NCC 2-542
IN-71-CR
320487
p. 18

Hearing in Three Dimensions:
Sound Localization

Frederic L. Wightman

Doris J. Kistler

Waisman Center, University of Wisconsin

(NASA-CR-187706) HEARING IN THREE DIMENSIONS: SOUND LOCALIZATION (Wisconsin Univ.) 18 p
CSCL 20A
N91-13971
Unclas
G3/71 0320487

ABSTRACT

Our ability to localize a source of sound in space is a fundamental component of the three-dimensional character of the "Sound of Audio." For over a century scientists have been trying to understand the physical and psychological processes and physiological mechanisms that subserve sound localization. This research has shown that important information about sound source position is provided by interaural differences in time of arrival, interaural differences in intensity, and direction-dependent filtering provided by the pinnae. Progress has been slow, primarily because experiments on localization are technically demanding. Control of stimulus parameters and quantification of the subjective experience are quite difficult problems. Recent advances, such as the ability to simulate a three-dimensional sound field over

headphones, seem to offer potential for rapid progress. Research using the new techniques has already produced new information. It now seems that interaural time differences are a much more salient and dominant localization cue than previously believed.

1. INTRODUCTION

The "Sound of Audio" is inherently three-dimensional. Almost regardless of how that sound gets to our ears, at a live concert or via our "walkman" headset, it has an undeniable three-dimensional character to it. The violins are on the left in front, and the tubas are on the right toward the rear. Even the words we use to describe the "sound of audio" convey a three-dimensional quality. We describe sound images as broad, thin, or flat, and as having width, height, and depth.

2. BASIC RESEARCH ISSUES

Researchers in psychoacoustics have long been interested in what it is about sounds and how they are processed by the human sensory system that gives them their three-dimensional quality. Most of our research has focussed on one aspect of that problem, namely the mechanisms and processes that underly our ability to localize, or to assign spatial positions to sound images. The general approach we follow in this research involves mapping relations between stimulus variables (acoustical characteristics of the sounds) and response variables (perceived directions, etc.). The aim, of course, is to learn about what goes on between stimulus and response, or, in other words, how the system works. Obviously, if we were to study the response to all possible stimuli, we would learn all there is to know about the system. The hope is that if we choose input stimuli correctly we will be able to reduce the scope of the problem considerably. Such an approach is familiar to anyone who has studied linear systems theory, where the input stimulus of choice is the sinusoid. We will call this the "linear systems" approach.

The success of the "linear systems" approach to the study of perception relies on accurate specification and control of the stimulus variables, and accurate measurement of the response variables. In many studies, these requirements are easy to meet. For example, if we are interested in the detectability of a sound, it is a relatively simple matter to specify and control the intensity of the sound, and while it is a much less simple matter, we are confident that we know how to quantify the detectability of the sound. In the case of sound localization, however, the problems of stimulus control and response quantification are formidible.

On the stimulus side we face two problems. One is that "the stimulus" consists of more than just the sound itself. In other words, sound localization depends not only on acoustical factors, but also on non-acoustical factors such as memory, context, vision, etc. Even if we restrict our study to the acoustical factors alone, we must deal with the very difficult matter of measuring and controlling the stimulus. It is now generally agreed that the acoustical stimulus that should be measured is the sound pressure waveform (or energy input) at the listener's eardrum. Measurement at a listener's eardrum is difficult at best. Moreover, the many reflections and complex interactions of sound waves in a typical room make control of the acoustical stimulus at the ears of a listener nearly impossible. The use of an anechoic room solves some of the problems, but even in this artificial environment control of the stimulus is a difficult matter.

Measurement of the response in a localization experiment is no less challenging. The problem here is that what we wish to measure, the apparent position (or any other quality, for that matter) of a sound image in auditory space, is a purely subjective thing that exists only in the head of the listener. Thus, our measurements must be indirect, relying on verbal report or some other kind of response (e.g., pointing) from the listener. There is ample evidence that responses such as these can be heavily influenced by factors that have little relevance to apparent image position, such as the range and distribution of stimulus and/or response alternatives presented in the experiment. The implication is that while apparent position may be invariant under certain experimental manipulations, the listener's report may well vary, as a result of other, apparently irrelevant manipulations. Great care must be taken to reduce the contaminating influence of these factors in localization experiments, and we must always be aware that the potential for contamination exists.

2.1 CLASSICAL STUDIES

In spite of all the difficulties, systematic research on sound localization has been going on for over a century. In the last decade alone, almost 50 experiments on the subject have been reported in major scientific journals. The early work attempted to determine the major acoustical cues to apparent image position and how those cues might be processed by the auditory system. To make the acoustical analysis tractable, the head was assumed to be a rigid sphere and the ears to be points on the surface of the sphere, separated by 180 degrees. These assumptions led to the hypothesis that there exist just two potential cues in a typical localization task (e.g., localization of sources on the horizontal plane). These were the interaural differences in time of arrival (sound reaches the closer ear as much as 700 microseconds before the opposite ear) and interaural differences in intensity (at high frequencies the head casts an acoustic "shadow", such that the sound is more intense at the ear closer to the source). Acoustical measurements on human listeners (e.g., Feddersen, et al., 1957) have verified the presence of these cues, and have quantified the dependence of these cues on the azimuth of sinusoidal sources. Psychophysical experiments, conducted with headphones to allow for independent manipulation of the cues, have shown that the interaural difference cues are indeed detectable (Zwislocki and Feldman, 1956; Mills, 1960). There is also considerable indirect evidence that these cues are important for localization. For example, at low frequencies, the interaural time (or phase) difference that is introduced when a stimulus is moved a just-noticeable angle off the midline (Mills, 1958) is about the same as the just-detectable interaural time (phase) difference measured under headphones. The same correspondence holds for interaural intensity differences at high frequencies (see Mills, 1960 for a summary of these points).

2.2 THE DUPLEX THEORY - LATERALIZATION EXPERIMENTS

The assumption of simplified geometry, the acoustic measurements, and the results of early psychophysical experiments form the basis of the so-called

"Duplex Theory" of localization, outlined as early as the turn of the century by Lord Rayleigh (Strutt, 1907). In its simplest form this theory holds that localization of low-frequency sounds is dependent on interaural time differences, and localization of high-frequency sounds on interaural intensity differences. Division of the frequency scale appeared necessary since temporal coding in the auditory system had been observed only at low frequencies, and interaural intensity differences exist only at high frequencies. A great deal of research was stimulated by the Duplex Theory, and as a result we have learned a lot about processing (e.g., detection and discrimination) of interaural time and intensity differences. The research almost always involved presentation of sounds to listeners over headphones, to allow precise control of interaural differences in time and intensity. Unfortunately, the extent to which the results of these experiments can be generalized to actual localization conditions may be quite limited. The headphone experiments were called "lateralization", as opposed to "localization" experiments, in recognition of the fact that stimuli presented over headphones are rarely externalized, even though interaural time and intensity differences appropriate to an externalized source are present. Thus, while lateralization experiments often claim to address issues of localization, the internalized character of the stimuli makes the claim questionable. For example, the fact that a subject listening over headphones can discriminate or detect interaural differences may say very little about how discriminations of azimuth and elevation changes are accomplished in free field. Similarly, lateralization paradigms can provide only indirect evidence on the viability of theories of localization such as the Duplex Theory.

2.3 RECENT ADVANCES

Progress during the last few years in the stimulus control and response measurement areas has brought both a recognition of the limitations of lateralization experiments and a flurry of new experiments on localization. Techniques have been developed to compensate digitally for individual loudspeaker characteristics (Wightman and Kistler, 1980), to position and to move sound sources in an anechoic room (Oldfield and Parker, 1984; Perrott and Musicant, 1977), to allow subjects to "point their heads" toward the apparent position of a sound image (Perrott, Ambarsoom, and Tucker, 1987; Mackous and Middlebrooks, 1990), or to point a "gun" at the apparent position (Oldfield and Parker, 1984) as means of responding. These developments at least partially solve some of the most difficult technical problems associated with localization research. A few of the general findings that have emerged from the new wave of localization research are: 1) complex, broadband sounds are localized best; 2) high frequencies must be present for accurate judgements of apparent source elevation; and 3) localization is most precise in front and at ear level, and least precise in the rear at high elevations.

2.4 IMPORTANCE OF PINNA CUES

Many of the recent experiments have emphasized the role of localization cues other than interaural time and intensity differences. Most notable, perhaps, are the studies of the cues provided by a listener's pinnae (Batteau, 1967; Wright, et al., 1974.) It has been known for some time that as a result of interactions of a sound with reflections from the convolutions of the pinnae, a direction-dependent filtering is imposed on an incoming stimulus. It is now clear that this spectral shaping is a very important cue for localization (see Butler, 1975, for a review of the research on this issue). One experimental demonstration of this is the fact that when the cavities of the pinnae are filled with putty, localization ability is markedly impaired (Gardner and Gardner, 1973.) Other recent experiments have considered the role of head movements (Thurlow and Runge, 1967), visual cues (Gardner 1968), a-priori knowledge of stimulus properties (Coleman, 1962), and postural variables (Lackner, 1983). The specific contributions of these factors to our perception of auditory space is not well understood, though it is agreed that in certain listening situations they are important.

While recent research recognizes the complexity of actual localization conditions, and the importance of cues such as those provided by the pinnae, there have been only a few attempts to manipulate these cues systematically. This is understandable, since until recently, it has not been technologically feasible. Schroeder and Atal (1963), and Morimoto and Ando (1982), have described a technique using two loudspeakers and digitally-generated stimuli whereby the illusion of a sound source at any arbitrary point in space can be created (so long as the position of the listener is known precisely). Bloom (1977) and Watkins (1978) have attempted to simulate source elevation changes by altering the spectrum of the source in a manner analogous to pinna filtering. Blauert (1969), and Butler and Planert (1976) have made similar attempts to alter the apparent location of a sound by modifying the spectrum. The success of these early attempts has been limited, especially since the experiments included no direct tests of the psychophysical adequacy of the manipulation.

As a consequence of the difficulties associated with systematic manipulation and control of localization cues there are still large gaps in our understanding of how localization works. Moreover, the areas of uncertainty are also the most basic. For example, it is still not entirely clear what characteristics of a sound cause it to be externalized. There are suggestions that the filtering action of the pinnae is important in this regard, but the issue is far from settled. Our inability to address such basic questions is almost certainly a result of the lack of necessary technology. This is exemplified by the fact that in spite of the overwhelming experimental advantages of headphone stimulus presentation, there are few empirically-validated reports of a duplication of the free-field experience with headphones (Wightman and Kistler, 1989a,b.)

3. SIMULATION OF AUDITORY SPACE WITH HEADPHONES

In our laboratory, we use digital signal processing techniques to synthesize stimuli that mimic those that actually reach a listener's ears in a free sound field. When these stimuli are presented over headphones, they produce faithful illusions of sound sources outside the listener's head (we call these "virtual sources"), at positions in space that we can specify in advance. The general aim of our technique is to use headphones to produce acoustic waveforms at a listener's two eardrums that are as close as possible to the acoustic waveforms produced by a sound source in real auditory space. First, using probe microphones and a sound source in an anechoic room, we measure, for each of a listener's ears, the free-field-to-eardrum transfer function at the desired point in auditory space. Next, we measure a comparable transfer function with our test sound transduced by the headphones. Then an FIR digital filter is computed by dividing the free-field transfer function by the headphone transfer function. Stimuli are then passed through this digital filter and transduced by the headphones. In this process, the headphone response should cancel and the free-field characteristics, consisting mostly of effects caused by the head and pinnae, should be superimposed on the stimulus. The resulting waveform at a listener's eardrums should be the same as if the stimulus had been produced by a loudspeaker at the desired position in auditory space. The results of actual measurements suggest that the error is quite small (Wightman and Kistler, 1989a.) All those who have listened to the synthesized stimuli report that the virtual sources are externalized, and located at the intended positions in auditory space. In our psychophysical experiments 10 listeners judged the apparent positions of both real and virtual sound sources; the results were consistent with the listeners' reports. The perceived locations of real and virtual sources were nearly identical (Wightman and Kistler, 1989b). Figure 1 shows sample results from the experiment.

3.1 RECENT RESEARCH IN SIMULATED AUDITORY SPACE

We have been using the virtual source techniques in a variety of experiments designed to answer some very basic questions regarding the cues used for sound localization and how those cues might be processed. The complete stimulus control offered by the virtual source techniques allows us to conduct experiments that would be impossible with real sources. For example, we can investigate the relative salience of interaural time and intensity differences by independently manipulating the amplitude and phase characteristics of the digital filters we use to produce the virtual sources. With free-field sources such independent control of the amplitude and phase characteristics of a sound at the listener's ears is nearly impossible.

One experiment we have conducted that takes advantage of the virtual source technique asked listeners to judge the apparent positions of sound images constructed such that interaural time cues and interaural intensity cues were in conflict. Thus, if the apparent position of a given stimulus was determined by interaural time cues, we would expect listeners to make the response (e.g., point in the direction) appropriate to the time cue, and if position was determined by interaural intensity cues they would make the response appropriate to the intensity cue. We fully expected that the results would suggest that both cues were operative, and thus that responses would be at some intermediate position, or spread out between the two positions. In fact, so long as low frequencies were present in the stimulus, apparent position was determined completely by the time cue.

Figure 2 shows sample results from this experiment. In the top panels (Fig. 2a) we show judgements of apparent position made by one listener to 36 wideband (200 Hz - 14 kHz) virtual sources. Each data point represents the average position judgement from eight presentations of the stimulus. Listeners report apparent position by verbally indicating apparent source azimuth, elevation, and distance (Wightman and Kistler, 1989b). The data on the left are from a condition in which time and intensity cues were normal. The fact that apparent azimuth and elevation agree well with intended ("target" on the figure) azimuth and elevation indicates the general adequacy of the virtual source technique. The data on the right are from a condition in which the interaural time difference cue was the same for all 36 stimuli while the interaural intensity difference cues were normal. The interaural time difference at each frequency was set to that value appropriate to a stimulus at 90 degrees azimuth and 0 degrees elevation (i.e., directly opposite the listener's right ear). Thus, we say that for all the stimuli, the interaural time cue "pointed" to "90,0", and as a result, for all but one of the stimuli (that one with a target position of "90,0") the interaural time and intensity cues were in conflict. Note that for all stimuli the listener's judgements of apparent source azimuth were consistent with the time cue, and were concentrated around values close to 90 degrees. Even when the target source position was at -90 degrees (on the opposite side of the head), the listener's judgements followed the time cue. In this case, large interaural intensity differences signalled a source position directly opposite that indicated by the time cue, but not a single judgement was ever made (by our 8 subjects) that followed the intensity cue. Note also that the listener's judgements of apparent source elevation were compressed around 0 degrees. This result is consistent with a view that interaural time difference is a "dominant" localization cue; the only source elevation that is consistent with the large interaural time difference present at "90,0" is zero.

With low frequencies removed from the stimulus, fixing the interaural time difference cue had no effect. The lower pair of panels in Figure 2 show data from a condition identical to that described above, except that the stimuli were high-pass filtered at 2.5 kHz. Note that in this case the interaural time cue modification had no apparent effect. The listener's judgements of apparent source position in the condition in which interaural time differences at each frequency "pointed to" "90,0" were the same as in the condition in which both time and intensity cues were normal.

The dominance of interaural time differences in determining the apparent azimuth and elevation of a sound image may have important implications for sound engineers. For wideband sources or sources that contain mostly low frequencies (2 kHz and below), modification of the intensity ratio between left and right channels of a stereo recording cannot be expected to have any influence on the apparent position of the resultant sound image. The group delay between channels, on the other hand, will dominate apparent position.

4. CONCLUSION

The physical, physiological, and psychological mechanisms and processes that subserve the three-dimensional character of the "Sound of Audio" are just beginning to be revealed by modern research on sound localization. We have come a long way since the Duplex Theory and the early experiments with headphones and sinusoids. While the picture grows increasingly complex, modern advances such as the virtual source technique represent powerful tools for use in our research. We can expect very rapid progress in this area during the next decade.

ACKNOWLEDGMENTS

This research was supported by a grant from the NIH (NIDCD, P01-NS12732) and by a Cooperative Research Agreement with NASA-Ames (NCC2-542). Special thanks go to Ms. Marianne Arruda, without whose tireless and expert assistance this work would not have been possible.

REFERENCES

Batteau, D. W. (1967). The role of the pinna in human localization. Proceedings of the Royal Society of London, 168 B, 158-179.

Blauert, J. (1969). Sound localization in the median plane. Acustica, 22, 205-213.

Bloom, P. J. (1977). Creating source elevation illusions by spectral manipulation. Journal of the Audio Engineering Society, 25 (9), 560-565.

Butler, R. A. (1975). The influence of the external and middle ear on auditory discriminations. In W. D. Keidel and W. D. Neff (Eds.), Handbook of Sensory Physiology. Berlin: Springer-Verlag.

Butler, R. A., and Planert, N. (1976). The influence of stimulus bandwidth on localization of sound in space. Perception and Psychophysics, 19, 103-108.

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

Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (1957). Localization of high-frequency tones. Journal of the Acoustical Society of America, 29, 988-991.

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

Gardner, M. B., and Gardner, R. S. (1973). Problem of localization in the median plane: effect of pinnae cavity occlusion. Journal of the Acoustical Society of America, 53, 400-408.

Lackner, J. R. (1983). Influence of posture on the spatial localization of sound. Journal of the Audio Engineering Society, 31, 650-661.

Makous, J. C. (1990). Two-dimensional sound localization by human listeners. Journal of the Acoustical Society of America, in press.

Mills, A. W. (1958). On the minimum audible angle. Journal of the Acoustical Society of America, 30, 237-246.

Mills, A. W. (1960). Lateralization of high-frequency tones. Journal of the Acoustical Society of America, 32, 132-134.

Morimoto, M., and Ando, Y. (1982). On the simulation of sound localization. In Localization of Sound: Theory and Applications, Gatehouse, R. W. (ed.). Groton, CT, Amphora Press.

Oldfield, S. R., and Parker, S. P. A. (1984). Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions. Perception, 13, 581-600.

Perrott, D. R., Ambarsoom, H., and Tucker, J. (1987). Changes in head position as a measure of auditory localization performance: Auditory psychomotor coordination under monaural and binaural listening conditions. Journal of the Acoustical Society of America, 82, 1637-1645.

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

Schroeder, M.R., and Atal, B.S. (1963). Computer simulation of sound transmission in rooms. IEEE Int. Conv. Rec. 11, 150-155

Strutt, J. W. (Lord Rayleigh). (1907). On our perception of sound direction. Philosophy Magazine, 13, 214-232.

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

Watkins, A. J. (1978). Psychoacoustical aspects of synthesized vertical locale cues. Journal of the Acoustical Society of America, 63, 1152-1165.

Wightman, F. L., and Kistler, D. (1980). A new "look" at auditory space perception. In G. vanden Brink and F. Bilsen (Eds.), Psychophysical, Physiological, and Behavioral Studies in Hearing. Delft: University Press.

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

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

Wright, D., Hebrank, J. H., and Wilson, B. (1974). Pinna reflections as cues for localization. Journal of the Acoustical Society of America, 56, 957-962.

Zwislocki, J., and Feldman, R. S. (1956). Just noticeable differences in dichotic phase. Journal of the Acoustical Society of America, 28, 860-864.

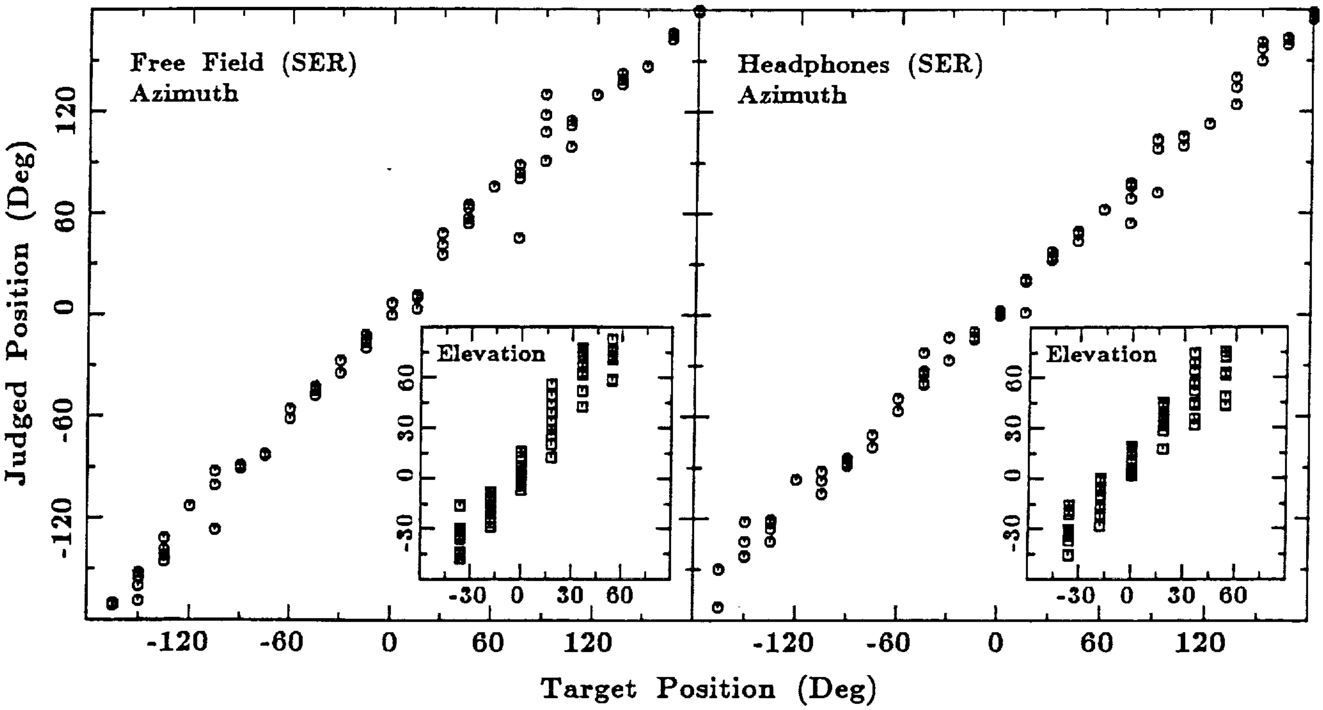

Figure 1. Scatterplots showing actual source azimuth (and, in the insets, elevation) versus judged source azimuth for subject SER in both the free-field and virtual source conditions. Each data point represents the centroid of at least 8 judgements. Seventy-two source positions are represented in each panel. Data from 6 different source elevations are combined in the azimuth panels, and data from 24 different azimuths are combined in the elevation panels. Note that the scale is the same for azimuth and elevation plots.

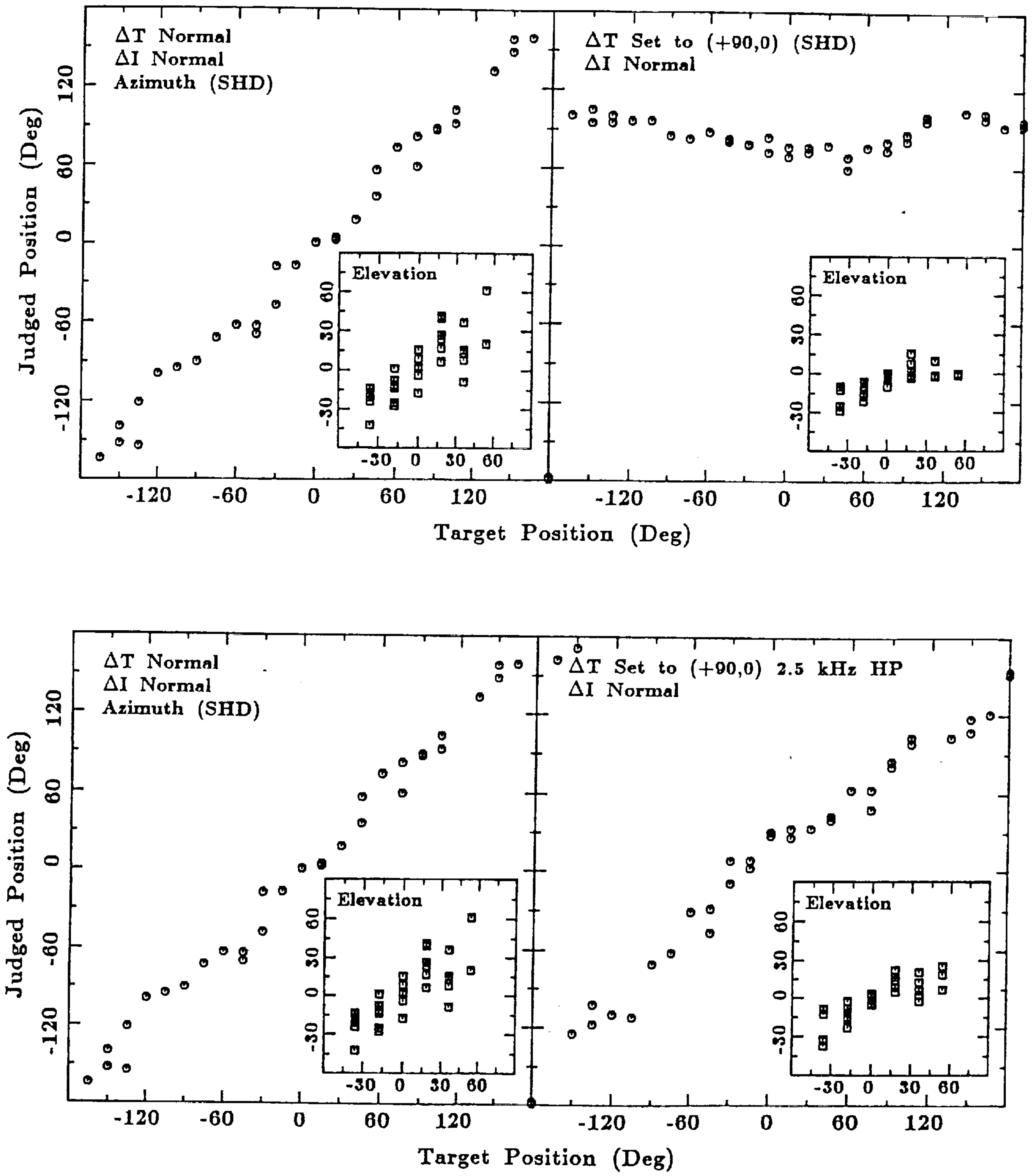

Figure 2. Scatterplots (similar to those in Figure 1.) showing data from conditions in which interaural time difference and pinna cues were in conflict. In the top panels, performance with normal virtual-source stimuli (left) is compared to performance when interaural time cues consistently "point to" a source at "90,0" (directly opposite the listener's right ear.) In the bottom right panel (bottom left panel is the same as the top left panel) performance is shown for the condition in which interaural time cues "point to" "90,0" and the stimulus is high-pass filtered at 2.5 kHz. Data from a single subject (SHD) are shown.