Defining and Redefining Limits on Human Performance in Auditory Spatial Displays

Barbara Shinn-Cunningham and Nathaniel Durlach
Research Laboratory of Electronics, MIT
77 Massachusetts Ave.
Cambridge, MA 02139
Email: shinn@cbgrle.mit.edu

Abstract

Design of an effective auditory display depends upon understanding the requirements of the task as well as understanding limitations imposed by available display technology. However, an effective display also must consider limitations that are imposed by the human.

The current chapter examines the use of spatial cues in an auditory display. First, normal perceptual limitations on spatial perception are discussed, with emphasis on some nonobvious ways in which perceptual limitations can affect the amount of information received by a human operator. An approach to overcoming such perceptual limitations is then introduced and its effectiveness examined. Although the focus of the chapter is on limitations in the perception of spatial auditory cues, the approach outlined here can be utilized in other display dimensions.

1 Introduction

The main goal in the development of new auditory displays is to increase the amount of information reaching a human operator. A number of guidelines derive from this basic goal. For instance, in order to maximize the information received by a human operator, cues used to encode information in one acoustic dimension should be chosen carefully so that information in other acoustic cues is not lost or distorted. As an example, spatial auditory cues can be used in conjunction with the pitch of a complex stimulus to encode different aspects of information presented; however, use of the pitch of pure sinusoids in place of more spectrally-complex stimuli would interfere with the perception of sound location, since sinusoids are often difficult to localize. Different dimensions in the information must be presented will be lost.

Increasing information transfer to the human also depends upon choosing cues that make use of natural talents of the user. For instance, using acoustic location cues to encode data which are inherently spatial allows the user access to normal spatial perceptual strategies for exploring and understanding the information presented. Information must also be mapped to cue dimensions which have the appropriate resolution and dynamic range to allow perception of relevant details in the data. In addition to these high-level considerations, there are many perceptual phenomena that affect performance in auditory localization tasks. These topics are considered in the next section.

2 Perceptual Limits in Auditory Localization

2.1 Peripheral Limits

The most basic limitations on human perception of the position of an acoustic source derive from peripheral, sensory limitations of the human. These limitations control the absolute limits on the resolution of spatial information that can be conveyed to the user, and derive from the basic structure of our hearing apparatus.

Perception of all of localization cues is limited to some degree by our sensory system. Perception of interaural time differences (ITDs—important in the perception of source azimuth) is limited by the ability of the sensory apparatus to faithfully encode timing information by phase-locking to the input stimulus. Phase-locking to individual cycles can only occur for stimuli below about 1-4 KHz; however, if high-frequency stimuli have sufficient bandwidth, cells exhibit firing patterns which are phase-locked to the envelope of the stimuli [1]. Psychophysical results show that these peripheral limitations correspond to the limits in our ability to perceive interaural phase differences [2]. Even when cells can phase-lock to a stimulus, ITD resolution is limited by the stochastic nature of auditory nerve firings.

Similarly, resolution of interaural intensity differences (IIDs—useful in the perception of source azimuth as well as elevation for source positions off the median plane) depends upon noise in the auditory nerve, since the reliability of the perceptual judgement depends upon the representation of both the left and right ear signal intensity. Perception of spectral shape (important in discerning source elevation and front-back location) depends upon the spectral resolution of the peripheral, bandpass filtering (of approximately 1/3-octave width) which is performed in the cochlea. The perception of source distance depends upon the extraction of higher-level cues (such as relative level of reverberation, changes in loudness and position with relative motion of the source, and comparison of source level and spectrum with expected level and spectrum) that do not correlate as closely with any single physical cue. However, even for source distance, the peripheral representation of the stimulus will restrict the information perceived by the listener, if in less obvious ways.

2.2 Central Processing Limits

Performance is not only limited by the peripheral representation of stimuli, but by the processing performed at more central stages of the auditory pathway. These more central mechanisms can be defined by physiological work and deduced from psychophysical experimentation. For instance, psychophysical studies of sensitivity to ITDs have shown that sensitivity is greatest for ITDs near 0 [2]. In addition, physiological exploration of the brainstem has found a site in the auditory pathway in which cells extract interaural timing differences [1]. Each cell in the Medial Superior Olive is "tuned" to a specific ITD and responds preferentially to ITDs of that value. To the extent that the distribution of ITDs to which the cells are tuned is "hard-wired," this distribution determines the dynamic range of ITDs that can be perceived as well as the resolution with which they can be determined.

Other more central processing limitations are deduced solely from psychophysical performance on various localization tasks. One aspect of auditory direction perception which is not obvious from peripheral cues is the extent to which the many possible localization cues are combined by these central mechanisms. This combination of different cue dimensions makes sense from an evolutionary viewpoint: by combining information across cue dimensions, the accuracy of source location judgements can be increased. However, the ways in which information is combined by the auditory system can prevent a listener from separately perceiving cues which are theoretically

available in the stimulus. In particular, using different localization cues to encode different data dimensions is a very poor idea, since most localization information is effectively averaged in the auditory pathway.

ITDs and IIDs are combined by the auditory system such that the perceived location of a dichotic source which has an ITD favoring the left ear and an IID favoring the right ear can be heard at the center of the head with appropriate choices of ITD and IID—so called time-intensity trading (for example, see [3]). The perceived source position may be so diffuse that it becomes ambiguous and/or breaks into two separate images, one depending most heavily on the ITD cues, and one which depends most heavily on the IID cues. However, most significantly, even under such circumstances, each of the two separate images is influenced by both the ITD and IID information in the stimulus.

Another example of the central combination of conceptually distinct acoustic location cues is demonstrated by the perception of multiple sources from different directions. Under many circumstances, localization information from multiple sources is combined and cannot be perceived independently. For instance, even simulated sources with differing frequency content may not be perceived as coming from different positions; the apparent location of the composite source depends upon a weighted average of the location information in the various frequencies [4]. If harmonically-related tones are presented, they are often fused into one event with a single location, even when tones in the complex have different ITDs [5]. This fusion is less likely to occur when tone complexes do not have a harmonic relationship. Similarly, when complex sounds or tones with different interaural information are amplitude-modulated by the same waveform or have similar onsets, they are often perceived as a fused source at a single location; when their envelopes differ, location information from the two sources is less likely to combine and the location of each of the sources is more likely to be accurately perceived [6]. The combination of location information thus is influenced by the some of the same properties that affect source segregation [7].

In the precedence effect, two sources from different directions are heard from a single position when the onsets of the two sources occur within about 10 ms (for simple sounds) to 50 ms (for complex waveforms like speech and music) of each other [8]. Under such circumstances, the apparent position of both sources is influenced strongly by the location information in the first-arriving wavefront, and only weakly by the position of the second source (i.e, location information in the second source is lost). When the onset delay between the two sounds is increased, two positions may be perceived; however, as in the previous examples, the apparent locations of both positions are influenced by location information in each of the two sources [9]. Some discussions of the precedence effect have focussed on the apparent temporal "sluggishness" of the binaural system and its difficulty in perceiving changes in interaural information over time. Another example of this sluggishness is in the perception of source motion: sensitivity to source motion is poor relative to that of the visual system (for instance, see Perrott [10]). Evidence supports the idea that such sluggishness arises from central mechanisms: for instance, the precedence effect works for sound sources which have energy in distinct spectral regions, arguing that the combination of interaural information occurs after within-critical-band processing has been performed [11].

2.3 Cognitive Limits

While sensory noise and processing mechanisms determine absolute limits for the best performance that can be achieved, on many tasks performance is far worse than would be predicted on the basis of these limits. For instance, performance on discrimination tasks in which a listener must simply discern differences between two stimuli can be predicted by taking into account peripheral noise. However, on identification tasks, where many possible stimuli are presented and the listener's task

to correctly name each stimulus, performance is far worse than would be predicted on the basis of sensory limitations (e.g., performance is worse when positions from −90 to 90 degrees must be identified by a listener than when he must simply determine whether a sound came from a position of 0 degrees or from 5 degrees [12, 13, 14]). These results indicate that, while lower limits on performance can be deduced from peripheral limitations, achievable performance on complex tasks depends strongly on higher-level functions.

2.4   Limits in Cue Salience

A final type of limitation arises from ambiguities in the physical cues themselves. Such restrictions arise commonly in the perception of source distance. Perceptually, source distance is often judged poorly, reflecting the ambiguities in the cues available to the listener. Source distance cues include the absolute level of the source, the relative energy of the source across frequencies, the change in angular direction of the source with listener movement, and the direct-to-reverberant energy ratio. Of these cues, only the latter two are unambiguous. For an unknown source, the absolute level can provide a relative distance cue but not an absolute cue [15, 16, 17]. Similarly, for distant sources, high-frequency energy is absorbed as it passes through the atmosphere more than is low-frequency energy, thus providing a spectral cue to distance [18]. Again, such cues are only unambiguous if the original spectrum of the sound at its source is known. Other cues are unambiguous in theory, but in practice are affected by perceptual limitations within the auditory system. For instance, parallax effects (change in angular source position with listener movement) are only poorly perceived [19] and are likely to be influenced by limits in the temporal resolution of the binaural system (binaural sluggishness and the subsequent poor resolution of source movement). Finally, the ratio of direct-to-reverberant energy can provide some absolute information about source distance [16, 20], but only when the listener is correctly able to discern the acoustic characteristics of the physical environment around him. Thus, in anechoic space, sources are often perceived as closer than their actual position, since such an environment has an unnaturally short reverberation time [21]. In addition, adding reverberation into a display, while providing source distance information, often decreases the salience of directional cues [22, 23, 24].

Of course, the salience of other source location cues is also limited by the physical cues which are heard by a listener; however, for most other cues, significant information can be heard given the naturally-occurring range of cues encountered: extreme values of ITDs are on the order of 700 ms for an average, adult headwidth [25, 26]; maximum IIDs caused by the acoustic shadow of the adult head can be as large as 20 dB [25], and the maximum size of notches in the received spectrum due to pinnae effects can be as large as 25 dB [26].

Unlike the perceptual effects discussed above, limitations in physical cues do not depend upon processing mechanisms in the auditory pathway. Thus, such limitations can be controlled externally, something which cannot be done to overcome peripheral, central, and cognitive effects in auditory localization.

3   Redefining Perceptual Limits in Auditory Localization

The above discussion highlights ways in which perceptual and physical phenomena affect the amount of spatial information received by the listener under a broad variety of circumstances. Information that is theoretically available to the listener may not be useful perceptually because of the way in which the information is processed in the auditory pathway. Under other conditions, limitations arise from the physical magnitude of the available cues, as in the discrimination of source direction in a simple acoustic environment. To the extent that performance on more

complex tasks (such as in an identification experiment) is proportional to the just noticeable difference (JND) in relevant cues, performance on even these more difficult tasks may be bounded by peripheral limitations.

In an auditory display in which localization cues are artificially synthesized, the magnitude of the physical cues produced is under the control of the designer of the system. In fact, the way in which location is encoded in physical stimuli can be modified by the system designer by making use of cues not normally used for the perception of source location. For instance, since distance cues are poorly perceived and ambiguous, a completely different cue dimension (such as pitch) could be used to encode distance. This approach to increasing cue resolution is imperfect, however. Under such circumstances, the user will have no experience in interpreting source pitch as a distance cue, and thus may need substantial training in order to make use of the distance information. Also, by using a new cue dimension for encoding source location, fewer dimensions are available for encoding other features of the data to be presented to the listener.

A different approach to improving resolution of source position is to emphasize the location cues which exist in natural environments. One way of emphasizing spatial cues is to increase the dynamic range of all the spatial cues which are heard by a listener, perhaps by simulating the cues that would reach a listener with a larger-than-normal head. Such a listener would receive larger-than-normal ITD and IID cues as well as larger-than-normal spectral cues (from his enlarged pinnae). As with the first approach, there are some possible problems with this scheme. The listener must be able to make use of the enlarged dynamic range of the resultant cues (e.g., must be able to accurately perceive larger-than-normal ITDs, a feat which may be limited by the distribution of ITD-sensitive cells in the auditory pathway). In addition, the change in the mapping from cue magnitude to source location will also affect the average (signed) error in perceived source direction (also known as bias). As with cue substitution, training is necessary in order to overcome response bias. However, although positional errors may arise because of the this type of remapping (especially prior to training), emphasized cues will at least be perceived immediately as spatial cues. There is therefore reason to believe that training a listener to use emphasized spatial cues would be easier than would training a listener to interpret an arbitrary cue (such as pitch) as a positional cue.

Work examining the perceptual effects of changing the mapping from physical cue to source position has been investigated for many modalities and under many different exposure conditions. Studies of the effects of changes in the mapping from location cues to sound source location have been reviewed in Shinn-Cunningham et al. [27]. Most previous studies of such changes have explored rotating the interaural axis and have focussed solely on how localization bias is affected. From the standpoint of the information received by a listener in an auditory display, changes in performance resolution are of equal, if not greater importance. In addition, none of these earlier studies has examined the extent to which it is possible to increase performance resolution by increasing physical-cue dynamic range.

3.1   Preliminary Results

The question of how performance resolution and bias are affected by increasing physical location cues has been examined using a real-time auditory display system consisting of a Convolvotron (made by Crystal River Engineering), a Bird electro-magnetic head tracker, and a 386-based PC. This real-time "spatialization" system includes compensation for listener movement in order to simulate externally-stable sound source locations.

Since the Convolvotron was designed to present only localization cues in the normal range, use of the system restricted the total dynamic range of localization cues. Therefore, for the exper-

iments presented here, we examined how subjects adapt to nonlinearly remapping the azimuthal auditory cues such that cue magnitudes were increased only in a limited region and overall dynamic range was equal to the normal dynamic range. The remapping function employed is shown in Figure 1 and is given by:

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

With this transformation, a source at azimuth Θ is presented with the normal location cues that would correspond to the position $f_n(\Theta)$. For the experiments described here, n was set to 3. As can be deduced from Figure 1, when $n = 3$ source positions are displaced laterally relative to the position heard using normal cues. The differences in localization cues for two sources in the frontal region (from -30 to +30 degrees in azimuth) are larger than normal with this remapping (and thus should yield better-than-normal resolution), while two locations off to the side would give rise to more similar cues than are normally heard (and would yield poorer-than-normal resolution), thus creating an artificial "acoustic fovea" in front of the listener. In addition to affecting resolution, however, this transformation was also expected to cause a bias whereby sources were perceived farther off-center than were their actual locations.

Subjects were seated in front of a 3-foot diameter arc of 13 lights spaced by 10 degrees, from 60 to +60 degrees and numbered left to right from 1 to 13. The lights were used to train subjects to interpret acoustic location cues appropriately for the remapping function employed. Subjects were tested using a 13-alternative identification paradigm in which they named the position of a 500-ms-long click train that was simulated from one of the 13 marked locations.

In all experiments, subjects were given "normal" cues (without the transformation described in Eq. (1) first, then altered cues, and finally normal cues again. This order of runs allowed the subjects' performance to be tracked over the course of the session, and provided "normal" control results against which performance with the transformed localization cues could be compared. Many test runs were performed by each subject in each session; however, results from only the four most crucial of these runs are presented here: (1) the first run using normal cues, (2) the first run using transformed cues, (3) the final transformed-cue run, and (4) the first normal-cue run following exposure to altered cues. Various conditions were tested, including experiments in which subjects were trained by turning their head to face a coordinated real-light/simulated-sound; experiments in which subjects were simply given correct answer feedback after each trial in the identification task; experiments in which the complexity of the acoustic background was varied; experiments in which the number of positions used in the experiment was altered; and experiments in which the absolute strength of the remapping function employed was varied. None of the changes in experimental protocol caused any changes in the basic pattern of results reported here.

Subjects completed 8 identical, 2-hour-long sessions. Results from each subject were found by combining responses for each of the four crucial runs across sessions. Data were processed in a variety of ways to find estimates of the resolution (accuracy with which adjacent positions could be discriminated by the subject) and bias (average signed error in response for each position). While the various processing schemes differed in their computational complexity (and in their ability to account for effects of the experimental paradigm employed in the study), the results were consistent for all processing methods. For this reason, results from the most simple processing scheme are described and reported here.

For each subject and run the average response and the standard deviation in response were found for each of the 13 possible locations from the confusion matrix for that run and subject. Resolution between adjacent pairs of positions was estimated as the difference in the average
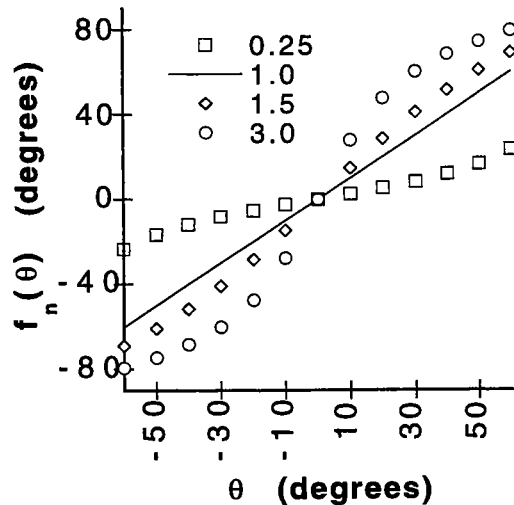
Figure 1: Transformation used to create better-than-normal resolution in front of the listener ($n > 1$) or better-than-normal resolution to the sides of the listener ($n < 1$). In experiments reported here, $n = 3$ (circles) and marked positions from -60 to $+60$ degrees were presented. For a position Θ, the normal HRTF cues used to present the source correspond to the normal position $f_n(\Theta)$.

responses normalized by the geometric average of the standard deviations for the two positions (similar to the standard sensitivity measure d'). Bias was estimated for each position as the difference between mean response and the correct response normalized by the standard deviation for the position. Resolution and bias were then averaged across subjects for each position to generate the results presented here.

The first test using normal cues was expected to show little systematic bias, and provided a baseline for resolution performance. The first test with the transformed localization cues was expected to show a strong bias whereby source locations were heard farther off center than they were in the remapped space. Given the remapping function, resolution in the first test with altered cues was expected to show improved resolution in the acoustic fovea and decreased resolution at the sides. After training with the altered cues, bias was expected to decrease in magnitude for all positions, and resolution was expected to either (1) remain enhanced within the acoustic fovea (if resolution depended solely on the difference between the magnitudes of the cues at adjacent positions) or (2) change with time (if performance depended upon high-level cognitive factors which were affected by training). Finally, results from the first, post-training, normal-cue test were expected to show either (1) a bias in the direction opposite that shown when the transformed cues were first presented (if changes in subject performance were unconscious, and therefore could not be immediately "turned off" by the listener following training) or (2) little bias (if subjects were capable of consciously interpreting cue as either normal or altered, as appropriate). Resolution on the final normal-cue test was predicted to be like resolution with normal cues prior to exposure, unless training affected resolution.

Regardless of the exact training method employed (and the method used to process the results), the bias and resolution results were similar. Figure 2 shows bias results as a function of source position for a typical experiment. Normal-cue runs are plotted with circles; altered-cue runs with exes. The solid lines represent runs prior to altered cue training while dashed lines correspond to the "adapted" results. In all bias results, there is an edge effect due to the experimental paradigm:
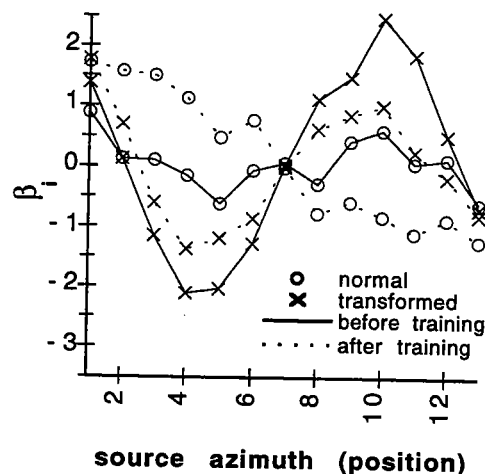
β_i vs **source azimuth (position)**

Legend:
○ normal
✕ transformed
—— before training
······ after training

Figure 2: Bias results for normal and transformed cues, prior to and following training with transformed cues.

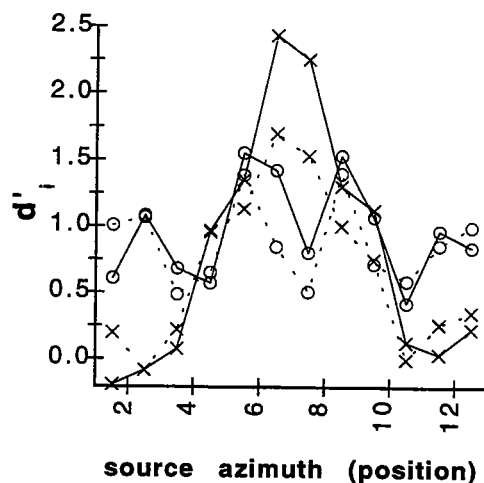**d'_i** vs **source azimuth (position)**

Figure 3: Resolution results for normal and transformed cues, prior to and following training with transformed cues.

since responses were limited to the 13 positions used, bias had to be positive (or zero) for the leftmost position and negative (or zero) for the rightmost position.

Results from the first normal cue run (solid circles) showed some biases, although these errors were significantly smaller than those found in other runs. A strong bias occurred in the first test with transformed cues (solid exes) in the direction predicted by the transformation and the aforementioned edge effect (subjects heard sources farther off-center than they were). Results for the test using transformed cues after training (dashed exes) showed a clear reduction in bias over the whole range of positions tested; however, this adaptation was not complete. Finally, a negative after-effect is seen in the results from the final normal cue test following exposure (dashed circles) indicating that performance was not controlled solely by conscious correction which could be "turned off" at will.

Resolution results from the same experiment are shown in Figure 3. Resolution for normal-cue runs showed a systematic pattern (which may be due to positional dependencies on the accuracy of the simulation as well as true differences in resolution arising from perceptual issues) which was

74

fairly consistent for pre- and post-exposure runs, although there appears to be a slight decrease in resolution following training, at least for the central positions. As expected for the transformation employed, resolution on the first run using the transformed cues was enhanced for positions in the central region and degraded at the edges of the range. Of most interest are the results for the final, altered-cue test: although resolution remains enhanced over that achieved with normal cues, there is a decrease in resolution compared to results for the first test using the transformed cues, demonstrating that resolution does not depend only on physical cues.

3.2 Discussion

The results demonstrate that it is possible to achieve better-than-normal resolution by altering the physical cues used to encode sound source location. In addition, subjects are able to learn remappings between acoustic cues and physical locations as demonstrated by their ability to partially overcome bias with training.

Although subjects can achieve better-than-normal resolution, there appear to be small decreases in resolution with training. One possibility is that the decrease in resolution may be explained by considering the effective range before and after adaptation; prior to adaptation, subjects expect positions to span 120 degrees (from -60 to $+60$ degrees), but during training, they hear positions covering a much larger range of physical cues [from $f_3(-60)$ to $f_3(60)$ or 158.2 degrees] and learn to attend to this broader range. In general, resolution for a given pair of stimuli decreases in experiments using large ranges relative to resolution in small-range experiments [12, 13]. This decrease is generally associated with cognitive limitations on memory. Earlier models of resolution performance (such as those by Durlach and Braida [12, 13]) may be extended to include specifications for how performance changes over time in an adaptation experiment by specifying how the effective mean response and the effective range depend upon training (e.g., see Shinn-Cunningham [29]).

Conclusions

Performance on all tasks depends upon a number of factors, including the physical stimuli received by the human and the way in which information is processed by the human operator. In selecting cues to be used in acoustic displays, consideration of these factors is crucial. Under many circumstances, cues which are theoretically available to a listener cannot, in practice, be utilized by the subject. In such cases, careful consideration of the way in which the available information is processed by the listener may give insight into why performance is poor. On the other hand, some tasks are not primarily limited by the processing performed within the auditory pathway, but by the physical magnitude of the cues reaching the listener. Since such cues are under the complete control of the designer of a human-machine interface, these types of limitations can be directly manipulated. This approach was investigated in some simple experiments of azimuthal localization using an acoustic display. Results were encouraging: remapping acoustic cues to new positions improved resolution for the task tested, and subjects were able to partially adapt to the imposed mapping. However, as in all tasks, performance is not limited solely by the physical stimuli. Even for the relatively simple task employed, it appears that resolution may depend on the state of the observer (e.g., cognitive factors).

These results further emphasize the importance of having a thorough understanding of the human receiver in trying to design an auditory display. The principles outlined here can be applied to a variety of interface questions, both in auditory display (in choosing encodings of other acoustic cue dimensions such as pitch, timbre, etc.) and in other modalities (in choosing how to create

cost-effective and useful visual displays; for instance, by taking into account the nonlinear spatial resolution of stimuli on the retina [30] and the relatively poor temporal resolution of the visual system [31]).

Acknowledgments

References

[1] Yin, T. C. T., and J. C. K. Chan. "Interaural Time Sensitivity in Medial Superior Olive of Cat." *J. Neurophys.* **64** (1990): 465–488.

[2] Mills, A. W. "On the Minimum Audible Angle." *J. Acous. Soc. Am.* **30** (1958): 237–246.

[3] Hafter, E. R., and L. A. Jeffress. "Two-Image Lateralization of Tones and Clicks." *J. Acous. Soc. Am.* **44(2)** (1968): 563–569.

[4] Dye, Jr., R. H. "The Combination of Interaural Information Across Frequencies: Lateralization on the Basis of Interaural Delay." *J. Acous. Soc. Am.* **88(5)** (1990): 2159–2170.

[5] Buell, T. N., and E. R. Hafter. "Combination of Binaural Information Across Frequency Bands." *J. Acous. Soc. Am.* **90** (1990): 1894–1900.

[6] Stellmack, M. A., and R. H. Dye Jr. "The Combination of Interaural Information Across Frequencies: The Effects of Number and Spacing of Componenets, Onset Asynchrony." *J. Acous. Soc. Am.* **93** (1993): 2933–2947.

[7] Bregman, A. S. *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press, 1990.

[8] Zurek, P. M. "The Precedence Effect." Directional Hearing. Eds. W. A. Yost and G. Gourevitch, 85–105. New York: Springer-Verlag, 1987.

[9] Shinn-Cunningham, B. G., P. M. Zurek, and N. I. Durlach. "Adjustment and Discrimination Measurements of the Precedence Effect." *J. Acous. Soc. Am.* **93** (1993): 2923–2932.

[10] Perrott, D. R. "Studies in the Perception of Auditory Motion." In *Localization of Sound: Theory and Applications in Room,* edited by R. W. Gatehouse, 169–193. Groton, CT: Amphora Press, 1982.

[11] Shinn-Cunningham, B. G., P. M. Zurek, R. K. Clifton, and N. I. Durlach. "Cross-Frequency Interations in the Precedence Effect." *J. Acous. Soc. Am.* (1995): in press.

[12] Durlach, N. I., and L. Braida. "Intensity Perception. I. Preliminary Theory of Intensity Resolution." *J. Acous. Soc. Am.* **46(2)** (1969): 372–383.

[13] Braida, L. D., and N. I. Durlach. "Intensity Perception. II. Resolution in One-interval Paradigms." *J. Acous. Soc. Am.* **51(2)** (1972): 483–502.

[14] Searle, C. L., L. D. Braida, M. F. Davis, and H. S. Colburn. "Model for Auditory Localization." *J. Acous. Soc. Am.* **60** (1976): 1164–1175.

[15] Mershon, D. H., and J. N. Bowers. "Absolute and Relative Cues for the Auditory Perception of Egocentric Distance." *Perception* **8** (1979): 311–322.

[16] Mershon, D. H., and L. E. King. "Intensity and Reverberation as Factors in Auditory Perception of Egocentric Distance." *Percept. Psychophys.* **18** (1975): 409–415.

[17] Coleman, P. D. "Failure to Localize the Source Distance of an Unfamiliar Sound." *J. Acous. Soc. Am.* **34** (1962): 345–346.

[18] Little, A. D., D. H. Mershon, and P. H. Cox. "Spectral Content as a Cue to Perceived Auditory Distance." *Perception* **21** (1992): 405–416.

[19] Speigle, J. M., and J. M. Loomis. "Auditory Distance Perception by Translating Observers." *Proceedings of the IEEE 1993 Symposium on Research Frontiers in Virtual Reality*, 92–99. San Jose, California: IEEE Computer Society, 1993.

[20] Mershon, D. H., W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan. "Effects of Room Reflectance and Background Noise on Perceived Auditory Distance." *Perception* **18** (1989): 403–416.

[21] Gardner, M. B. "Distance Estimation of 0-degree or Apparent 0-degree-Oriented Speech Signals in Anechoic Space." *J. Acous. Soc. Am.* **45(1)** (1969): 47–53.

[22] Giguere, C., and S. M. Abel. "Sound Localization: Effects of Reverberation Time, Speaker Array, Stimulus Frequency, and Stimulus Rise/Decay." *J. Acous. Soc. Am.* **94(2)** (1993): 769–776.

[23] Hartmann, W. M. "Localization of Sound in Rooms." *J. Acous. Soc. Am.* **74** (1983): 1380–1391.

[24] Rakerd, B., and W. Hartmann. "Localization of Sound in Rooms II. The Effects of a Single Reflecting Surface." *J. Acous. Soc. Am.* **78** (1985): 524–533.

[25] Mills, A. W. "Auditory Localization." In *Foundations of Modern Auditory Theory*, edited by J. V. Tobias, 303–348. New York: Academic Press, 1972.

[26] Wightman, F., and D. Kistler. "Headphone Simulation of Free-Field Listening. I. Stimulus Synthesis." *J. Acous. Soc. Am.* **85** (1989): 858–867.

[27] Shinn-Cunningham, B. G., H. Lehnert, G. Kramer, E. M. Wenzel, and N. I. Durlach. "Auditory Displays." In *Spatial and Binaural Hearing*, edited by R. Gilkey and T. Anderson. New York: Erlbaum, 1995.

[28] Lippmann, R. P., L. D. Braida, and N. I. Durlach. "Intensity Perception. V. Effect of Payoff Matrix on Absolute Identification." *J. Acous. Soc. Am.* **59(1)** (1976): 129–134.

[29] Shinn-Cunningham, B. G. "Adaptation to Supernormal Auditory Localization Cues in an Auditory Virtual Environment." Ph.D. Thesis in Electrical Engineering, Massachusetts Institute of Technology, 1994.

[30] Schwartz, E. L. "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception." *Biol. Cyber.* **25** (1977): 181–194.

[31] Welch, R., and D. H. Warren. "Intersensory Interactions." In *Handbook of Perception and Human Performance*, edited by K. R. Boff, et al., I: 25.1-25.36. John Wiley and Sons, 1986.