–107.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America, 94,* 111–123.

Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening: I. Stimulus synthesis. *Journal of the Acoustical Society of America, 85,* 858–867.

Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening. II. Psychophysical validation. *Journal of the Acoustical Society of America, 85,* 868–878.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time

differences in sound localization. *Journal of the Acoustical Society of America, 91*(3), 1648–1661.

Wightman, F. L., & Kistler, D. J. (1993). Sound localization. In R. Fay, A. Popper, & W. Yost (Eds.), *Springer series in auditory research: Human psychophysics* (pp. 155–192). New York: Springer-Verlag.

Wightman, F. L., Kistler, D. J., & Andersen, K. J. (1994). Reassessment of the role of head movements in human sound localization. *Journal of the Acoustical Society of America, 95*, 3003.

Woodworth, R. S. (1938). *Experimental Psychology*. New York: Holt.

Yin, T. C. T., & Chan, J. C. K. (1988). Neural mechanisms underlying interaural time sensitivity to tones and noise. In G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 385–430). New York: Wiley.

Young, P. T. (1931). The role of head movements in auditory localization. *Journal of Experimental Psychology, XIV* (2), 95–124.

Zakarauskas, P., & Cynader, M. S. (1991). Aural intensity for a moving source. *Hearing Research, 52*, 233–244.

Zurek, P. (1980). The precedence effect and its possible role in the avoidance of interaural ambiguities. *Journal of the Acoustical Society of America, 67*, 952–964.