CHAPTER 1

AUDITORY SPACE

Simon Carlile

1. PERCEIVING REAL AND VIRTUAL SOUND FIELDS

1.1. PERCEIVING THE WORLD

One of the greatest and most enduring of intellectual quests is that of self understanding. What we understand and the intellectual models that we manipulate in the process of applying that understanding are intimately related to what we perceive of the world. Our perceptions are in turn related to the structure of our sense organs and to the brain itself. The neurosciences represent a rapidly growing body of knowledge and ideas about the marvelous machinery of the brain and are making an increasingly important contribution to this process. There is now a considerable understanding of the basic operation of the five senses of extero-reception: vision, hearing, touch, taste and smell. Our perception of our environment necessarily involves these five senses together with the senses of balance and body position (proprioception). The richness of our perception is clearly heightened by the complex combinations of these senses. For example, the successful restaurant generates a sensual experience that goes well beyond the simple satiation of hunger. The lighting and furnishings generate a mood that is relaxed and comfortable, the smells relate to the food and the conversation of other diners is muted and combines with the background music to generate a sense of communion and yet privacy.

In this book we are interested principally in the mechanisms by which the perception of an illusory or phantom space can be generated; in particular, the generation of virtual auditory space. In most cases this is achieved by presenting over headphones sounds that appear to come from locations in space that are distant from the listener. On the face of it, this might not appear too daunting a task.
An engineer might argue, quite reasonably, that by simply ensuring that the pattern of sound waves delivered over headphones to the ear drum was the same as when the individual was listening to a sound in free space, then the auditory experience should be identical. Indeed this is the very basis of the generation of virtual auditory space (Fig. 1.1). However, as we shall see, this is beset by a number of nontrivial problems that result in compromises in design and implementation. As a consequence, this becomes an issue where engineering solutions need to be guided by our understanding of the processes of hearing that lead to our perception of sounds. Due to a number of biological and evolutionary constraints, many of the operations of the auditory nervous system are quite nonlinear. Therefore, the challenge is to build efficient devices which result in this illusion of auditory space by matching up the necessary engineering compromises and biological constraints. This kind of challenge can only be effectively met when there is a close association between auditory neuroscientists, psychophysicists and engineers. It is hoped that this book may make something of a contribution to this association.

1.2. DIMENSIONS OF THE PERCEPTION OF AUDITORY SPACE

Under normal listening conditions, the perception generated in a listener by a sound emitted from a single source is generally that of a particular auditory object. It has been argued that the auditory system has evolved for the detection of objects which generally correspond to sources of sounds. Consequently, an auditory object is mapped onto the physical attributes of its source. A talker is a person (or possibly an electronic sound source), a bark is a dog, a snap can be a twig breaking etc. However, when we as observers and commentators begin to classify the perceptual qualities of an experience we begin to indulge in a theory-dependent exercise. That is, we necessarily make assumptions about the nature of the stimulus and begin to map perceptual quality onto presumed physical attributes. The interpretation that we place on our perception then is inextricably linked to our expectations about the world. For instance we can easily attribute the source with a spatial position with respect to the listener. In localizing a sound source we assign a two dimensional direction to the sound source and we estimate how far away the source is.

Things get a little more complicated when we consider factors such as the extent of the source. The idea that a sound has a spatial extent could possibly be mapped onto some notion of the size of the object emitting the sound. However, there is a body of psychophysical work

---

*a This is a point sometimes missed in the design of high fidelity audio systems where the drive for system linearity can result in over-engineering when compared to resolution of the final receiver, the human ear.*
which indicates that extent tells us something about the environment within which we are listening to the source.\textsuperscript{4} The term ‘spaciousness’ has been coined, particularly by architectural acousticians, to describe this perceptual quality (see Blauert\textsuperscript{5} for discussion). The great concert halls of the world are designed with the physical attributes necessary

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{(a) When we listen to sounds over headphones the source of the sound is generally perceived to be inside the head. If we vary the signals at each headphone so that, as in the case illustrated in the figure, the signal is of greater amplitude in the left ear and arrives earlier in the left ear, the apparent source of the sound will appear closer to the left ear. (b) If a sound source is located off the midline in free space, in this case close to the left ear of the listener, the sound will be of greater amplitude in the left ear and arrive earlier in the left ear. In contrast to the figure above, under normal listening conditions, the sound is also filtered by the outer ear before it is encoded by the auditory nervous system. These effects are illustrated by the time/pressure graphs on each side of the head and represent the pressure waves generated by a particular sound source. In this case we perceive the sound to be located in free space away from the head. (c) To generate the illusion of a sound in free space, the pattern of sound waves that would have been produced by a sound in free space is generated over headphones. This is achieved by taking into account the normal filtering effects of the outer ear. In this case, the illusion is generated of a sound source at a particular location outside the head. Reprinted with permission from Carlile S and King AJ, Cur Biol 1993; 3:446-448.}
\end{figure}
to generate this quality. In a large concert hall the sense of spaciousness results primarily from an interaction between the primary incident wavefront, generated by the performer, and the lateral reflections combined with the reverberation. When a sound is ‘spacious’ the listener feels surrounded by or immersed in the sound and, at least for music, this tends to increase the emotional impact of the sound. The amount of reverberance in an environment determines to some extent the ability to localize a single source. Therefore, in some sense, the spaciousness of a sound is at one end of a perceptual dimension where accurate localization of a discrete source is at the other.

The foregoing discussion serves to underline the important notion that sound alone is not necessarily sufficient for the generation of our perception of our auditory world. Certainly, sounds generate in us certain sensations but the perception that results from these sensations can be dependent on other factors. These can include the expectations we have about the nature of the sound sources and the environment within which we are listening to these sources. These are sometimes referred to as ‘cognitive’ factors or ‘top down’ elements of perception. However, as we shall see later, auditory perception is also dependent on other factors which are not necessarily ‘cognitive’ in origin.

With these cautions in mind we can start to draw some preliminary conclusions about the dimensionality of our perception of auditory space. If we initially restrict our considerations to a single sound source in an anechoic field, then the two obvious perceptual dimensions are direction and distance of the source relative to the head. The direction can be indicated using a familiar coordinate system such as azimuth angle with respect to the frontal midline and elevation angle with respect to the audio-visual horizon. The perception of distance is relative to our egocentric center.

---

*b A highly reverberant environment results in an increase in the incoherence of the sounds waves at each ear, thereby degrading the acoustic cues used by the auditory system in determining spatial position (see chapter 2, section 1).

*c The emotional impact of music in such situations may be related to our restricted ability to localize the source. In an evolutionary context, accurate localization of a particular source might have very important survival consequences. In fact the principal evolutionary pressure on hearing may well be the ability to localize a presumed predator (or prey). In the musical context the inability to localize the source and indeed the notion of being immersed or consumed by the source may add some emotional frisson to the experience.

*d This co-ordinate system relies on a single pole system like that used to describe global location on the planet. Other coordinate systems are sometimes employed to describe sound location and these are described in greater detail in chapter 2, section 1.4.2.
1.3. The Nature of the Auditory Stimulus

Much of the previous research effort in auditory neuroscience has followed a formal reductionist line and concentrated on how the auditory system encodes simple sounds. The stimuli used in such experiments are typically short bursts of white noise or pure tones, sometimes modulated in frequency or amplitude. These are generally presented using headphones or closed field sound systems sealed into the external canal. The underlying methodological philosophy assumes that if the system acts linearly and superposition applies, then the encoding of more complex sounds could be understood in terms of the encoding of ‘elementary’ sounds. Over the last 100 years or so this approach has provided a considerable amount of important information about auditory processing of simple sounds but there is not yet a complete picture. However, what is beginning to become clearer is that this simple ‘bottom up’ approach may not be able to provide the necessary intellectual and methodological tools to examine the processing of ecologically appropriate stimuli.

Recent work has reinforced the notion that the principle of superposition does not necessarily apply to the analysis of many combinations of sounds. The auditory system tends to analyze sounds differently depending on various parameters of the sound; for instance, when we are dealing with very short duration sounds the binaural auditory system tends to analyze sounds synthetically rather than analytically.\(^6\) In the latter type of processing, sound is broken up into various components (frequency, level, time of arrival) and then parsed into different potential auditory objects. In contrast, synthetic processing tends to result in a single auditory object whose characteristics are dependent on some type of vector sum of the components. These different modes of processing may well have an ecological rationale. If we accept that very short or transient sounds are unlikely to have resulted from a combination of sources—the inadvertent sounds made by a predator may well fit into this category, for instance—the efficiency of processing the location of this sound may be of paramount importance. Synthetic processing of a sound may be the result of such a strategy. On the other hand, analytic processing is likely to be more computationally expensive and therefore, time consuming. Such a strategy may be reserved for longer duration sounds such as communication sounds which require discrimination along other dimensions.

A further limitation of a simple reductionist approach is that the elementary stimuli are unlike most sounds that are likely to be encountered in the ‘real world.’ One point of view is that the auditory system never evolved to detect and analyze such sounds.\(^3\) Following from this, it would seem questionable as to whether probing the system with such sounds will lead to clear picture of its normal processing. There is no doubt that the system can encode such sounds but the question is whether the responses one elicits with such stimuli
bear much relationship to the kinds of processing of more ecologically valid stimuli. Certainly, the perceptual experience generated by such stimuli is clearly impoverished. Sounds presented over headphones are generally perceived as coming from a phantom source within the head rather than outside; that is, they have zero egocentric distance. By varying a number of characteristics of the sounds at each ear, the apparent source of the sound can be made to move closer to one ear or the other but still lacks any 3D directional quality. Such sounds are said to be lateralized within the head rather than localized in external space. There are very few natural listening experiences that result in such an auditory illusion.

So what are the advantages of using a headphone stimulus system? The short answer is one of stimulus control. By delivering sounds over headphones it is possible to carefully control the characteristics of the sound delivered to each ear. This makes possible a highly reproducible stimulus and greater rigor in experimental design. This kind of stimulus control also makes possible a whole class of experiments which would be impossible using a sound presented from a loudspeaker in the free field. As we shall consider in some detail below, the mammalian auditory system has two ears, each sampling the sound field under slightly different conditions. The differences in the inputs to each ear are used by the auditory system in a variety of tasks; for instance, determining the location of a sound source or separating out a sound of interest from background noise. Using headphones, the differences in the sounds at each ear can be manipulated in ways which would be very difficult using sound sources placed away from ears in the free field. So, despite its obvious perceptual limitations, the closed field or headphone presentation of stimuli still provides a powerful experimental tool.

1.4. A VIRTUAL SOUND FIELD

1.4.1. Generation and utility

If we use headphones to generate a sound field at a listener’s ear-drums that is identical to the sound field that is normally generated by a sound source in the free field, then the listener should perceive the sound source as existing in the free field; that is, in virtual auditory space (VAS; Fig. 1.1). In contrast to the stimulus methodology described in the previous section, a complex sound presented in virtual auditory space is a highly ecological stimulus. Under properly controlled conditions, the percept generated in the listener is of a sound emanating from a source located away from the head at a particular location in space. Clearly, this is also an illusion, but in this case the illusion is one which better approximates the normal listening experience. From a research point of view, stimuli presented in virtual auditory space promise to provide a most powerful tool for investigating many important and outstanding questions. Such a presentation method
combines the stimulus control offered by headphones together with the ecological validity of a real free field sound source. Additionally, as these signals are usually generated using digital signal processing techniques and fast digital-to-analog converters, it is a relatively simple task to perform complex manipulations of the signals before presentation (chapter 3).

The usefulness of this technique for auditory research is almost entirely dependent on how well the illusory sound field corresponds to the real free field. Clearly, any experiment which relies on manipulation of the virtual sound field to expose auditory processing strategies will be confounded if the original virtual field is a poor approximation to a real free field. Chapter 4 canvases some of the difficult acoustical issues involved in generating high fidelity VAS which result in acoustical compromises in its implementation. Therefore, the question of fidelity of a virtual sound field is principally a perceptual issue rather than an acoustical issue. As such, it becomes operationally defined and based on some behavioral or psychophysical test. In the remainder of this section we will consider what kind of psychophysical tests might be used to determine the fidelity of a virtual sound field.

1.4.2. Tests of fidelity

One of the most clearly understood aspects of auditory behavior relating to a sound field is the capacity of a subject to localize the source of a sound within that field. Thus, the fidelity of VAS could be determined by comparing the ability of a subject to localize an auditory target within VAS with that in the free field. However, there are a number of important factors that need to be considered if such an approach is to be useful. For instance, it is well known that there are differences between individuals in their accuracy of sound localization\(^7\)\(^-\)\(^10\) (chapter 1, section 2), therefore this factor should also be taken into account when assessing VAS fidelity.

The type of localization task used in such a test is also important. Clearly, the power of any test is related to the specificity of the question that is asked and, in the context of auditory localization, the mechanisms that are tested are intimately related to the kind of stimulus that is employed. The simplest form of localization relies on a *homing* strategy. In this case the sound detector need only be able to code stimulus level and its output integrated with movement of the detector throughout the sound field. The only requirement for the target stimulus is that it be continuous or at least very repetitive. *Scanning* the sound field is a second and slightly more sophisticated localization strategy. In this case the sound receiver has to be directional but it need only be rotated in the sound field. Thus, scanning is not dependent on translocation of the receiver with respect to the source. Again, sound level is encoded and integrated with rotation of the receiver to provide the directional information.
If the duration of the stimulus is very short and nonrepetitive these two localization strategies will fail. The capability to localize the source of a transient stimulus represents a much more sophisticated capability than that of homing or scanning. Biologically, this is achieved by using two receivers which sample the sound field under slightly different conditions; in the case of the mammal, the two ears are generally found on each side of the head. The inputs to the two ears are compared by the auditory system (binaural processing; see chapter 2) to extract a variety of cues to the location of the sound source. A whole range of auditory stimuli can be localized by such a mechanism but its particular utility is in the localization of sounds which are so brief that homing and scanning strategies are not possible. However, for some stimuli, such as pure tones with no amplitude or frequency modulations, even this localization mechanism is not perfect and can lead to large errors. Its not surprising that narrow frequency band sounds are often exploited as warning signals by groups of animals. There is a clear evolutionary advantage for a group of animals in being made aware of the presence of danger such as a predator. However, there is clearly no individual advantage if such a warning signal can be easily localized and the hapless sentry exposes himself to attack! So the choice of warning signals which are particularly difficult to localize represents the evolutionary compromise.

Transient stimuli also represent a special class of stimuli which are likely to have a high ecological significance. The inadvertent sounds of approach, particularly in an environment with plenty of vegetation, are most likely to be generated by snapping twigs or rustling of leaves. Both are short duration sounds containing a wide range of frequencies. Indeed, the shorter the duration of the transient, the closer it approximates a delta function and the broader the range of frequencies that it contains (see chapter 3). Such sounds might result from inefficient stalking by a predator and are thus highly significant in terms of survival. In survival terms the most important attribute of such a sound is its location.

The forgoing discussion suggests that the clearest test of the fidelity of a particular virtual sound field would be the capacity of a subject to localize a transient stimulus. Such a stimulus places the greatest processing demands on the auditory system and is dependent upon the widest range of acoustic cues to source location. If a particular virtual sound field fails to provide these cues, presumably because of the compromises made in its implementation, then there should be greater localization error in the virtual field compared to localization by the same subject in the free field. In the following chapters the methods by which localization ability can be assessed will be reviewed and questions of the spatial resolution of these methods will also be considered. Obviously, the methodology employed in such a test of fidelity must be sufficiently sensitive to be capable of detecting
perceptually relevant differences in the virtual and free sound fields. Clearly, if VAS is to be used in advanced auditory research or in mission critical applications, it is insufficient for a designer or engineer to simply listen to the virtual sound field and decide that it satisfies the necessary criteria because the sounds appear to come from outside the head and from roughly the correct locations.

1.5. THE REPRESENTATION OF AUDITORY SPACE IN THE CENTRAL NERVOUS SYSTEM

In the central nervous system the way in which auditory space is coded is very different from the other sensory representations of external space, particularly those of visual space or of the body surface. This has important implications for the way in which we might expect the auditory system to process information and for the specific characteristics of a sound that are important in generating our perception of the auditory world. These fundamental differences in processing also flag a caution about using analogies imported from different sensory systems in our attempts to understand processing by the auditory system.

The fundamental differences between these systems has its origin in how the sensory information itself is encoded. In the visual system, light from in front of the eye enters through the pupil and strikes the light sensitive receptors in the retina at the back of the eye. Thus, the resulting pattern of neural activity in the retina corresponds to the spatial pattern of light entering the eye. Broadly speaking the visual system is working like a camera and takes a picture of the outside world. That is, the visual field is mapped directly onto the retina which then makes connections with the brain in an ordered and topographic manner. Thus, visual representations are said to be topographic in that there is a direct correspondence between the location of activity in the neural array and the spatial location of the visual stimulus. In other words, the spatial patterns of neural activity that occur in the visual cortex correspond directly to the patterns of activity in the retina which in turn correspond to the pattern of light entering the eye.\(^e\)

The primary sensory coding by the auditory system is very different from the visual system. Sound is converted from mechanical energy to neural signals in the inner ear. The inner ear, however, breaks down the sounds as running spectra and encodes the amplitude and phase of each frequency component. Due to a number of biological

\(^e\) This topographic pattern of activity is preserved across a large number of cortical fields but as processing becomes more advanced from neural field to neural field, the topographical pattern tends to become increasingly blurred as this topographic map is sacrificed for the extraction of other important visual features such as motion, form or colour (see refs. 74 and 75).
limitations, the ability to encode phase decreases as a function of increasing frequency. What is most different with this encoding scheme compared to the visual system is that the spatial pattern of neural activity across the auditory receptors (and subsequently the auditory nuclei in the central nervous system), reflects the frequency content of the sound and not the spatial location of the source. Therefore, the processes that give rise to neural representations of auditory space and indeed our perception of auditory space must be based on other information that is extracted from the auditory inputs to one or both ears. That is to say, space perception is based upon a highly computational neuronal process. In the visual system, the sensory ‘primitive’ is the location of the origin of a ray of light and the emergent perceptual components are, say form or motion. By contrast, for the auditory system the sensory primitive is sound frequency and space is one emergent component. Thus, while the auditory nervous system clearly generates some kind of a representation of auditory space, the mechanisms by which this arises are very different to how space is encoded in the other senses that deal with the place of objects in the external world.

1.6. AN OVERVIEW OF THE FOLLOWING REVIEW

The foregoing introduction has been necessarily eclectic and discursive in an attempt to illustrate the range of issues that should be considered when the proper implementation and applications of virtual auditory space are considered. In the following sections of this chapter and in the following chapter we shall consider, in a more systematic and comprehensive manner, many of the issues that have been touched on above.

Auditory localization of single sound sources under anechoic conditions is probably the best understood process involving a free field sound field. As is suggested above, performance testing based on this process probably provides the best test of the fidelity of a virtual sound field. For this reason, in the second section of this chapter we will review the current state of knowledge of human sound localization abilities. Humans seem to localize a sound source quite accurately although some other nocturnal predators, notably the owl, do somewhat better. It may be that such performance differences result from differences in processing strategies by the auditory nervous systems of these animals. However, while there are known to be structural differences in the auditory nervous systems of the owl compared to that of the human, it is not clear whether these differences simply reflect a different evolutionary heritage or real differences in the processing strategies. Another very important difference underlying the variations in localization performance between species is likely to be in the quality of the acoustic cues to spatial location that are generated at the outer ears. For instance, there are major differences in the structures of the ears of owls and humans and the acoustics of these structures are known
in some detail. In chapter 2 we will consider in some detail the physical cues to sound location that are generated at the auditory periphery of the human.

The major theories of how we perceive sounds in auditory space have been built upon an understanding of the physical cues to sound location and capacity of the auditory system to encode those cues. There are a number of important biological limitations to the processes by which these cues are encoded by the central nervous system. Therefore, the fact that a possible cue is present at the auditory periphery by no means indicates that this cue is utilized by the auditory nervous system. In the second section of chapter 2 some of the experiments that have examined the sensitivity of the auditory system to the physical cues to a sound’s location are examined.

2. SOUND LOCALIZATION BY HUMAN LISTENERS

2.1. ACCURACY AND RESOLUTION IN AUDITORY LOCALIZATION

2.1.1. Introductory observations

There are two main approaches to assessing the capacity of the auditory system to localize a sound source; (i) assessing absolute localization accuracy,\(^9,12\) or (ii) determining the minimum audible change in the location of a stimulus, the so-called minimum audible angle (MAA).\(^13,14\) An important distinction between these two approaches is that the first examines localization ability per se, while the second examines the capacity of the auditory system to detect changes in any or all of the cues to a sound’s location. That is, in a MAA experiment, two stimuli may be distinguished as being different (by virtue of the small differences in spatial location) but the subject may still be incapable of accurately localizing either of the stimuli or even assessing the magnitude or direction of the vector of difference. While the assessment of MAA can provide important information about the quantum of information in auditory processing, it does not necessarily relate to the processes that lead to our perception of auditory space. On the other hand, the detection of small differences associated with slight variations in the locations of sources may provide insights into other auditory processes that rely on differences in the signals arriving at each ear. For the remainder of this section we will concentrate on experiments that have examined absolute localization accuracy rather than MAA.

There have been a number of studies examining sound localization accuracy and several excellent recent reviews.\(^15-18\) Rather than going systematically through this large literature I will concentrate on some of the general issues that have importance for the generation, validation and applications of virtual auditory space.
One general observation is that, to date, most localization experiments have been conducted under quite controlled acoustic conditions. Clearly the motivation for such an experimental approach is the desire to move from the simple to the complex in experimental design. In these experiments the testing environment is generally anechoic, the stimuli are generally broadband and of short duration and presented from a fixed number of source locations to a subject whose position is also generally fixed in space. As a consequence, after the first few stimulus presentations, the subject will have considerable knowledge about the stimulus spectrum and the acoustic environment. This is of course a very unnatural listening situation in that most sounds of interest are likely to have time variant spectra and the listening conditions are also likely to be constantly changing with head movements, variations in the number and locations of other sound sources and the variation in the geometry of reflecting surfaces as one moves about the environment. Thus while current work may provide insights into the limits of our sensory coding of auditory space, we should remain cautious about what the current state of knowledge can tell us about sound localization in a real world situation.

2.1.2. Methodological issues

There are two main methodological issues that need to be considered: (i) how the position of a sound source is varied; and (ii) how the subject indicates where the sound source is perceived to be. These issues are discussed in some detail in chapter 4 (section 5.1) and are only briefly considered here. Varying the location of a test stimulus has been achieved by using either a static array of possible sources or by using a single moveable sound source placed at a number of locations about the subject. In the first case it is often possible for the subject to simply indicate a number identifying which speaker a stimulus was perceived to have come from. In the second case, as there is only a single target, the localization experiments are usually carried out in complete darkness and the subject is required to indicate the location of the source by pointing or noting the location coordinates in some way. A number of studies have shown that localization performance can be influenced by foreknowledge of the potential target locations as would be the case when a subject is faced with an array of speakers from which it is known that the target will come. Under normal conditions localization is a continuous spatial process so that constraining or quantizing the subject’s responses places artificial bounds on the subject’s behavior and may also bias our analyses of this behavior (chapter 4, section 5.1).

For these and other reasons discussed later we will be principally concerned here with those studies that have used continuous variations in the location of the target. These studies have used a variety of techniques to indicate the sound location including point-
ing with a hand held gun, pointing the face towards the target and tracking the position of the head or simply having the subject call out the coordinates of the apparent location of the source. We have found that, with appropriate training, pointing the head towards the target and tracking the head location is a highly efficient and reliable method for indicating perceived target location\textsuperscript{10} (Fig. 1.2).

2.1.3. Two types of errors in absolute localization

Using brief bursts of broadband noise, two different types of localization errors can be demonstrated:

(i) large localization errors associated with a front-to-back or back-to-front reversal of the apparent target location; that is, the location of the target is indicated correctly with respect to the median plane but the front-back hemisphere is confused.

---

Fig. 1.2. The figure shows a subject inside the anechoic chamber at the Auditory Neuroscience Laboratory (University of Sydney). The subject stands on a raised platform in the center of the chamber. The robot arm, which carries the sound source, is suspended from the ceiling such that rotation of the vertical frame varies the azimuth location of the source. The inner hoop, which actually carries the speaker, is driven by small stepper motors on either side of the main frame; one such motor and its gearing can be seen to the left of the picture. The task of the subject in a localization experiment is to turn and point her nose at the speaker at each test location (the experiments are carried out in complete darkness). The small cap on the head of the subject carries a 6 degrees of freedom tracking receiver which indicates the location of the head and the direction towards which the subject is pointing. The subject indicates completion of each task by pressing the button seen in this subject’s left hand.
(ii) variations in the perceived location relatively close to the actual target.

The difference in the character of these errors implies the failure of different localization processes. When broadband stimuli are used, front-back localization error is found in between 6%\(^{5,9}\) and 3%\(^{23}\) (see ref. 12, 4%—also Parker, personal communication; ref. 7, 5.6%) of localization judgments.\(^{8}\) This indicates that, under these listening conditions, there is some ambiguity in the perceived location of the sound source. It is important to note that these kind of errors are likely to be strongly affected by the foreknowledge that subjects may have about the configuration of potential stimulus locations, particularly where these are limited in number or limited to a particular spatial plane.\(^{24}\) Such foreknowledge may help the subject in resolving the perceptual ambiguity of some stimuli so that their performance no longer represents the simple perceptual aspects of the task.

Regardless of the experimental approach, a general observation for broadband stimuli is that the accuracy of localization varies as a function of the location of the target (Fig. 1.3). In general, human subjects demonstrate the smallest localization errors and the smallest minimum audible angles for targets located about the frontal midline at around the level of the audio-visual horizon (the plane containing the eyes and the interaural axis). In studies using continuous variation of sound locations, the absolute accuracy of localization varies across studies, presumably reflecting methodological differences such as the spectral content of the stimulus and the method of indicating the location of the stimulus.

We have found,\(^{10}\) using the head pointing technique described above, that for sound locations on the anterior midline and ±20° about the audio-visual (AV) horizon the variation of the horizontal component of localization is between 2° and 3° with the variation in the vertical estimates between 4° and 9°. However, for locations around the interaural axis the horizontal variation increases to between 8.5° and 13° and variation in the estimates of the vertical locations is between 6° and 9°. For posterior locations close to the AV horizon the variation in

\(^{5}\) This may represent an underestimate of the number of front-back confusions in this study as Makous and Middlebrooks did not test locations directly behind the subject.

\(^{8}\) There is no generally accepted definition of what constitutes a front-back confusion for locations close to the interaural axis. For instance if a sound is located 10° behind the interaural axis but is perceived to be located in front of the interaural axis, does this represent a front-back confusion? The small differences in the front-back confusion rate may well reflect variations in the criteria of identification between studies. The main point from these data however, is that the front-back confusion rate is relatively low across all of the studies.
the estimates for horizontal components ranges between $8^\circ$ and $12^\circ$ and for the vertical components between $7^\circ$ and $10.5^\circ$. In general the errors in localization increase towards the extremes of elevation. Makous and Middlebrooks\textsuperscript{9} report similar variations in localization accuracy to those we have found, although in that study the errors reported for the posterior locations were generally larger. Although there are a number of differences in the stimuli between these and previous studies, the

![Fig. 1.3. The mean localization accuracy from 9 subjects is shown together with an estimate of the variance of the location estimates. Each sphere represents the hemisphere of space surrounding the subject as indicated on each plot. The filled circle indicates the location directly in front of the subject (Azimuth 0° Elevation 0°). The actual location of the target is indicated by the small cross at the origin of each ray. The center of each ellipse indicates the mean location (azimuth and elevation) of six localization trials for each subject. The variance of the azimuth and elevation components estimate is indicated by the extent of the ellipse. The distributions of the localization estimates for each target position are described by a Kent distribution.\textsuperscript{72} Data from Carlile et al.\textsuperscript{10}](image-url)
smaller localization errors found in these two studies compared to previous studies probably reflect differences in the methods by which the subjects indicated the perceived location of the target (see chapter 4, section 5.1). Furthermore, it is also not entirely clear to what extent the spatial variations in the localization acuity can be attributed to sensory limitations or to the methods employed by the subjects to indicate the perceived location (however see refs. 9, 10). However, the fact that a consistent general pattern of the spatial variation of localization accuracy is seen across all studies using very different methodologies supports the notion that these differences are, in a large part, attributable to sensory effects.

2.1.4. Localization accuracy is dependent on the stimulus characteristics

Another general finding that emerges from numerous different studies is that the ambiguity of a sound’s location increases when the bandwidth of the stimulus is restricted. This is manifest as an increase in the number of front-back confusions. Decreasing stimulus bandwidth also results in a general decrease in localization accuracy. Butler found that, following correction of the front-back confusions, there was a progressive increase in localization error as the bandwidth of a noise centered on 8 kHz was decreased from 8 kHz to 2 kHz. These observations indicate that, for accurate localization, spectral information across a wide range of frequencies is required.

The complex folded structure of the outer ear has also been shown to play a very important role in this process (Fig. 1.4). Increases in the number of front-back confusion errors have also been reported when the concavities of the outer ear were filled with plasticine but the auditory canal was left patent. This further demonstrates that the interactions between a broadband stimulus and the structures of the outer ear also provide important localization cues. As is discussed in detail in chapter 2, the outer ear filters the sound across a wide range of frequencies. The exact characteristics of this filtering vary as a function of the location of the source, so providing the so-called ‘spectral cues’ to sound location.

The link between the spectral characteristics of a sound and its location has been examined extensively in the context of sound locations on the median vertical plane. On the basis that the head and ears are symmetrically arranged, it has been generally argued that interaural differences are uniformly zero for median plane locations; thus the elevation of a sound’s location on this plane must be indicated by variations in the spectral content of the signal produced by pinna filtering. However, from an analysis of localization data using a decision theory approach and careful acoustical recording from each ear it is clear that, at least for the subjects examined, there are often marked acoustical asymmetries that lead to significant interaural level differences for sounds on the median plane. Notwithstanding this
A number of studies have indicated that the apparent location of a sound source on the median plane can be varied by manipulating the stimulus spectrum rather than the actual location of the source.\textsuperscript{4,35,36} The perception of the vertical location of sounds presented over headphones is associated with the spectral ripple produced by comb filtering using a delay and add procedure.\textsuperscript{37} Such an approach was suggested by the work of Batteau\textsuperscript{38} who argued that sound locations could be coded by multiple delays provided by the complex sound paths of the outer ear. Although he suggested a time domain analysis of the input signal it seems more likely that the auditory system analyzes the resulting comb filtered inputs in the frequency domain\textsuperscript{39} (see chapter 2 section 1.8.2 and chapter 6 section 2.3.2). Consistent with the role of the outer ear in providing these spectral cues, manipulation of the outer ear by filling the concavities of the pinna has also been found to reduce localization accuracy for sounds on the median plane.\textsuperscript{19,35,40,41} However, some care must be taken in interpreting many of these data as most studies have employed a small number of visible sound sources and thus constrained the subject’s response choices (see section 2.1.2).

Fig. 1.4. A simple line drawing showing the main features of the complexly convoluted structure of the outer ear. The main functional components are (a) the pinna flange comprising helix, anti helix and lobule, (b) the concha including the cymba and the cavum and (c) the ear canal connecting to the floor of the concha (not shown). Adapted with permission from Shaw EAG. In: Keidel WD et al, Handbook of Sensory physiology. Berlin: Springer-Verlag, 1974:455-490.
2.2. Localizing Sound Sources with One Ear

So far we have considered sound localization using two ears, but it has been known for some time that individuals with only one functional ear can also localize sounds with reasonable accuracy. Many studies using normal hearing subjects but with one ear plugged have also demonstrated reasonable localization accuracy for targets in both horizontal and vertical planes. In subjects who were artificially deafened in one ear, sound locations along the horizontal plane tended to be displaced towards the functional ear so that localization accuracy was good for locations about the interaural axis on the unblocked side and increasingly less accurate for locations displaced from these locations. Where subject responses were unconstrained, vertical localization does not seem to be as affected as the perception of horizontal location.

Monaural localization has also been found to be dependent on the spectral content of the stimulus and is very inaccurate for stimuli low-passed at 5 kHz. Monaural localization is also disrupted by manipulation of the pinna. For the monaural subject, the apparent locations of narrow band stimuli seemed to be determined by their center frequency rather than their actual location. However, both practice effects and context effects have also been shown to influence the subject’s responses. There is evidence that some subjects with a long history of unilateral deafness perform better than subjects with one ear blocked, particularly with respect to displacement of apparent sound locations towards the hearing ear. Mixed localization accuracy was also reported for a group of 44 unilaterally impaired children when compared to 40 normally hearing subjects with the hearing impaired children showing a greater range of localization errors for a noise high-passed at 3 kHz. Whatever the basis of these differences in accuracy, the principal finding is that monaural subjects can localize reasonably well and must do so on the basis of the filtering effects of the outer ear.

2.3. Dynamic Cues to the Source of a Sound

2.3.1. Head motion as a cue to a sound’s location

There have been a number of studies examining the contribution that head movements or movement of the sound source play in localization accuracy. The basic idea is that multiple sequential sampling of the sound field with ears in different locations with respect to the source would provide systematic variation in the cues to a sound’s location. For instance, the pattern of the variations in the binaural cues could be used to help resolve the front-back confusion. The availability of such cues is, of course, dependent on a relatively sustained or repetitive stimulus to allow multiple samples. When naive subjects are attempting to localize long duration stimuli they do tend
to make spontaneous head movements of the order of ±10°, particularly when the sound is narrow band. However, some what surprisingly, there is little evidence that, for a binaurally hearing individual, head movements contribute significantly to localization accuracy under normal listening conditions. Some improvements are seen where the cues to a sound’s location are impoverished in some way.

Head movements have been shown to substantially increase monaural localization accuracy of a 3.7 kHz pure tone. Induced head movement (in contrast to self induced movements) showed some increase in localization accuracy where the noise or click stimuli were high- or low-pass filtered. Fisher and Freedman showed that self induced head movement produced no improvement in the localization of a small number of fixed sound sources. The elevation estimates of low-pass noise were reported to be very poor and, despite expectations to the contrary, allowing subjects to move their heads during the presentation of a long duration stimulus did not result in any improvement. Pollack and Rose confirmed the finding that small head movements have no effect on localization accuracy but found that when a subject turned to face the source of the sound there was an increase in localization accuracy. This last result may have more to do with the location dependent variations in localization accuracy discussed above rather than a contribution of a head motion cue to source location per se.

Thus, despite strong theoretical expectations to the contrary, there is almost no evidence that head movements are useful in localizing a free field sound source unless the bandwidth of the sound is narrow or the spectral cues to location are degraded in some other way. This suggests that, at least under the experimental conditions examined so far, the auditory system does not re-sample the sound field as a cue to location. This may be related to the fact that most subjects already have two simultaneous samples of the sound field (one from each ear). Furthermore, the system is only likely to gain more information by re-sampling if the characteristics of the stimulus are stationary. In contrast to the kinds of sounds used in these experiments, natural sounds are highly nonstationary in both their temporal and spectral characteristics. Under such conditions, variations in the subsequent samples of the sounds that result from rotation of the head could be confounded by the variation in the characteristics of the source. Thus, in the ecological context in which this system has evolved, re-sampling probably represents a computationally expensive and yet largely redundant strategy.

2.3.2. Perception of the motion of a sound source

In contrast to what is known about localization of static sources, considerably less effort has been expended examining issues of auditory motion. Furthermore, many previous studies of auditory motion are limited by a number of technical and theoretical problems outlined
below. There is also considerable disagreement as to the mechanisms of motion analysis, some of which may be traced to differences in methodology.

Many studies have employed simulated auditory motion using head-phones because of the technical difficulties associated with silently moving a physical sound source in the free field. Variations in the binaural stimulus parameters result in variations in the lateralized sound image within the head and therefore this experimental paradigm suffers from the limitation that sound image has no externalized 3D spatial location. As with studies of auditory localization, the generalizability of such experiments to free field listening conditions is questionable. Other methods of simulating motion using static sources rely on stereo-balancing between two widely spaced free field speakers or rapid switching between relatively closely spaced speakers. These methods generate a more compelling percept of auditory motion in that the sound image occupies extra-personal space. However, as discussed above, the generation of the percept does not necessarily demonstrate that a particular simulation method generates all of the relevant cues to auditory motion. Stereo-balancing involves a reciprocal variation in the level of the stimulus at each speaker. This results in a continuous variation in the loudness of the sounds in each ear. However, this method will not produce appropriate variation of the location dependent filtering effects of the outer ears that results during an actual variation in the location of a source. Accurate localization processing of a static stimulus requires the conjunction of appropriate binaural and monaural cues so that the cue mismatches produced by stereo balancing might also disrupt some aspects of motion processing.

Cue mismatch may not be a problem for movement simulations relying on rapid switching between closely spaced speakers. This technique assumes that the distance between speaker is within discriminable limits and that the rate of switching is within the ‘sampling period’ of the auditory system. The first assumption can be confirmed by studies of localization of static signals discussed above but the second has yet to be experimentally demonstrated. A problem that can arise with rapid switching is the “ringing” that this produces in each speaker at signal onset and offset. This will produce spectral splatter resulting in significant side lobes in the spectra of narrow band sounds or smearing of the spectra of complex sounds. This can be avoided to a large extent by appropriate time domain windowing of the signal. However, both windowing or the off-set ringing in individual speakers will result in a sound ‘source’ that has a much larger spatial extent than a real moving point source.

There have been a number of studies employing a moving loudspeaker. The speaker was attached to the end of a boom anchored above the head which could be rotated around the subject. These studies have necessarily been limited in the trajectories of movement and the
range of velocities that could be examined (e.g., ref. 67) or were restricted to pure tone stimuli (e.g., refs. 68, 69). Pure tones are not a particularly ecological stimulus as most natural sounds are rich in spectral features. More importantly, their use may have negated an important potential cue for auditory motion, namely the location-dependent variations in the peaks and notches in the filter functions of the outer ear. These spectral features are only available with complex, broadband stimuli. Only recently have broadband sounds been employed with actual moving sources\textsuperscript{70} or simulated movements using multiple sources.\textsuperscript{64,65} It is noteworthy that, where comparison between studies is possible, the minimum audible movement angle (MAMA) is considerably less with broadband stimuli than with pure tones.

In this chapter we have seen that our perception of auditory space is dependent on a range of auditory and non auditory factors. The generation of virtual auditory space promises to provide a very powerful research tool for the study of this important perceptual ability. A key determinant of the utility of VAS is its fidelity. While the generation of VAS is simple in conception, its implementation involves a number of acoustic compromises. It has been proposed here that a behavioral measurement of the localization ability of subjects listening to short duration noise stimuli presented in VAS represents an appropriate measure of that fidelity. A review of the literature examining such localization behavior reveals that there are individual differences in ability and that nonauditory factors can play an important role in localization performance. Therefore, adequate tests of VAS fidelity using auditory localization tasks need to take these factors into account.

In the next chapter we will consider the physical cues to a sound’s location that are available to the auditory system. However, demonstrating the presence of a particular physical cue does not necessarily imply that the auditory system utilizes this cue. Some physiological models of the encoding of these cues will be described and psycho-physical tests examining the sensitivity of subjects to particular physical cues will also be considered. Such studies also provide insights into the limits of sensory coding of these cues and provide important benchmarks for the acoustic precision with which VAS needs to be generated.

**ACKNOWLEDGMENTS**

I would like to acknowledge Drs. Martin, Morey, Parker, Pralong and Professor Irvine for comments on a previous version of this chapter. The recent auditory localization work from the Auditory Neuroscience Laboratory reported in this chapter was supported by the National Health and Medical Research Council (Australia), the Australian Research Council and the University of Sydney. The Auditory Neuroscience Laboratory maintains a Web page outlining the laboratory facilities and current research work at http://www.physiol.usyd.edu.au/simonc.
REFERENCES
65. Perrott DR, Costantino B, Ball J. Discrimination of moving events which accelerate or decelerate over the listening interval. J Acoust Soc Am 1993; 93:1053-1057.