N.I. Durlach
B.G. Shinn-Cunningham
and
R.M. Held
Virtual Environment and
Teleoperator Research Consortium
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Supernormal Auditory Localization

I. General Background

Abstract

In this series of papers, we consider human auditory localization and how its deficiencies can be reduced by appropriate processing and coding of acoustical signals in teleoperator and virtual-environment systems. Attention is given to how localization cues can be altered to improve the just-noticeable-difference (JND) in spatial position and to phenomena related to the use of such altered localization cues for the identification of spatial position. Unlike most current studies of synthetic auditory localization, our study includes consideration of distance as well as direction. In this first paper of the series, we provide general background material. In subsequent papers, we will present a variety of empirical results.

1 Introductory Remarks

The normal human auditory localization system has relatively good resolution in azimuth (about 1°) for most kinds of signals and reference angles in the neighborhood of 0° azimuth (i.e., straight ahead). However, it has poor resolution in azimuth at angles off to the side, in elevation, and in distance. In all these latter cases, the just-noticeable-difference (JND) in the variable in question constitutes a substantial fraction (i.e., one-fifth) of the total meaningful range of the variable. In other words, the total range encompasses only a few JNDs. This poor resolution raises the question of whether it might be possible to design processing for teleoperator and virtual environment systems that enhances the effective resolution artificially. Even for azimuthal resolution in the neighborhood of 0° azimuth, where natural resolution is best, there might be applications for which still better resolution is desirable.[1] Certainly, the existence of foveal vision did not preclude the development of binoculars.

In a previous paper, Wenzel (1992) summarized the results of research directed toward providing listeners who are being stimulated with acoustic signals over earphones with spatial perceptions similar to those that would occur with real sources located in real environments (so-called "auditory spatialization"). The work described in the present series of papers is motivated by the just-cited deficiencies in normal localization abilities and is directed toward the development of methods for providing listeners with better-than-normal localization ability, so-called "supernormal auditory localization." This first paper in the series provides general background on supernormal auditory localization. Subsequent papers will describe results obtained from psychophysical

1. For a picturesque illustration of a World War I application of "localization enhancement," see Figure 4 in Wenzel (1992).

Presence, Vol. 2, No. 2, Spring 1993, 89–103

experiments on localization performance using altered localization cues and from application studies in which the human–machine interface includes supernormal auditory localization features.

For purposes of this paper, teleoperator and virtual-environment systems can be defined (and distinguished) as illustrated in Figure 1. In a teleoperator system, the human operator interacts with a real environment via a human–machine interface and a telerobot.[2] In a virtual-environment system, the real environment and the telerobot are replaced by a computer simulation. Whereas the purpose of a teleoperator system is to sense, navigate through, and/or operate on the real environment, the purpose of a virtual-environment system is to alter the state of the human operator (e.g., by training or entertaining the operator) or the information environment (i.e., by modifying the state of the computer). Although there now exist blends of the two types of systems (e.g., teleoperator systems in which virtual-environment subsystems are used to construct predictions that are then used to guide actions of the human operator), the basic conceptual distinctions are still valid. A useful diagram of a general modern "telesystem" that includes teleoperator, virtual-environment, and robotic systems as special cases (after Robinett, 1993) is shown in Figure 2.

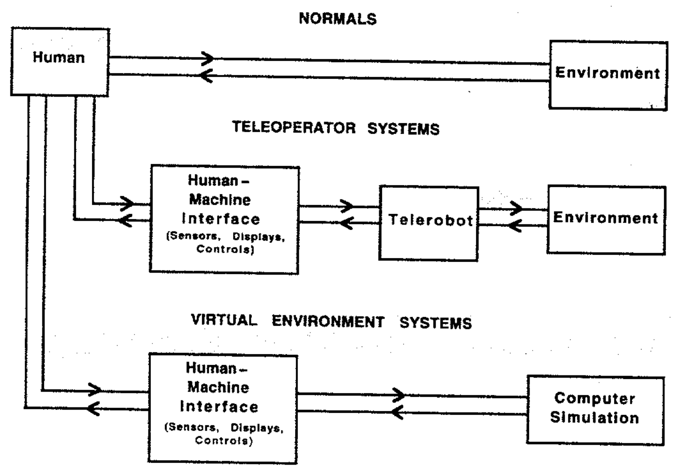

Figure 1. *Schematic outline comparing normal human system, teleoperator system, and virtual-environment system.*

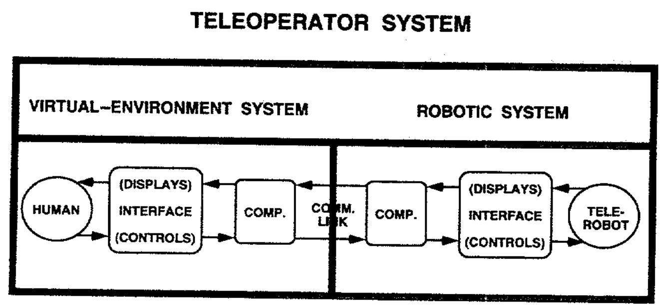

Figure 2. *General "telesystem" (after Robinett, 1993). Whole system provides outline of modern teleoperator system with inclusion of computers on both sides of the communication link; left half alone provides outline of virtual-environment system; right half alone provides outline of autonomous robotic system.*

Two important distinctions between teleoperator systems and virtual-environment systems with respect to supernormal auditory localization are the following.

In a teleoperator system that does not use sensory substitution (i.e., one in which the auditory display conveys information about the acoustic environment of the telerobot), the resolution depends both on the acoustic sensory system employed in the telerobot and on the display of this acoustic information to the auditory system of the human operator. If nonacoustic information is displayed to the auditory system (e.g., optically detected information indicating immediate danger is displayed as an acoustic warning signal or force information indicating mechanical contact is displayed to the auditory system), then, of course, the role played by the acoustic sensing system will be replaced by other sensing systems and the meaning of the auditory localization function will depend on how this nonacoustic information is coded into auditory signals. In a virtual-environment system, the question of resolution at the sensing end of the system is meaningless because the environment is synthesized rather than sensed, but resolution in the display remains important.

2. Sheridan has recommended using the word "teleoperator" for the component of the teleoperator system that we have labelled the "telerobot" in Figure 1 (Sheridan, 1992). We have not adopted this nomenclature because we think it is confusing. Not only do many people, both experts and nonexperts, use the word "teleoperator" (as a contraction of "teleoperator system") to refer to the whole system rather than to a single component, but many nonexperts tend to think "teleoperator" refers to the human-operator component.

Considering first the overall teleoperator problem of sensing and displaying the acoustic environment, we briefly summarize the goals and human capabilities for achieving these goals via the natural human auditory system as shown in Figure 3. First, the system must spatially resolve the acoustic sources. As indicated above, the human auditory system has only a limited capacity to achieve such resolution.[3] Second, for each of these sources, the system must factor (or, in the time domain, deconvolve) the received signal into the transmitted signal and the filtering action performed by the environment (where we include the head and body of the listener, as well as other objects in the acoustic path from source to eardrum, as part of the environment). Remarkably, the ability of the human to exploit a priori information about the world to correctly factor signals (i.e., to determine from the signal received at the two ears the properties of both the transmitted signal and the acoustics of the space through which the signal has passed) has never been studied. Finally, the system must extract information on the sources and environment that is relevant to the task at hand and to the subject's world model, and interpret these elements in terms of this model. Clearly, compared to artificial systems, the human is relatively good at this part of the processing.

A schematic diagram of an idealized sensing system is shown in Figure 4. In this scheme, an array of sensors (microphones), together with appropriate parallel processing, is used to analyze the incoming acoustic field in such a way that each spatial or directional cell (i.e., region in 3-space) is represented by a distinct channel that carries all the signals from that cell, and only the signals from that cell. Subsequent to this spatial analysis, the signal in each channel is subject to a content analysis (e.g., a frequency analysis) to determine the characteristics of the acoustic source in that cell. Perfect spatial resolution is achieved to the extent that no more than one source appears in any given cell (and the volume of each cell is vanishingly small). Information on the acoustic properties of the environment can be obtained by analyzing the results across cells (and, as always, exploiting whatever a priori information is available on the sources and the environment).

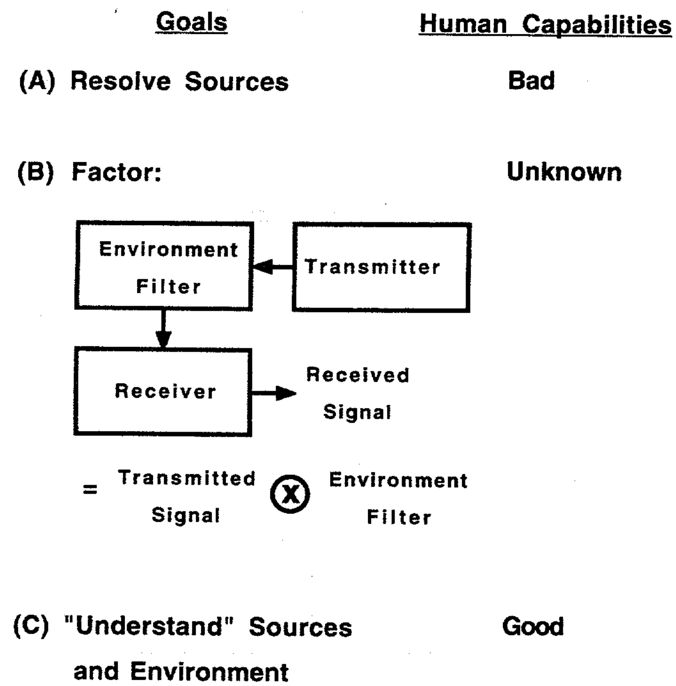

Figure 3. *Outline of general goals and human performance in achieving those goals for a generalized acoustic sensing system.*

A simplified schematic diagram of how the human system is constructed is shown in Figure 5. In this system, only two sensors (ears) are employed; they are spaced relatively close together (relative to the longer wavelengths to which the human auditory system is sensitive), and, unlike the situation in the idealized system, the content (frequency) analysis is performed prior to the spatial analysis. Because the frequency analysis is performed before the spatial analysis, each frequency channel contains energy from all the sources (assuming, of course, that the various sources transmit energy at these frequencies) and the system is faced with the problem of synthesizing the identity and location of the sources from these frequency components.[4] Among the factors

3. In this paper, we use the word resolution to refer both to the ability to detect small changes in the location of a single source and to the ability to separate out multiple simultaneous sources. Although it is possible for a system to have good resolution in one sense but not the other, this distinction is not important for the points made in this paper. In other words, the transformations we consider in subsequent sections of this paper affect both types of resolution in the same manner.

4. If it were not for this feature of the auditory system, and the resultant analysis/synthesis problem, it is extremely unlikely that the current interest in "auditory scene analysis" would have developed (see Bregman, 1990; Yost, 1991).

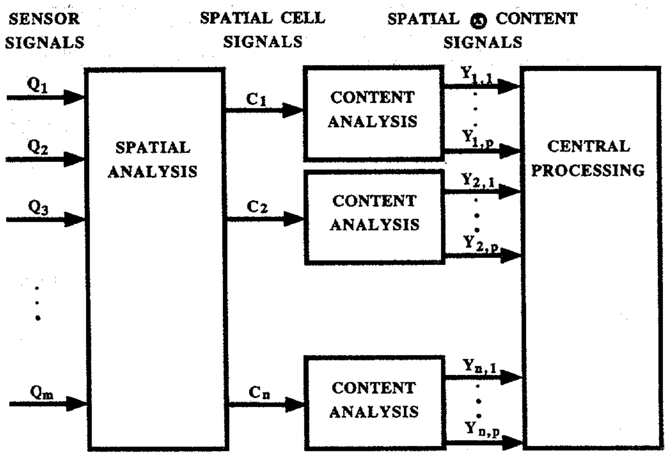

Figure 4. *Idealized acoustic sensing system. In this system, the* m *signals* $Q_1, \ldots, Q_m$ *arising from a microphone array containing* m *microphones are processed to form* n *signals* $C_1, \ldots, C_n$ *corresponding to* n *spatial cells; each of the* $C_1, \ldots, C_n$ *signals is then content analyzed (e.g., by frequency analysis) into* p *channels, and the resulting* n $\times$ p *channels are fed into a central processor.*

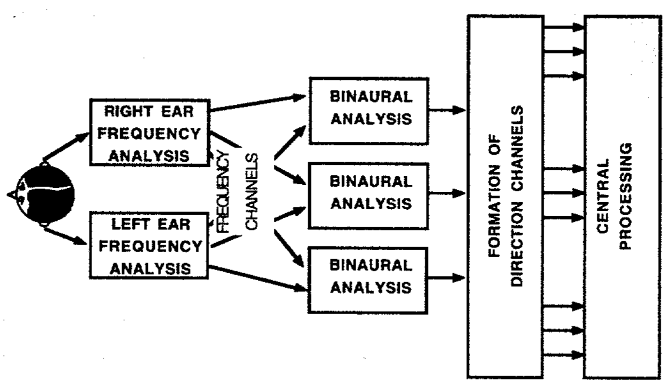

Figure 5. *Schematic illustration of human sensing system. In this system, a frequency analysis is performed prior to a spatial analysis. (For simplicity, the diagram shows only three frequency channels at the output of each ear.) Corresponding frequency channels from the two ears are then fed into a common binaural processor, and the outputs from all such processors are then used to form spatial (or directional) channels and a spectral* $\times$ *spatial map.*

that have been cited as possible underlying causes for this peculiar structure are the economy of mechanical as opposed to neural frequency analysis and/or constraints on head size. In any case, and as can be seen by comparing Figures 4 and 5 (adapted from Colburn, Zurek, & Durlach, 1987), the design of the human system is radically different from the design that would be employed by most engineers for most types of telerobots.

2 Normal Human Localization

In this section, we provide a brief summary of how the natural human localization system performs, independent of how it is constructed. For those readers who find this summary too dense, a more leisurely paced summary (with explanatory diagrams) is available in a previous issue of this journal (Wenzel, 1992).

Consider initially the task of estimating the direction of a single source in anechoic space (i.e., space in which there are no reflections or echoes). The primary method for achieving such an estimate is to compare the signals received in the two ears to determine the interaural amplitude ratio and the interaural phase difference as a function of frequency. Letting

ω = frequency (in radians/sec)
θ = azimuth
ϕ = elevation
$X(\omega)$ = complex spectrum of transmitted signal
$Y_L(\omega, \theta, \phi)$ = complex spectrum of signal received in left ear
$Y_R(\omega, \theta, \phi)$ = complex spectrum of signal received in right ear

one has

$$\begin{aligned} Y_L(\omega, \theta, \phi) &= X(\omega)S_L(\omega, \theta, \phi) \\ Y_R(\omega, \theta, \phi) &= X(\omega)S_R(\omega, \theta, \phi) \end{aligned} \tag{1}$$

where $S_L(\omega, \theta, \phi)$ and $S_R(\omega, \theta, \phi)$, which we refer to as *space filters,* denote the complex transfer functions describing the filtering actions that occur as the sound propagates from the source to the left and right eardrums of the listener. These equations (which specify both the amplitude and phase relations because all factors have been defined in the complex domain) assume that the distance from the head to the source is sufficiently great that it affects only the overall level of Y_L and Y_R (and this dependence is ignored), and that the source can be assumed isotropic. We use the term *space* *filter* here rather than *head-related-transfer function* (HRTF) because (1) in nonanechoic environments the space filter includes the effects of reflections in the environment as well as the effects of the listener's body, head, and pinnae, and (2) when considering super localization systems in Section 2 of this report, we will specify a variety of space filters that are different than the natural HRTFs.

For most purposes, the comparison between the signals received at the two ears can be represented by forming the ratio

$$\frac{\Upsilon_L(\omega, \theta, \phi)}{\Upsilon_R(\omega, \theta, \phi)} = \frac{S_L(\omega, \theta, \phi)}{S_R(\omega, \theta, \phi)} \qquad (2)$$

The absolute value of this ratio gives the interaural amplitude ratio and the phase of this ratio gives the phase difference between the two ears. To complete the binaural processing, the system must use this ratio to form an estimate of the direction (θ, ϕ).

For narrowband signals, directional discrimination by interaural comparison is achieved primarily from interaural phase information at low frequencies (<1500 Hz) and interaural amplitude information at high frequencies (>1500 Hz). Inasmuch as interaural phase becomes ambiguous at the higher frequencies and amplitude differences become negligible at the lower frequencies, these results are theoretically predictable. For signals that have significant bandwidth (so that the phase ambiguities can be eliminated by looking across frequency), interaural time delays can also be effective at high frequencies.

Although interaural processing has the virtue of providing useful information over all frequency regions and of not requiring a priori information about the transmitted signal [because $X(\omega)$ cancels out when the interaural ratio $\Upsilon_L(\omega, \theta, \phi)/\Upsilon_R(\omega, \theta, \phi)$ is formed], it suffers from directional ambiguities (e.g., front–back confusions). To a rough approximation, these ambiguities can be predicted by assuming the head is a sphere with sensors located at the ends of a diameter and then determining the three-dimensional surfaces over which the ratio $\Upsilon_L(\omega, \theta, \phi)/\Upsilon_R(\omega, \theta, \phi)$ is a constant. Under this assumption, the surfaces of ambiguity are cones whose axes are collinear with the interaural axis. The median plane, throughout which the interaural ratio is unity, constitutes a limiting case of this family of cones. With the exception of these ambiguities, and the exclusion of a small frequency region around 1500 Hz and an angular region off to the side of the head (where the just-noticeable-difference in angle, the so-called "minimum audible angle," may be as large as 20°), interaural processing leads to minimum audible angles in the horizontal plane (i.e., in azimuth) of roughly 1–5°.

The natural human direction-finding system uses two methods for resolving the above-mentioned ambiguities: (1) head movements and (2) monaural processing. Thus, for example, front–back ambiguities can be eliminated by rotating the head in azimuth: if the source is in front, a rotation to the right will cause the signal in the left ear to lead in time and to be louder; if the source is in back, the opposite will occur. Directional information from monaural processing is achieved by using a priori information on the transmitted signal $X(\omega)$ to estimate certain properties of $S_L(\omega, \theta, \phi)$ from the received signal $\Upsilon_L(\omega, \theta, \phi)$ [or $S_R(\omega, \theta, \phi)$ from $\Upsilon_R(\omega, \theta, \phi)$]. To obtain useful information in this manner, however, it is necessary that the transmitted spectrum $X(\omega)$ contain energy above 5 kHz [where the structure of the pinna causes significant effects in $\Upsilon_L(\omega, \theta, \phi)$ and $\Upsilon_R(\omega, \theta, \phi)$]. Although there are many conditions under which monaural localization is useless and, under the best of conditions, it enables listeners to determine elevation in the median plane with an accuracy of only about 20° (assuming the head is held fixed), it nevertheless provides an important adjunct to interaural localization. Particularly noteworthy is the performance achieved in discriminating azimuth in the horizontal plane by certain individuals who are unilaterally deaf (Hausler, Colburn, & Marr, 1983).

The existence of reflected energy (echos and reverberation), which occurs in most acoustic environments, tends to degrade direction finding. However, this degradation is limited by the tendency of the auditory system to enhance perception of the direct acoustic wave and suppress the later arriving echos (the "precedence effect"). Of all the phenomena associated with normal human direction finding, the precedence effect is perhaps the least understood and most interesting as a research topic (e.g., see Zurek, 1987; Shinn-Cunningham, Durlach, & Zurek, 1993).[5]

Finally, it should be noted that reflected energy also plays a role in the perception of sound–source distance. Some information on distance is obtained from loudness by exploiting a priori information about the intensity of the source or by exploiting the change in loudness with motion (Ashmead, LeRoy, and Odom, 1990). Information on distance can also be obtained from spectral shape (because high frequencies tend to be attenuated with distance more than low frequencies) by exploiting a priori information about the spectrum of the source (Coleman, 1968; Little, Mershon, and Cox, 1992). Additional important information, however, is obtained by estimating the ratio of reflected to direct energy. In a nonanechoic environment, this ratio tends to increase as the source moves further away. Since both of these methods of estimating distance are relatively crude, however, our ability to discriminate distance, like our ability to discriminate elevation in the median plane, is relatively poor.

Further, more detailed discussions of many of the above-mentioned issues can be found in Blauert (1983) and Yost and Gourevitch (1987).

Although the ability to *discriminate* the location of a single source constitutes an important measure of localization performance, it is certainly not the only measure. A further measure, that is at least as important, concerns the ability to *identify* source location. Whereas discrimination performance is limited solely by basic auditory sensitivity, identification performance is limited also by inadequacies of short-term auditory memory. In particular, to the extent that the perceptual space in which the identification is taking place is subjectively unidimensional, the best information transfer that can be achieved is restricted to roughly 2 bits (e.g., Miller, 1956). An extensive series of studies concerning this limitation, which includes development and evaluation of a general quantitative model for discrimination and identification (that includes both memory and sensory factors), as well as substantial empirical data for the special case of auditory intensity, is available in the series of papers by Durlach, Braida, and their associates (e.g., see Durlach & Braida, 1969; Braida, Lim, Berliner, Durlach, Rabinowitz, & Purks, 1984). Application of this model to localization and to the identification of interaural time delay and interaural amplitude ratio can be found in Searle, Braida, Davis, and Colburn (1976) and in Koehnke and Durlach (1989). Based on these studies, as well as other studies in which location identification data are reported (e.g., see the references in Searle et al., 1976), we estimate that the total information transfer achievable with a fixed head in anechoic localization space is only about 3 bits if the transmitted signal is randomized (and no greater than 6 bits if the transmitted signal is held fixed).[6]

Further issues arise when the environment contains a number of simultaneous independent sources. One such issue concerns the ability to detect a given source in a background of interference emanating from sources at other locations. Substantial research has been conducted on this masking problem, particularly on the advantages of binaural hearing ("binaural unmasking"), and particularly for cases in which the signals are presented through earphones to achieve improved stimulus control. Extensive summaries of both theory and data in this area are available in Durlach and Colburn (1978) and Colburn and Durlach (1978).

Finally, it should be noted that compared to binaural unmasking, relatively little attention has been given to localization in multiple target situations. Even if a given target source can be reliably detected, it may not be well localized in the presence of other (background) sources. Fortunately, current work by Yost and his associates at Loyola University (personal communication) may soon provide data on this problem of masked localization. (Some results on masked discrimination of interaural

5. As Bob Berkovitz (of Sensimetrics, Inc.) has pointed out to the authors, the idea that most acoustic environments involve echoes and reverberation is valid only for modern environments. In the good old days (evolutionarily speaking), and discounting cave-dwelling groups, most environments were probably quite anechoic. Thus, the notion that any properties of the auditory system, such as those relating to the precedence effect, can be thought of in terms of evolutionary need to perform well in reverberant environments, is probably nonsense.

6. If the transmitted signal is randomized, no substantial information can be transmitted via variation in elevation or distance and there is no indication that more than 3 bits can be transmitted via azimuth. If the transmitted signal is held fixed, then no more than roughly 3 bits should be obtainable from the distance and elevation variables combined, even if 2 bits were obtained from each variable alone (e.g., Durlach, Tan, MacMillan, Rabinowitz, & Braida, 1989).

differences are available in Ito, Colburn, and Thompson, 1982.)

3 Localization Displays

Consider now the problem of displaying information to the human operator under the assumption that all the acoustical sources are resolved at the input to the display. This beneficial condition can arise in a teleoperator system as a consequence of employing a high-resolution sensory system and in a virtual environment system by construction. In either case, the problem we wish to consider now is how best to present the various sources in such a way that their resolution is usefully maintained. In particular, the display should enable the operator to scan the whole array of sources for purposes of monitoring and also to focus on any individual source for careful listening at will.

In principle, the display can be presented to (1) the monaural auditory system, (2) the binaural auditory system, (3) modalities other than hearing, or (4) a team of listeners. Although there are many interesting questions associated with paths (1), (3), and (4), in this paper we focus attention on displays to the binaural auditory system. [Discussion of path (1), in the context of multimicrophone hearing aids for unilaterally deaf subjects, can be found in Durlach, Corbett, McConnell, Rabinowitz, Peterson, & Zurek (1987).]

We also assume that the display is a localization display. In the present context, a "localization display" is one in which the values of certain parameters used to differentiate the sources covary in an orderly manner with the positions/movements of the listener. In other words, it is a display that permits the listener to derive spatial attributes of the source by moving his/her ears through the space.

A localization display is "natural" to the extent that the display parameters that covary with the listener's movements are the same as those that occur in the real world (as described in Section 2) and thus provide spatial attributes to the sources without any learning by the listener. Essentially all of the work on localization displays reported to date in the literature has been directed toward such natural displays (e.g., see the review by Wenzel, 1992). In other words, the focus has been on reproducing with earphone listening the localization cues normally encountered in field listening. This involves filtering the signals presented to the two ears with space filters $S_L(\omega, p)$ and $S_L(\omega, p)$ that are equal to the HRTFs that would be measured at the listener's eardrums for a source located at point P (and varying these filters appropriately in real time as the source and/or ears are moved).

Ideally, a supernormal localization display should improve spatial resolution without incurring increased response bias.[7] However, since any alteration of normal localization cues will, of necessity, introduce a response bias when first encountered, the best that one can hope for with respect to response bias is that the alteration is one to which the subject can easily adapt. Unfortunately, our understanding of sensorimotor adaptation (e.g., see the review by Welch, 1978) is not yet sufficiently adequate to enable one to predict ease of adaptation for arbitrary alterations. The construction of a quantitative model of adaptation that would facilitate such predictions is thus a primary research goal in this area. It should also be noted that practical applications of such super localization displays will require that the operator be able to shift back and forth between these displays and the normal environment on a more or less routine basis. Inasmuch as significant response bias in either environment would be detrimental, it is necessary that the subject be able to adapt back and forth between the two states quickly and easily. Recent investigations of conditional adaptation (by Welch and associates at NASA Ames, personal communication), as well as common experience in the visual domain with the alterations imposed by eyeglasses and microscopes, suggest that such adaptation behavior is realizable.

In addition to the alteration being of such a nature that response bias can be effectively eliminated by sensorimotor adaptation, it is important that the alteration not seriously reduce the extent to which the relevant nonspatial information carried by the source can be per-

7. Whereas spatial resolution refers to the extent to which different physical positions can be distinguished perceptually, response bias refers to the extent to which physical position and perceptual position deviate.

ceived. In other words, the altered coding of position should not seriously interfere with reception of the nonspatial message emitted by the source. Thus, for example, if the source were emitting a speech message, no alteration of localization cues should be used that makes the speech unintelligible. (In the area of hearing aids for monaural listeners, the task of improving spatial resolution without degrading speech intelligibility has proved to be a significant problem.) Keeping both the response bias problem and the source intelligibility problem in mind, we now turn our attention to alterations of localization cues that will improve spatial resolution. Initially, we consider resolution at the simplest level: the just-noticeable difference (JND) in each spatial parameter (azimuth, elevation, and distance) for a single, isolated, source. Subsequently, we consider problems related to the absolute identification of spatial location (a task in which, as discussed in Section 2, performance is likely to be limited by inadequacies of short-term memory as well as of basic sensitivity) and to interference among multiple, simultaneous, targets.

4 Improving the JND in Direction and Distance

4.1 Direction

4.1.1 Enlarging the Head. Perhaps the most obvious approach to increasing directional resolution is to simulate the results that would be obtained in field listening with a larger head. As pointed out by Clifton, Clarkson, Gwiazda, Bauer, and Held (1988), adaptation to small changes in head size is already required in connection with natural development in infants.

One method of achieving such a simulation is by frequency scaling the space filters. In other words, define a new pair of space filters $S'_L(\omega, \theta, \phi)$ and $S'_R(\omega, \theta, \phi)$ by the relations

$$\begin{aligned} S'_L(\omega, \theta, \phi) &= S_L(K\omega, \theta, \phi) \\ S'_R(\omega, \theta, \phi) &= S_R(K\omega, \theta, \phi) \end{aligned} \qquad (3)$$

where S_L and S_R are the normal HRTFs and K exceeds unity. [Note that we are scaling only the space filters here; the transmitted spectrum $X(\omega)$ is, of course, not scaled.] Since the physical interactions of the incoming acoustic wave with the body, head, and pinna of the listener tend to remain invariant if both the wavelength and the body, head, and pinna of the listener are scaled in size by the same amount,[8] increasing the frequency by the factor K is equivalent to increasing the size of the body, head, and pinna by the factor K. Such a transformation has the advantage of both magnifying interaural differences to help binaural direction finding and magnifying monaural spectral effects to enhance monaural direction finding. And it does both of these in a coherent and unified manner. The most obvious disadvantage of such a scheme is that it leaves the new space filters S'_L and S'_R undefined at the higher frequencies. For example, if $K = 2$ and the normal filters are known up to 20 kHz, the modified filters are only defined up to 10 kHz. Thus, one would either have to ignore the higher frequencies or find some method for filling them in.

Another possible method of simulating expanded size is to exploit previous work on physical modeling of head and pinnae (Genuit, 1984). To the extent that an adequate model exists for relating the stimuli at the ear drums to geometric parameters of the physical structure, one can compute the physical stimuli that would occur when the size of the head and pinnae are increased.

4.1.2 Faking an Enlarged Head. Transformations that do not truly simulate an enlarged head in the sense achieved by frequency scaling, but nevertheless capture certain aspects of head enlargement and may be quite useful for special purposes, have previously been discussed by Durlach and Pang (1986) and Van Veen and Jenison (1991).

Durlach and Pang (1986) considered transforming the received signals $\Upsilon_L(\omega, \theta, \phi)$ and $\Upsilon_R(\omega, \theta, \phi)$ into

8. Strictly speaking, this statement is true only if the distance between the head and source is also scaled by the same amount. A paper describing the conditions under which this distance scaling can be ignored is presented in later pages of this issue of *Presence* (Rabinowitz, Maxwell, Shao, & Wei, 1993).

new signals $\Upsilon'_L(\omega, \theta, \phi)$ and $\Upsilon'_R(\omega, \theta, \phi)$ by exponentiation:

$$\begin{aligned}\Upsilon'_L(\omega, \theta, \phi) &= [\Upsilon_L(\omega, \theta, \phi)]^n \\ \Upsilon'_R(\omega, \theta, \phi) &= [\Upsilon_R(\omega, \theta, \phi)]^n\end{aligned} \tag{4}$$

Making use of Eq. (1), we see that

$$\begin{aligned}\Upsilon'_L(\omega, \theta, \phi) &= [X(\omega)]^n[S_L(\omega, \theta, \phi)]^n \\ \Upsilon'_R(\omega, \theta, \phi) &= [X(\omega)]^n[S_R(\omega, \theta, \phi)]^n\end{aligned} \tag{5}$$

and

$$\frac{\Upsilon'_L(\omega, \theta, \phi)}{\Upsilon'_R(\omega, \theta, \phi)} = \left[\frac{S_L(\omega, \theta, \phi)}{S_R(\omega, \theta, \phi)}\right]^n \tag{6}$$

where $S_L(\omega, \theta, \phi)$ and $S_R(\omega, \theta, \phi)$ are the relevant space filters. For $n > 1$, this transformation magnifies the interaural amplitude and phase difference at all frequencies ω, thus improving binaural localization. Also, of course, it magnifies the dependence of both the left and right space filters on ω, thus improving monaural as well as binaural localization. Although Durlach and Pang introduced this transformation as a possible means for improving the localization capabilities of humans with impaired hearing (specifically, a degraded ability to discriminate interaural differences), the idea is equally applicable to teleoperator systems with anthropomorphic telerobots, provided the functions Υ_L and Υ_R are assumed to be the received signals at the ears of the telerobot and the functions Υ'_L and Υ'_R are assumed to be the signals displayed to the human operator. If the telerobot is the same size as the human, then the transformation will provide directional cues that surpass those normally encountered. On the other hand, if the telerobot is very small (e.g., to permit exploration of small spaces), then the transformation can be used to magnify the reduced directional cues associated with a small head and thereby return these cues to normal. Further discussion of these exponentiation transformations can be found in Durlach and Pang (1986). Experimental results on benefits in speech intelligibility in a background of noise for listeners with impaired hearing achieved by use of this transformation can be found in Kollmeier and Peissig (1989).

Van Veen and Jenison (1991), concerned about the nonlinearity of the exponentiation transformation and the resultant complications that such a nonlinearity would introduce when more than one acoustic source contributes to the received signals Υ_L and Υ_R, suggested a linear alternative to magnification by exponentiation. Their processing scheme assumed two inputs $X_L(\omega)$ and $X_R(\omega)$, two outputs $\Upsilon_L(\omega)$ and $\Upsilon_R(\omega)$, and four linear filters $H_{LL}(\omega)$, $H_{RR}(\omega)$, $H_{LR}(\omega)$, and $H_{RL}(\omega)$ arranged in the following manner:

$$\begin{aligned}\Upsilon_L(\omega) &= X_L(\omega)H_{LL}(\omega) + X_R(\omega)H_{RL}(\omega) \\ \Upsilon_R(\omega) &= X_L(\omega)H_{LR}(\omega) + X_R(\omega)H_{RR}(\omega)\end{aligned} \tag{7}$$

The filters were then chosen to minimize the average squared error between a user-specified space warping function and the actual space warping determined by the filters. In this paper, attention was confined to manipulating only interaural time delay (as opposed to interaural amplitude difference or both interaural time delay and interaural amplitude difference) and to frequencies below 1000 Hz. The results clearly indicated that, at least under the assumed conditions, magnification using linear filters was feasible.

4.1.3 Remapping the Normal Space Filters. A very different approach to those described in the preceding two sections for altering directional resolution involves retention of the same (natural) set of space filters, but introduction of a new mapping between these filters and the directions (θ, ϕ). Using this approach, one defines a transformation

$$\theta' = f(\theta, \phi)$$

$$\phi' = g(\theta, \phi)$$

and then defines

$$\begin{aligned}S'_L(\omega, \theta, \phi) &= S_L(\omega, \theta', \phi') = S_L[\omega, f(\theta, \phi), g(\theta, \phi)] \\ S'_R(\omega, \theta, \phi) &= S_R(\omega, \theta', \phi') = S_R[\omega, f(\theta, \phi), g(\theta, \phi)]\end{aligned} \tag{8}$$

In such a transformation, no new space filters are created; instead, the old space filters are reassigned to different angles. Moreover, they are assigned consistently with respect to the L–R variable so that not only are the same set of space filters used for each ear, but the same set of interaural ratios is preserved. In general, the use of such a remapping transformation will increase resolution in some regions of (θ, ϕ) space and decrease it in others.

As an illustration of this general class of transformations, confine attention to the horizontal plane (i.e., assume $\phi = \phi' = 0$) and let

$$\begin{aligned} S_L'(\omega, \theta, 0) &= S_L(\omega, \theta', 0) \\ S_R'(\omega, \theta, 0) &= S_R(\omega, \theta', 0), \end{aligned} \tag{9}$$

where

$$\theta' = \frac{1}{2}\tan^{-1}\left[\frac{2n\sin(2\theta)}{1 - n^2 + (1 + n^2)\cos(2\theta)}\right]$$

Pictures of this transformation[9] are shown in Figure 6 for the cases $n = 1, 2, 3$, and ½. For $n > 1$ this transformation increases resolution in the neighborhood of $\theta = 0°$ and decreases it in the neighborhood of $\theta = +90°$ and $-90°$. For $n < 1$, the opposite occurs. Transformations of this type may present less of a challenge to sensorimotor adaptation mechanisms than those considered in the previous two sections because no new auditory stimuli are introduced in these remapping transformations. Experimental results obtained with this transformation will be presented in the next paper of this series.

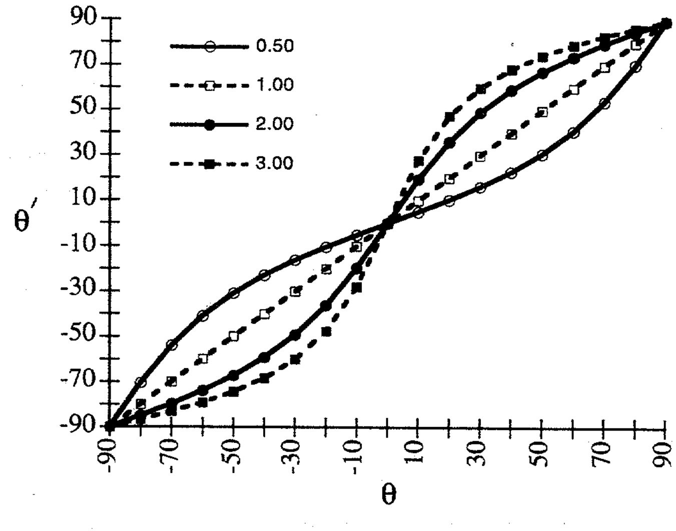

Figure 6. *A plot of the transformation specified by Eqs. 9.*

4.2 Distance

Despite the use of the term "3-D" in essentially all descriptions of current auditory spatialization systems designed to simulate normal spatial hearing, none of these systems is actually a 3-D system: the filters used to synthesize spatial cues depend only on two dimensions, azimuth and elevation. Although it is possible to scale the overall intensity with virtual distance, such scaling is not very effective (in part, because changes in intensity can be interpreted as changes in source intensity rather than in source distance). In general, the inadequacy of these systems with respect to distance is clearly demonstrated by the fact that the auditory images these systems create often appear to be located inside the head (the degenerate limit of distance–cue failure) and the actual spatialization achieved is limited to "in-head spatialization."[10] A detailed discussion of the factors influencing the externalization of auditory images can be found in Durlach, Rigopulos Pang, Woods, Kulkarni, Colburn, and Wenzel (1992). A comprehensive list of references on distance perception can be found in Little et al. (1992) and Brungart (1993).

As indicated in Section 2, the physical cues underlying distance perception in a natural environment include sound intensity, high-frequency-to-low-frequency energy ratio, and direct-to-reverberant energy ratio. Also, at very small distances, the wavefront impinging on the head will not be planar and some information should be extractable from the curvature of the wavefront. To date, however, and to the best of our knowledge, none of the previous efforts to measure HRTFs in anechoic space has attempted to measure how these filters change with distance at distances close to the listener's head.

For the relatively simple case of distance coding in a virtual environment in which the reflective properties of the space are completely free to be used for artificial distance coding (rather than being used to achieve a certain general acoustic ambience or to simulate a specific struc-

9. We are indebted to Xiao Dong Pang for designing this transformation.

10. Even when the image appears inside the head, it can be localized to a certain degree. Most of this localization, however, takes place along a line joining the two ears (so-called "in-head lateralization). The degree to which the locations of in-head images vary along the up–down or front–back dimensions is relatively minor compared to variations in lateralization.

ture, such as an existing concert hall), we have previously suggested a class of coding schemes based on an idealized single-echo structure (Durlach, 1991). In these schemes, the space filters $S_{A,m,\tau}(\omega)$ are given by

$$S_{A,m,\tau}(\omega) = A[1 + me^{-j\omega\tau}] \tag{10}$$

where A gives the overall amplitude, m the strength of the echo relative to the direct sound, and τ the time delay between the direct sound and the echo (the phase shift in the reflection process is assumed to be zero). The corresponding amplitude and phase characteristics of this "comb filter" are given by

$$\begin{aligned} |S_{A,m,\tau}(\omega)|^2 &= A^2[1 + m^2 + 2m\cos(\omega\tau)] \\ \text{phase } [S_{A,m,\tau}(\omega)] &= \tan^{-1}\left[\frac{-m\sin(\omega\tau)}{1 + m\cos(\omega\tau)}\right] \end{aligned} \tag{11}$$

Assuming for simplicity that $m = 1$ (i.e., the reflection coefficient is unity), one obtains the simplified characteristics

$$\begin{aligned} |S_{A,1,\tau}(\omega)|^2 &= 4A^2\cos^2\left(\frac{\omega\tau}{2}\right) \\ \text{phase } [S_{A,1,\tau}(\omega)] &= \frac{\omega\tau}{2} \end{aligned} \tag{12}$$

Assuming further, again for simplicity, that the virtual sound source and the virtual sensor (the virtual ear) are the same distance h above (or below or beside) the reflecting surface, and that the distance D between source and sensor is large with respect to h, $D \gg h$, one then obtains the approximation

$$\tau = \frac{h^2}{VD} \tag{13}$$

and the filter characteristics can be rewritten as

$$\begin{aligned} |S_{A,1,D}(\omega)|^2 &= 4A^2\cos^2\left(\frac{\omega h^2}{VD}\right) \\ \text{phase } [S_{A,1,D}(\omega)] &= \frac{\omega h^2}{VD} \end{aligned} \tag{14}$$

where V is the velocity of sound. In this sort of distance coding, the information on distance is presented in terms of spectral modulation in a manner similar to that which would occur in a space with one specular reflection. The parameter h can be chosen so that the range of modulation frequencies encountered as the distance D is varied is optimum for the purpose of discriminating between different values of D.

Further encoding of the source–sensor distance D can be achieved of course by employing redundant, multidimensional encoding and specifying relationships between D and the overall amplitude A and the relative echo strength m as well as between D and the echo delay τ. Such redundant encoding will increase the resolution in D, but will interfere with the use of these parameters for encoding other types of information. If one wants to use the parameter A in a natural manner, then a relation of the form $A = K/D$ (K a constant) and the corresponding amplitude characteristic

$$|S_D(\omega)|^2 = \left(\frac{4K^2}{D^2}\right)\cos^2\left(\frac{\omega h^2}{VD}\right) \tag{15}$$

would make sense because it corresponds to the natural attenuation that occurs with distance in free space. Inasmuch as the magnitude of reflection coefficients tends to decrease with an increase in the angle of incidence, natural encoding of m would require that m increase monotonically with an increase in D.

In general, in considering use of the comb filters specified by Eqs. (11) for spatialization, two further points should be noted. First, inasmuch as (1) elevation, as well as distance, is normally perceived very poorly, and (2) these filters have a natural relationship to elevation as well as distance, these filters could also be usefully employed to improve resolution in elevation. In other words, rather than using m and τ to code distance redundantly, these parameters could be used to code both distance and elevation (or even to code elevation redundantly, where A is used to code distance). Since filters of this type often occur naturally as a consequence of the reflection process in natural environments and do not usually interfere too seriously with understanding the nonspatial information in the acoustic signal, it seems likely that they could be used in virtual environments without causing too much interference or requiring too much learning time.

Second, in the recent distance estimation study by Brungart (1993), in which it was found that distance estimation was not much improved by the inclusion of single-echo reflection information, transmitted intensity was always held fixed and the attenuation factor A was always used as a coding variable. No effort was made to study the effectiveness of coding by the use of τ and/or m under conditions in which the transmitted intensity was randomly varied. Also, no attempt was made to study discrimination or identification performance, or to train the subjects through the use of correct-answer feedback. Thus, at present, we have little empirical information about the information transfer that could be obtained with the single-echo filter set when the factor A is not used for coding, or with all three variables A, τ, and m when feedback training is employed.

Finally, it should be noted that the coding of distance and/or elevation by use of these comb filters can be combined with the coding of distance by use of filters that smoothly and progressively attenuate the high frequencies or with the coding of elevation by the insertion of notches or peaks in the spectrum at high frequencies (to exaggerate the normal notches and peaks at these frequencies). As with the comb filters, however, the detailed filtering schemes must be chosen so as not to interfere too strongly with the reception of non-spatial information in the signal.

5 Beyond the JND

In previous sections, our discussion of the various localization transformations was restricted in two fundamental ways. First, attention was focused on the issue of resolution rather than response bias. Transformations were introduced to increase resolution without regard for how they would affect response bias and without regard for the extent to which this bias could be overcome by sensorimotor adaptation. Second, within the domain of resolution, attention was focused on improving the JND. No consideration was given to the constraints on resolution associated with limits on information transfer in absolute identification experiments resulting from inadequate short-term memory for sensations (e.g., Pollack, 1956, 1961; Garner, 1962; Miller, 1956; Durlach & Braida, 1969; Braida, Lim, Berliner, Durlach, Rabinowitz, & Purks, 1984). Similarly, except for the work cited by Kollmeier and Peissig (1989) on the Durlach and Pang (1986) interaural-magnification transformation, no attention was given to the problem of interaction among multiple simultaneous targets. In the following subsections, we consider briefly all of these issues.

5.1 Discrimination vs. Identification

To achieve good resolution in any space of variables, it is necessary that the number of JNDs in the space be large. Obviously, if one cannot discriminate between two elements in the space, it will not be possible to respond differentially to these elements in identification paradigms (i.e., paradigms in which the subject is required to identify which of *N* stimuli is presented on each trial). However, as is well known, and as is discussed extensively in the references just cited, the condition of a large number of JNDs in the space is not sufficient to ensure that good identification performance can be achieved. To achieve such performance, in addition to containing a large number of JNDs, the perceptual space must have characteristics that enable the subject to combat deficiencies in the human memorial system (e.g., as discussed in the above-mentioned references, the space must have appropriate "perceptual anchors" or be "multidimensional" or be highly susceptible to "chunking").

In principle, it is perfectly possible to develop a transformation of the auditory space in question that increases the number of JNDs in the space (by decreasing the size of the JNDs) but fails to increase the ability of the human listener to identify various elements of the space. This phenomenon can be easily illustrated by considering past results on the discrimination and identification of auditory intensity. Suppose, for example, one is presented with a space of signals consisting of a given fixed waveform (say a "click") and an intensity variable that covers the range 20–52 dB SPL. If one performs no transformation on this space, the number of JNDs in the space will be roughly 50. Now suppose one transforms this space by stretching the intensity variable to cover the range 20–100 dB SPL. This transformation will roughly triple the number of JNDs in the space. However, identification performance will remain invariant. For example, the identification performance achieved with the stimulus set consisting of the nine intensities 20, 30, 40, 50, 60, 70, 80, 90, 100 db will be no better than that achieved with the nine intensities 20, 24, 28, 32, 36, 40, 44, 48, 52. Detailed experiments, analyses, and theoretical models concerned with this phenomenon (as well as other phenomena related to memory limitations) can be found in the extensive series of papers by Durlach and Braida and their associates cited previously.

Returning now to the context of supernormal auditory localization, one must consider the extent to which the types of transformations considered above will improve resolution not only in pairwise discrimination experiments but also in multiple-stimulus identification. Although such transformations may prove useful even if their advantages are not maintained when going from discrimination to identification, the potential is much greater if these advantages are maintained. Although one cannot adequately address this issue with the data currently available, based on the work by Koehnke and Durlach (1989), as well as by Searle et al. (1976), it appears that one must proceed cautiously here. In other words, in the language of Pollack/Garner/Miller, it may be very important to incorporate multidimensional stimulus sets in the encoding process.

5.2 Response Bias and Adaptation

In general, when auditory localization cues are altered, both the listener's resolution and the listener's response bias will be affected. Thus, for example, returning to the cases considered previously in Section 4.1 in which interaural localization cues are magnified in the region about the median plane (i.e., an enlarged head is simulated), it is obvious that sources that are slightly off center will appear more off center than if the magnification had not been inserted. This will not only lead to an improved ability to discriminate between source directions close to the median plane, but it will cause the listener to mistakenly perceive sources that are off the median plane to be further off the median plane then they actually are. Accordingly, if a source is initially located on the median plane, and the listener turns his/her head slightly to the right (left), the source will appear to move to the left (right). Although in many applications it may be possible to correct for these biases by incorporating a compensating network in the nonhuman portion of the system, superior results could undoubtedly be obtained if these biases could be corrected within the human organism itself by means of sensorimotor adaptation in the sensorimotor loops involving auditory localization sensations, motor control of head position, and proprioceptive/kinesthetic sensations related to head position.

Unfortunately, although a substantial amount of previous research has been conducted on sensorimotor adaptation (e.g., see the review by Welch, 1978), this research has not yet led to an adequate theory of sensorimotor adaptation. Not only is it currently impossible to predict how the rate and asymptote of adaptation depend on the choice of transformation and training procedure, but there are no studies available in which both resolution and response bias are tracked over time together. In other words, those investigators who have been interested in resolution (mainly communication engineers) have tended to ignore response bias, and those investigators interested in response bias (mainly cognitive scientists) have tended to ignore resolution. Obviously, to develop theories that will facilitate the design of effective human–machine interfaces for teleoperator and virtual-environment systems, it is necessary to study both resolution and bias, and to study how both of these measures of performance change with time and with exposure experience.

5.3 Multiple Simultaneous Targets

In addition to the issues discussed in Secs. 5.1 and 5.2 above, a whole set of further issues arise in the case of multiple simultaneous targets. In particular, the occurrence of multiple targets is likely to reduce the listener's ability to detect the presence of one or more of the component targets. Furthermore, even if a given target in the complex is detectable, the presence of the other (background) targets may limit one's ability to localize it properly.

For the case of the normal binaural auditory system (i.e., when no transformation is inserted), substantial previous research has been directed towards the characterization and understanding of the detection problem. Of particular relevance here is the wealth of data and theory on the ability of the binaural auditory system to detect targets located at one position in the presence of "jammers" located at other positions (see, for example, the reviews by Durlach and Colburn, 1978, and Colburn and Durlach, 1978, on the phenomenon of binaural unmasking). In contrast, relatively little quantitative research has been done on the effect such interfering background signals have on target localization. For most applications, the ability of a transformation to suppress the deleterious effects of background interference on both detection and localization will be of crucial importance in evaluating the usefulness of the transformation. Although transformations that improve the JND and the identification of single-source location are likely to improve suppression of background interference, such results are certainly not guaranteed. In any case, the magnitude of the suppression under various stimulus conditions cannot be predicted solely on the basis of single target discrimination and identification performance.

6 Concluding Remarks

In the preceding, we have introduced the idea of supernormal auditory localization for virtual-environment and teleoperator systems, surveyed some transformations for creating improved (supernormal) localization performance, and pointed out some issues and problems that are related to the realization of this improved performance.

Needed future work in this area includes (1) the development of appropriate facilities for constructing the specified transformations of localization cues, (2) studying the perceptual effects of these alterations and how these perceptual effects change over time and with exposure experience, and (3) determining the practical benefits of these alterations in various application contexts.

Finally, it should be noted that most of the ideas discussed in this paper are equally relevant to other sense modalities. Thus, for example, just as increased separation of the ears can improve discrimination of azimuth in the auditory sense, increased separation of the eyes can improve discrimination of depth in the visual sense.

More detailed discussions of many of the ideas, issues, and problems mentioned in this paper will be presented in later papers of this series.

References

Ashmead, D. H., LeRoy, D., & Odom, R. D. (1990). Perception of the relative distances of nearby sound sources. *Perception and Psychophysics, 47,* 326.

Blauert, J. (1983). *Spatial hearing.* Cambridge, MA: MIT Press.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception XIII. Perceptual anchor model of context coding. *Journal of the Acoustic Society of America, 76,* 722–781.

Bregman, A. S. (1990). *Auditory scene analyses: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Brungart, D. S. (1993). Distance simulation in virtual auditory displays. Unpublished paper, Wright State University.

Clifton, R. K., Clarkson, M. G., Gwiazda, J., Bauer, J. A., & Held, R. M. (1988). Growth in head size during infancy: Implications for sound localization. *Developmental Psychology, 24,* 477–483.

Colburn, H. C., & Durlach, N. I. (1978). Models of binaural interaction. In Carterette & Friedman (Eds.), *Handbook of perception* (Vol. IV, pp. 467–515).

Colburn, H. S., Zurek, P. M., & Durlach, N. I. (1987). Binaural directional hearing—Impairments and aids. In W. A. Yost & G. Gourevitch (Eds.), *Directional hearing.* New York: Springer-Verlag.

Coleman, P. D. (1968). Dual role of frequency spectrum in determination of auditory distance. *Journal of the Acoustical Society of America, 44,* 631–632.

Durlach, N. I. (1991). Auditory localization in teleoperator and virtual-environment systems: Ideas, issues, and problems. *Perception, 20,* 543–554.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustic Society of America, 46,* 372–383.

Durlach, N. I., & Colburn, H. C. (1978). Binaural phenomena. In Carterette & Friedman (Eds.), *Handbook of Perception* (Vol. IV, pp. 365–466).

Durlach, N. I., Corbett, C. R., McConnell, M. V. L., Rabinowitz, W. M., Peterson, P. M., & Zurek, P. M. (1987). Multimicrophone monaural hearing aids. *Proceedings of RESNA 10th Annual Conference* (pp. 395–397). San Jose, CA.

Durlach, N. I., & Pang, X. D. (1986). Interaural magnification. *Journal of the Acoustic Society of America, 80,* 1849–1850.

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,* 251–257.

Durlach, N. I., Tan, H. Z., MacMillan, N. A., Rabinowitz, W. M., & Braida, L. D. (1989). Resolution in one dimension with random variations in background dimensions. *Perception and Psychophysics, 46,* 293–296.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* New York: John Wiley.

Genuit, K. (1984). *Ein Modell zur Beschreibung von Aussenohrubertragungseigenschaften.* Dr.-Ing. Thesis, Aachen, Germany.

Hausler, R., Colburn, H. S., & Marr, E. (1983). Sound localization in subjects with impaired hearing. *Octa Oto-Laryngologica, Supplementum 400.*

Ito, Y., Colburn, H. S., & Thompson, C. L. (1982). Masked discrimination of interaural time delays with narrow-band signal. *Journal of the Acoustical Society of America, 72,* 1821–1826.

Koehnke, J., & Durlach, N. I. (1989). Range effects in the identification of lateral position. *Journal of the Acoustic Society of America, 86,* 1176–1178.

Kollmeier, B., & Peissig, J. (1989). Speech intelligibility enhancement by interaural magnification. *Acta Oto-Laryngologica, Supplement 469,* 215–223.

Little, A. D., Mershon, D. H., & Cox, P. H. (1992). Spectral content as a cue to perceived auditory distance. *Perception, 21,* 405–416.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Pollack, I. (1956). Identification and discrimination of components of elementary auditory displays. *Journal of the Acoustic Society of America, 28,* 906–909.

Pollack, I. (1961). Selected developments in psychophysics with implications for sensory organization. In W. A. Rosenblith (Ed.), *Sensory communication.* Cambridge, MA: MIT Press.

Rabinowitz, W. M., Maxwell, J., Shao, Y., & Wei, M. (1993). Sound localization cues from a magnified head: Implications from sound diffraction around a rigid sphere. *Presence,* in press.

Robinett, W. (1993). Presentation at National Academy of Sciences meeting on Virtual Environments and Teleoperators, February 1993.

Searle, C. L., Braida, L. D., Davis, M. S., & Colburn, H. S. (1976). Model for auditory localization. *Journal of the Acoustic Society of America, 60,* 1164–1175.

Sheridan, T. B. (1992). Defining our terms. *Presence, 1,* 272–274.

Shinn-Cunningham, B. G., Zurek, P. M., & Durlach, N. I. (1993). Adjustment and discrimination measurements of the precedence effect. *Journal of the Acoustic Society of America,* in press.

Van Veen, V. D., & Jenison, R. L. (1991). Auditory space expansion by linear filtering. *Journal of the Acoustic Society of America, 90,* 231–240.

Welch, R. B. (1978). *Perceptual modification: Adapting to altered sensory environments.* New York: Academic Press.

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

Yost, W. A. (1991). Auditory image perception and analysis: The basis for hearing. *Hearing Research, 56,* 8–18.

Yost, W. A., & Gourevitch, G. (Eds). (1987). *Directional hearing.* New York: Springer-Verlag.

Zurek, P. M. (1987). The precedence effect. In W. A. Yost & G. Gourevitch (Eds.), *Directional Hearing.* New York: Springer-Verlag. –