NASA-CR-204683

Monaural sound localization revisited

Frederic L. Wightman
Department of Psychology and Waisman Center, University of Wisconsin—Madison, Madison, Wisconsin 53705

Doris J. Kistler
Waisman Center, University of Wisconsin—Madison, Madison, Wisconsin 53705

(Received 23 February 1996; revised 2 August 1996; accepted 25 September 1996)

Research reported during the past few decades has revealed the importance for human sound localization of the so-called "monaural spectral cues." These cues are the result of the direction-dependent filtering of incoming sound waves accomplished by the pinnae. One point of view about how these cues are extracted places great emphasis on the spectrum of the received sound at each ear individually. This leads to the suggestion that an effective way of studying the influence of these cues is to measure the ability of listeners to localize sounds when one of their ears is plugged. Numerous studies have appeared using this monaural localization paradigm. Three experiments are described here which are intended to clarify the results of the previous monaural localization studies and provide new data on how monaural spectral cues might be processed. Virtual sound sources are used in the experiments in order to manipulate and control the stimuli independently at the two ears. Two of the experiments deal with the consequences of the incomplete monauralization that may have contaminated previous work. The results suggest that even very low sound levels in the occluded ear provide access to interaural localization cues. The presence of these cues complicates the interpretation of the results of nominally monaural localization studies. The third experiment concerns the role of prior knowledge of the source spectrum, which is required if monaural cues are to be useful. The results of this last experiment demonstrate that extraction of monaural spectral cues can be severely disrupted by trial-to-trial fluctuations in the source spectrum. The general conclusion of the experiments is that, while monaural spectral cues are important, the monaural localization paradigm may not be the most appropriate way to study their role. [S0001-4966(97)02902-0]

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Yw [RHD]

INTRODUCTION

While human sound localization is generally acknowledged to be a process that depends predominately on acoustical stimulation of both ears, the study of monaural sound localization has captured the interest of hearing scientists since the turn of the century (Angell and Fite, 1901). In the past three decades, for example, more than 25 empirical studies have been published that deal explicitly with monaural localization. These studies are typically motivated by referring to weaknesses in the well-entrenched "duplex theory" of sound localization (Strutt, 1907). This theory holds that the apparent position of a sound is determined entirely by interaural time and level differences (ITDs and ILDs, respectively). It has been clear for some time that there are essential features of human sound localization that cannot be explained by ITDs and ILDs alone. That localization does not seem to be dramatically impaired on the median plane, where ITDs and ILDs are minimal, is one obvious example. The direction-dependent filtering provided by the pinnae is now acknowledged to be one of the most salient of the localization cues not incorporated in the duplex theory. Pinna filtering provides spectral shape cues at each ear individually, and the monaural localization paradigm, which typically requires normal hearing listeners to localize sound sources while one ear is plugged, is used as a way of studying how these monaural spectral cues are processed.

The monaural localization paradigm has some significant weaknesses that lead us to question the extent to which the results of such experiments can inform us about the mechanisms and processes that subserve sound localization in normal binaural conditions. The first problem is that complete "monauralization" of a listener is difficult to achieve, and this leads to the choice of very low stimulus levels (20–30 dB SL) in most monaural localization studies. While it is difficult to know the amount of attenuation provided by the typical "plug and muff" used in monaural localization studies, it is almost certainly lowest in the low frequencies. Given the documented importance of low frequencies for determining the extent to which listeners rely on ITD cues (Wightman and Kistler, 1992), a small amount of low-frequency energy leaking through the "plug and muff" could complicate interpretation of the results considerably. An additional complication is that sound will reach the occluded ear via bone conduction, and while the bone-conducted components would be more than 45 dB below the air-conducted sound at all frequencies (e.g., Hood, 1962), they cannot be ignored at high stimulus levels. Also, given the importance of both the low frequencies (for ITD coding) and the high frequencies (where monaural spectral cues are represented), the use of very low overall stimulus levels to circumvent the leakage and bone conduction issues is problematic. If the stimulus is wideband, its threshold would be determined primarily by the mid frequencies, where the auditory system is most sensitive. Thus, an overall stimulus level of 30 dB SL would limit the availability of cues at low and high frequencies, since these frequencies would be close to or below threshold. The second problem is that while monauralization is usually described as "removing" ITD and ILD cues, thus forcing listeners to attend to monaural spectral cues, it is probably more accurate to say that monauralization produces very unnatural ITD and ILD cues. Plugging one ear obviously causes a large ILD. It is an unnatural localization cue because the pattern of ILD across frequency produced by a plug is very different from that produced by a real sound source at any position in auditory space. The effect of monauralization on ITD is less obvious, but it seems just as appropriate to describe it as producing an infinite ITD as to say it removes ITD altogether. In any case, the result of monauralization is a situation in which the monaural spectral cues are usually in conflict with (i.e., signal different spatial positions) one or both of the interaural difference cues. Whether or not listeners will attend to the former and disregard the latter may depend on other factors such as task variables (range of stimulus positions and response alternatives, experience, expectation, context) or stimulus frequency content (bandwidth, low-frequency content, trial-to-trial spectral uncertainty). A third complicating factor in previous studies of monaural localization is the frequent emphasis on localization *accuracy*, typically measured by the extent to which a listener successfully identifies the specific loudspeaker in a small set of loudspeakers that actually produced the stimulus. Localization accuracy can be a useful metric, but in some conditions, monaural listening being one of them, reporting accuracy alone conceals large perceptual or response biases. For example, it is often reported that monaural localization accuracy is high for sources directly opposite the functioning ear and low for sources in front or behind. While true, the statement obscures the fact that the apparent origin of nearly all sounds heard monaurally is pulled strongly toward the unoccluded ear. Thus, accurate localization on the unoccluded side may be little more than an epiphenomenon produced by the large perceptual bias. Because of the issues raised above, it is difficult to interpret the results of many previous monaural localization studies as reflecting the salience of monaural spectral cues in normal binaural localization.

Monaural spectral cues are produced by the directionality of pinna filtering. Since the characteristics of pinna filtering change dramatically with changes in source position, those characteristics could potentially serve as cues (monaural spectral cues) to source position. The viability of monaural spectral cues depends on a listener's ability both to extract the pinna filtering characteristics from an incoming sound and to associate those characteristics with the appropriate source position. The latter process is usually thought to involve some form of comparison between the extracted pinna characteristics and a set of templates or feature lists stored in memory (e.g., Middlebrooks, 1992). Whether the stored representations of pinna characteristics are built up through experience or hard wired in the neural circuitry is not of concern here. However, there is ample evidence for the existence of some kind of stored representation that links apparent sound position and pinna characteristics.

Extraction of pinna filtering characteristics from an incoming sound requires knowledge of the spectrum of the sound source. The spectrum of a sound at the eardrum is the product of the pinna filter and the source spectrum. Thus, the only way a listener could deconvolve the two in order to process the characteristics of the pinna filter is by knowing the spectrum of the source. It is clearly unreasonable to postulate that listeners know, in any precise sense, the spectra of all potential sounds. However, it may not be unreasonable to suggest that laboratory experiments which require listeners to localize a noise burst or click, the spectrum of which is simple and constant for many trials, may offer listeners an opportunity to learn the source spectrum. In everyday life, when the source spectrum is uncertain and highly variable, listeners may make certain assumptions about the source spectrum in order to accomplish the deconvolution.

A large body of work on monaural localization shows that under certain circumstances information about sound source position is extracted from the sound at one ear. This clearly suggests that the auditory system is deconvolving from the sound transduced at the eardrum the separate contributions of the sound source and the pinna filtering. Whether the deconvolution is based on assumptions about or prior knowledge of source characteristics is unclear. Some studies, such as those in which narrow bands of noise were used as the stimulus (Belendiuk and Butler, 1977; Butler and Flannery, 1980; Flannery and Butler, 1981; Musicant and Butler, 1984, 1985; Butler, 1986), suggest that assumptions are made about the source spectrum. Others, such as those in which a white noise was the stimulus. (Oldfield and Parker, 1986; Butler *et al.*, 1990), are inconclusive, since white noise, which has a flat spectrum with minimal trial-to-trial spectral uncertainty, may allow listeners to learn the source spectrum. The fact that some process like deconvolution can occur to extract monaural spectral cues is an important result that emerges from past work on monaural sound localization.

Another important finding contributed by previous monaural localization experiments is that in certain conditions some or all features of monaural localization are nearly normal, as if the listener was binaural. For example, when the sound source is on the side of the functioning ear, the elevation component of the apparent position is near normal (Oldfield and Parker, 1986; Butler *et al.*, 1990; Slattery and Middlebrooks, 1994). With long-term experience, some monaural listeners, such as the unilaterally deaf listeners studied by Slattery and Middlebrooks (1994), demonstrate near normal localization in both azimuth and elevation components and not only on the side of the functioning ear, but on the side of the occluded ear as well. Clearly these listeners have learned sophisticated strategies for extracting and processing the monaural spectral cues.

The research described here revisits the monaural localization paradigm. Our purpose is not only to address some of the problems with the earlier work, but also to use the monaural paradigm to learn more about how the monaural cues contribute to normal binaural localization. The hallmark of our approach is the use of the virtual sources, i.e., sounds presented over headphones that include nearly all of the spatial attributes of sounds presented in free field and that evoke realistic, externalized spatial percepts (Wightman and Kistler, 1989a, b). The use of virtual sources provides considerably more interaural attenuation than a plug for monaural presentation (see below for data on this point) and allows for stimulus configurations not possible with real sources. The experiments described below will exploit these advantages.

Three experiments are described. The first measures the apparent positions of both real and virtual sources in monaural listening conditions. This is essentially a replication of previous work, with the added feature that in the virtual source conditions monaural stimuli are intermingled with binaural stimuli in an attempt to promote natural binaural localization strategies. The second experiment explores the influence of the spectral uncertainty of the stimulus to be localized in monaural listening conditions. The rationale is that some degree of spectral uncertainty is always present in everyday listening conditions, and this spectral uncertainty must interfere with a listener's ability to extract monaural spectral cues. The third experiment examines the influence on apparent position judgments of increasing amounts of unilateral attenuation. The aim of this experiment is to better understand the effects of various degrees of monauralization (such as obtained with a plug or over headphones).

I. GENERAL METHOD

A. Listeners

University of Wisconsin students participated as paid listeners in these experiments. Selection criteria consisted of normal hearing (as verified by complete audiometric exam), clean ear canals, and willingness to participate for 4–6 h per week for at least a semester. There were different numbers of listeners in each experiment; not all listeners participated in all three experiments. Most of the listeners were experienced, having participated in other localization experiments conducted in this laboratory.

B. Stimuli

In order to produce the virtual sources, a set of head-related transfer functions was measured on each listener. The measurement procedure was nearly identical to that described by Wightman and Kistler (1989a); the reader is referred to the earlier article for complete details. In short, a small (1-mm-diam) probe tube was held in position close to the listener's eardrum, while a wideband periodic noise test stimulus was presented from a loudspeaker. A microphone connected to the probe tube recorded the response to the test stimulus and a computer averaged the responses to multiple periods to improve signal-to-noise ratio. The two ears were measured simultaneously and the HRTFs from 266 source positions (roughly evenly spaced on the sphere, at 15° azimuth intervals all around the listener and at 12° elevation intervals from −48° to +72° relative to the horizontal plane) were measured during a single session. The transfer characteristics of the headphones used in the experiments (Sennheiser HD430) were measured in a similar way on each listener.

The procedures used to produce the virtual sources used as stimuli in these experiments were identical to those described in a previous publication (Wightman and Kistler, 1989a), so they will only be summarized here. A virtual source is synthesized by passing the desired stimulus (in these experiments a noise burst) through a pair of digital filters. Each digital filter consists primarily of the listener's own HRTF for the desired source position and ear divided by the headphone characteristic for that listener and ear. The result is two stimulus waveforms, one for each ear, which when presented simultaneously to the listener over the headphones produce an externalized sound image at an apparent spatial position very close to that which would have been produced by the comparable free-field source (Wightman and Kistler, 1989b).

The basic stimulus in all the experiments was a 250-ms noise burst with 20-ms cosine-squared on/off ramps. The noise was bandpassed between 200 Hz and 14 kHz and in the passband its spectrum was either flat or "scrambled." The scrambled spectrum was produced by randomizing the noise spectrum level within each critical band from trial-to-trial (uniform distribution, 20-dB or 40-dB range). Thus, in the case of 20-dB scrambling, adjacent critical bands could differ in level by as much as 20 dB. The noise burst was repeated four times on each trial with 300 ms of silence between the bursts. In free-field conditions the stimulus was presented from one of 12 small loudspeakers (Realistic Minimus 3.5) mounted on a vertical semicircular arc (as described in Wightman and Kistler, 1989b), at 12° elevation intervals. Since the arc could be rotated around the listener, the free-field stimulus could be presented from any azimuth and from one of 12 elevations ranging from −48 to +72. In virtual source conditions the stimulus was presented over Sennheiser headphones (HD 430). The overall stimulus level in both free-field and virtual source conditions was approximately 70 dB SPL.

C. Procedure

In free-field conditions listeners were seated in the anechoic chamber with their heads at the center of the loudspeaker arc and asked to keep their heads as still as possible. In the virtual source conditions, listeners were tested either in the anechoic chamber or in an IAC soundproof chamber. In both conditions the listeners were blindfolded. The task required listeners to report verbally, using standard spherical coordinates, the apparent azimuth and elevation (in degrees) of each stimulus immediately following the four noise bursts. In some, but not all, of the virtual source conditions, a few listeners were also asked to report apparent source distance in feet. A short training session was used to familiarize the listeners with the coordinate system. This session was conducted informally outside the anechoic chamber and included visual and auditory cues and feedback. Following the familiarization session, listeners were given about 6 h of experience listening and responding to free-field stimuli before any data were taken. No feedback was given either during this "training" phase or during any of the experimental conditions. A single 1.5-h session typically involved 180 trials, presented in 5 blocks of 36. All stimulus conditions were constant during a block, but could be changed between blocks. In many of the virtual source conditions, the test stimuli, which were either monaural or otherwise abnormal, were "interlaced" with normal stimuli. The normal stimuli were virtual sources with all the natural localization cues intact. The interlacing was random so that on any one trial there was a 0.5 probability of a test stimulus and a 0.5 probability of a normal stimulus being presented. At least eight blocks of trials were completed by each listener in each test condition. In conditions in which test stimuli were interlaced with normal stimuli, 16 blocks per condition were completed.

D. Data analysis

Data from localization experiments frequently include substantial numbers of what have come to be known as "front–back" confusions. These are responses indicating a perceived position in the front hemifield (azimuths between −90 on the left and +90 on the right) for a rear hemifield (azimuths from −90 to −180 on the left and between 90 and 180 on the right) target position. Given the roughly conical symmetry of the ITD cue such confusions are not entirely unexpected (cf. the "cone of confusion" described in Mills, 1972). However, the rate of front–back confusions varies considerably from listener to listener and from condition to condition (Wightman and Kistler, 1989b, Makous and Middlebrooks, 1990), and it is often difficult to distinguish between confusions and true error variance (e.g., for target positions near +90° and −90° azimuth). Consequently, analysis of apparent position data is problematic. Our choice is to avoid measures of central tendency and variability (which would be inappropriate with bimodal response distributions) and to restrict analysis of the data to the descriptive level. Thus, we display the raw data and draw conclusions on the basis of the appearance of those displays.

Data are displayed, condition by condition and listener by listener, on a three-pole coordinate system (Kistler and Wightman, 1992). Thus, each individual response is represented by a point on three separate graphs. The azimuth component of the response is decomposed into a "left–right" component and a "front–back" component, each expressed in degrees and plotted in separate graphs. The left–right component is the angle between the judgment vector and the median plane, and the front–back component is the angle between the judgment vector and the transverse plane (the vertical plane that goes through the ears). The elevation component of each response is plotted untransformed and is called the up–down component. In this coordinate system the extremes on each of the three dimensions are represented similarly, by angles of +90° and −90°.

II. EXPERIMENT 1: MONAURAL LOCALIZATION OF REAL AND VIRTUAL SOURCES

The general conclusion of all recent studies of monaural localization is that apparent azimuth is dramatically affected by the monauralization and apparent elevation is less affected (Oldfield and Parker, 1986; Butler *et al.*, 1990; Slattery and Middlebrooks, 1994). Apparent azimuth is pulled strongly toward a position directly opposite the open ear. While only one of these experiments (Slattery and Middlebrooks, 1994) evaluated the effect of monauralization on apparent positions of sources on the occluded side, the azimuth effect there was the same as on the unoccluded side. Our first experiment constitutes a replication of the essential features of the previous studies with both free-field and virtual sources.

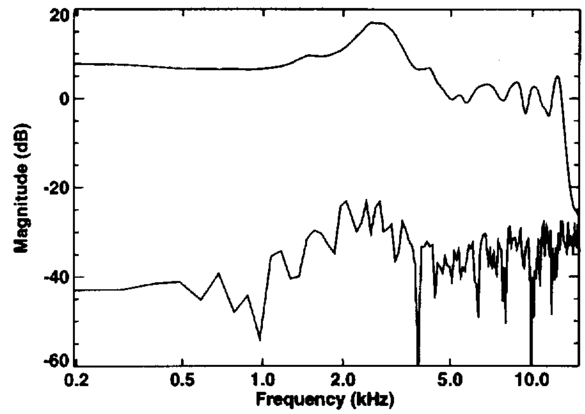

FIG. 1. The magnitude spectrum of a virtual monaural flat-spectrum noise stimulus measured in the ear canals of a representative listener. The upper curve shows the measurement made in the stimulated ear and the lower curve shows the measurement made in the nonstimulated ear. The nonstimulated ear measurements are obviously corrupted by the noise level of the measuring system (and sound room) which had a spectrum level of approximately −40 dB on this scale.

A. Method

All general aspects of stimulus generation and presentation and listener response were as described above. There were six conditions in this experiment, four free-field conditions involving real sources, and two virtual source condition. Of the four free-field conditions, two involved binaural listening, one at an overall stimulus level of approximately 70 dB SPL, and one at an overall level 40 dB lower. The other two free-field conditions required listeners to localize with the left ear occluded. Occlusion was accomplished in the usual way by plugging the ear with an EAR compressible foam plug, and covering it with a muff (EAR NRR26). Stimuli for the two "monaural" free-field conditions were at the same levels (about 70 dB SPL and 40 dB lower) as in the comparable binaural conditions. The two virtual source conditions involved binaural and monaural stimulus presentation (achieved by disconnecting the left headphone) at an overall level of approximately 70 dB SPL. In the monaural virtual source condition, monaural stimuli were interlaced with binaural stimuli as described above.

The monaural virtual source condition achieves excellent isolation of the nonstimulated ear, probably better than is possible with any plug-muff combination in the free field. Figure 1 shows measurements of ear canal sound pressure produced by a flat-spectrum noise stimulus in both the stimulated and nonstimulated ears of a typical listener in this experiment. Note that even at low frequencies the isolation exceeds 50 dB. This analysis does not consider the influence of bone-conducted sound which would effectively reduce the

TABLE I. Listener participation in the test conditions of experiment 1.

Conditions			Listeners									
		Level dB SPL	SDL	SDO	SDP	SER	SET	SGE	SGG	SHD	SHG	SIK
Free field	Binaural	70	X	X	X	X	X	X	X	X	X	X
Free field	Binaural	30	X					X	X	X		X
Free field	Monaural	70	X	X	X			X	X	X	X	X
Free field	Monaural	30	X		X			X	X	X	X	X
Virtual	Binaural	70	X	X	X	X	X	X	X	X	X	X
Virtual	Monaural	70	X	X	X	X	X	X	X	X	X	X

isolation at the lowest frequencies to about 40 dB (Hood, 1962). Thus, with the 70 dB SPL stimulus, which would have a spectrum level of less than 30 dB, the level in the nonstimulated ear is close to or below threshold at all frequencies. The plug-muff combination conventionally used to monauralize listeners cannot be expected to produce the same degree of isolation, especially at low frequencies.

The spectra of the noise-burst stimuli in this experiment were scrambled in an effort to approximate the spectral uncertainty typical of everyday listening. The scrambling was as described previously (Wightman and Kistler, 1989b). In this experiment the level in each critical band was randomized (from trial to trial) within a 20-dB range. The potential effect of this spectral scrambling on monaural localization is the subject of experiment 2.

Because the experiment was conducted over a long period of time, not all listeners participated in all conditions. However, we feel that enough listeners participated in each condition to represent the full range of individual differences we observed. Ten listeners in all were tested and Table I lists the conditions in which each listener participated. Six of the ten listeners contributed distance judgments in the virtual source conditions. No distance judgments were obtained in the free-field conditions since these were run before distance reporting was implemented.

B. Results

The data from the high level (70 dB SPL) binaural conditions are unremarkable, and data from comparable conditions have been described before (Wightman and Kistler, 1989b). The apparent positions of free-field sources match their actual positions reasonably well, with the exception of a few front–back confusions and greater variance in the up–down dimension than in the other two dimensions. The results from the virtual source condition are nearly identical to those from the free-field condition, attesting to the adequacy of the simulation. Figure 2 shows the results from a typical listener in the high-level binaural conditions.

The high level monaural conditions produced several intriguing results. In contrast to the binaural conditions, the monaural conditions revealed considerable individual differences. Figures 3 and 4 show data from two listeners (SGG and SIK, respectively) that represent the range of performance we obtained from the listeners who participated in these conditions. Note that in the case of the monaural free-field condition, responses to sources on the side of the open ear are plotted separately from the responses to the sources on the occluded side.

It is clear that, in general, localization as reflected by the match between target and response position is degraded in the monaural condition. One obvious effect is that the variance of the responses is much larger in the monaural conditions. A second is that in many cases the responses do not cluster along the major diagonal. In these cases there is little correspondence between target and response positions. In the free-field conditions, this is true primarily for listener SGG (Fig. 3). Note that for this listener, on both the open and

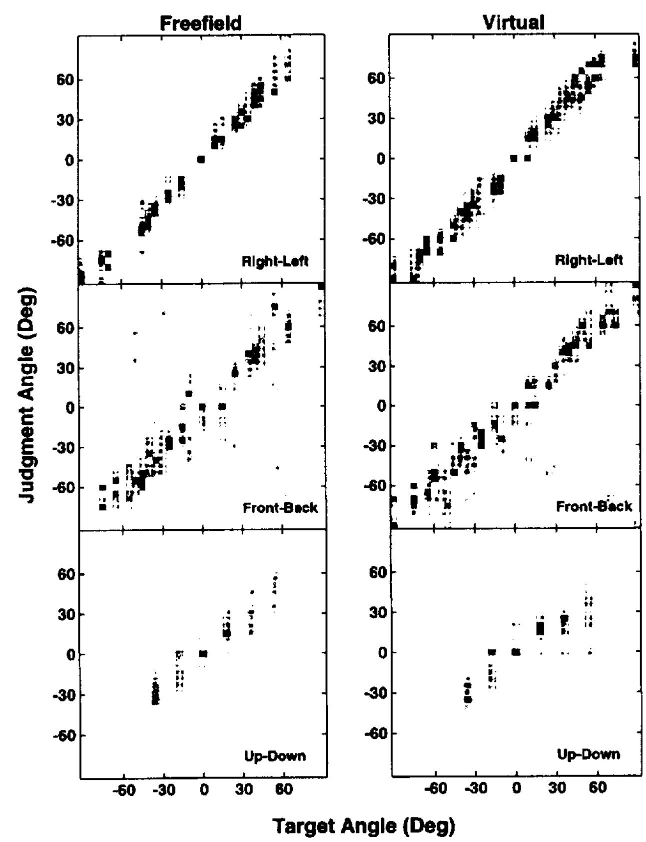

FIG. 2. Scatter plots of judged direction versus target direction from a typical listener in the high-level binaural condition of experiment 1. As described in the text, the judged and target directions are represented in terms of three angles, right–left, front–back, and up–down. Each data cell includes all the judgments within a 5° wide interval. The darker the cell, the more judgments represented in that cell. The lightest cells represent a single judgment. There are at least 288 judgments shown in each panel.

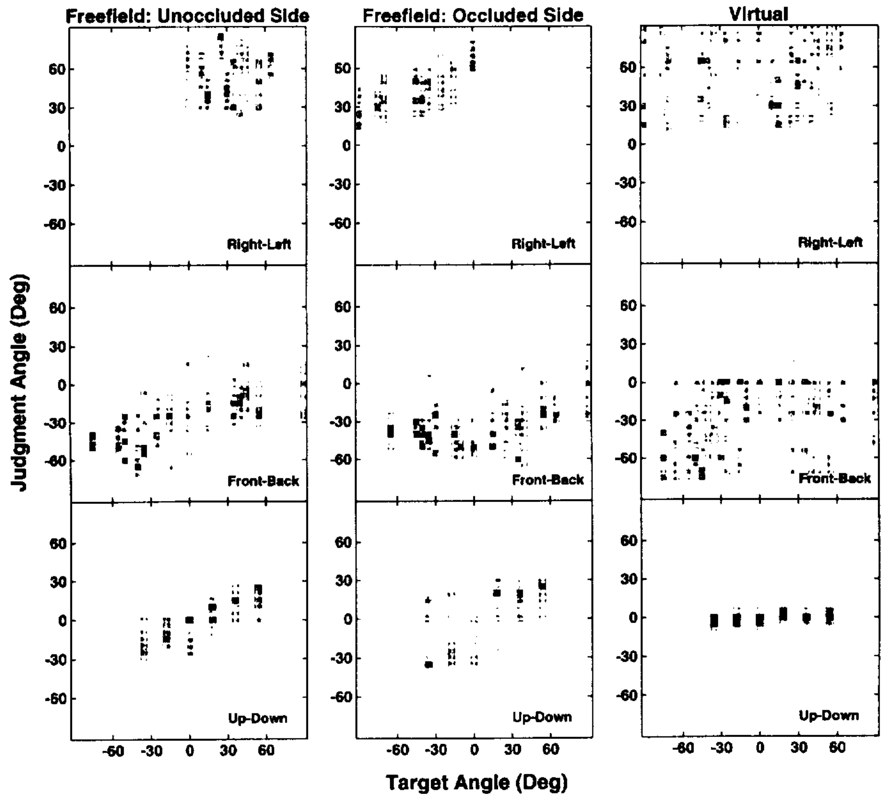

FIG. 3. Scatter plots of data from listener SGG in the high-level monaural conditions. Note that in the case of stimuli presented in free field, responses to stimuli on the occluded and unoccluded sides are plotted separately.

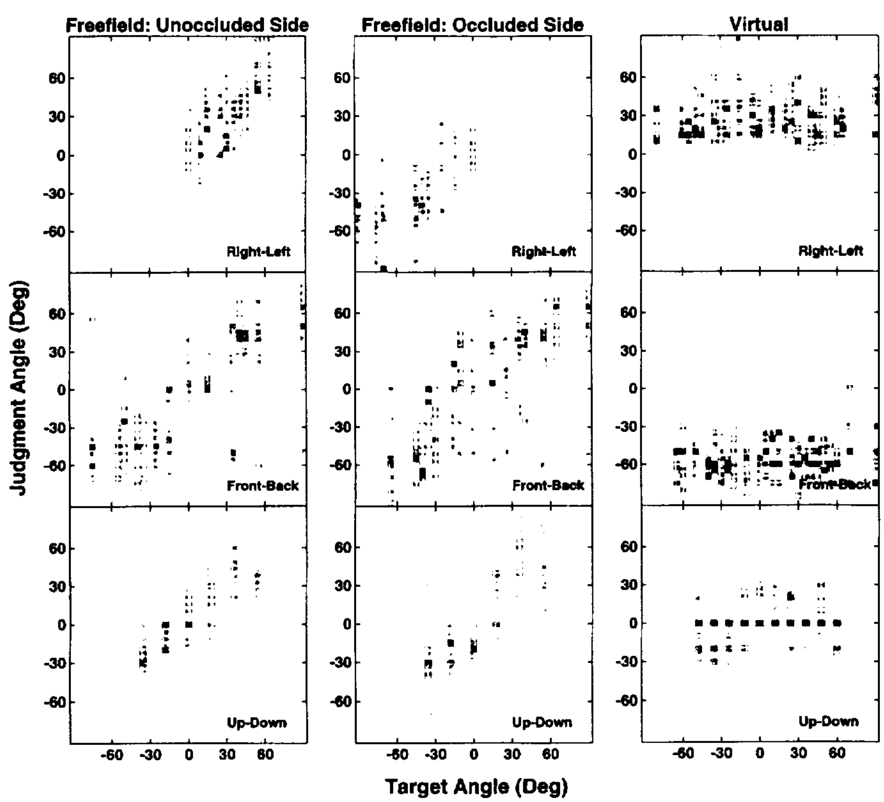

FIG. 4. Same as Fig. 3, but data are from listener SIK.

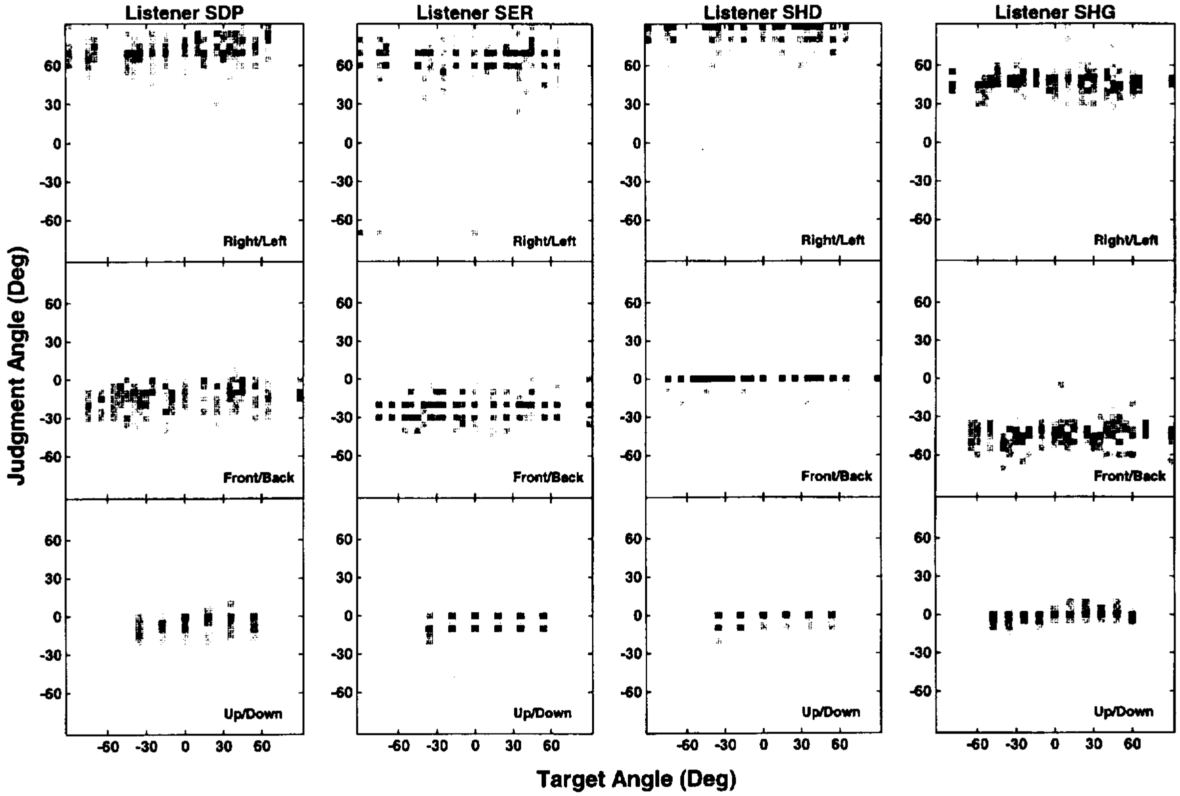

FIG. 5. Scatter plots of data from four additional listeners in the monaural virtual source condition.

occluded side, the responses, while showing the usual bias toward the open side and a hint of up–down perception on the open side, appear randomly distributed around a position roughly centered on the front–back and up–down dimensions. In the case of listener SIK (Fig. 4), however, responses to stimuli on the open ear side suggest nearly normal localization (although with increased variance in the responses). Even more remarkable is this listener's responses to stimuli on the occluded side. Not only are the up–down and front–back components of the responses nearly the same as on the open side, but the left–right components of the responses do not show the usual bias toward the open side (right, or positive angles on the left–right dimension). The reason for this is almost certainly inadequate "monauralization" by the plug and muff, an issue that will be discussed in connection with the results from the low-level free-filed conditions.

The apparent position judgments from the monaural virtual source condition are quite different than those from the comparable free-field condition. For both of the listeners whose data are shown in Figs. 3 (SGG) and 4 (SIK), responses to all stimuli are more or less randomly distributed on the side of the stimulated ear (positive angles on the left–right dimension), toward the rear of the interaural axis (negative angles on the front–back dimension), and more or less close to zero elevation (zero angle on the up–down dimension). Thus, we conclude that localization is essentially abolished in the monaural virtual source condition. Distance judgments were obtained from both SGG and SIK in the virtual source conditions. In the binaural virtual source condition (data not shown) the mean source distance reported by SGG was 4.5 ft (s.d.=0.8) and by SIK it was 3.9 ft (s.d. =0.7). In the monaural virtual source condition (Figs. 3 and 4), SGG reported a mean source distance of 2.6 ft (s.d.=0.9), and SIK reported a mean distance of 2.7 ft. (s.d.=1.6). Thus, even though the monaural virtual sources were not localizable, they were apparently externalized by these listeners.

Not all listeners in the monaural virtual source condition distributed their azimuth judgments as widely as those shown in Figs. 3 and 4. In fact, a more typical pattern was a tight clustering of judgments around a single azimuth. To illustrate this trend the data from four additional listeners in the monaural virtual source condition are shown in Fig. 5. The overall conclusion that localization is abolished in the monaural virtual source condition is the same for these listeners as for those whose data are shown in Figs. 3 and 4. Distance judgments are available for three of these four listeners (all but SER) and confirm that all monaural virtual sources were externalized. The mean reported source distances were 3.1, 28.5, and 0.5 ft (s.d=1.7, 18.0, and 1.5) for SDP, SHD, and SHG, respectively.

The results from the low-level free-field conditions reveal the inadequacy of the plug and muff in achieving effective monauralization. Figures 6 and 7 show the data from two listeners, SGG and SIK, respectively, whose data from the high-level free-field condition were displayed in Figs. 3 and 4. Note first that the apparent position judgments in the low-level binaural condition are nearly identical to the judgments in the high-level binaural condition (cf. Figs. 2 and 6, both from listener SGG). While this comparison is shown for only one listener, the data from all the other listeners are consistent with this observation. Also note that for listener SGG, the monaural judgments are the same at low and high

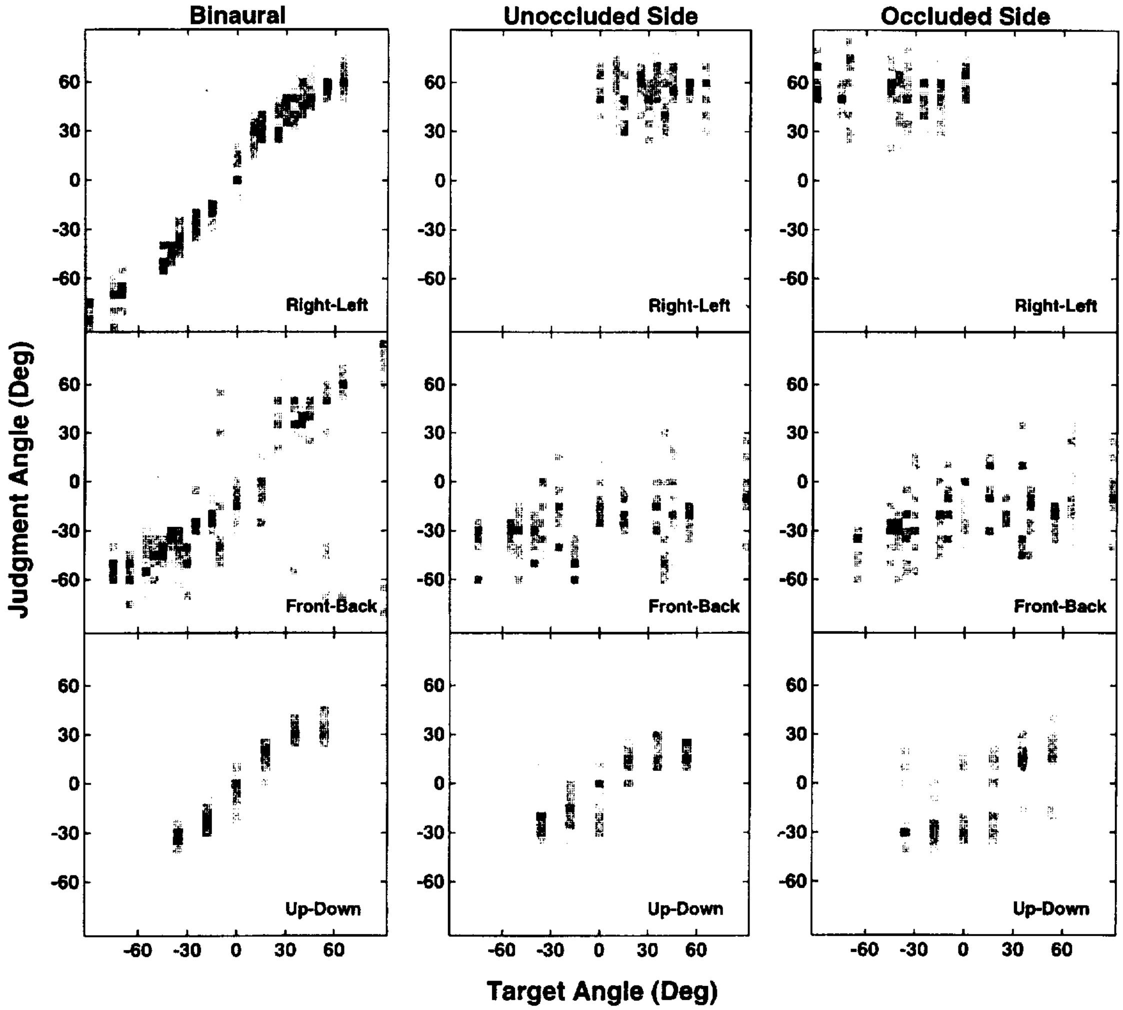

FIG. 6. Scatter plots of data from listener SGG in the low-level free-field conditions.

levels for stimuli on both the open and occluded side (cf. Figs. 3 and 6). However, for listener SIK, whose monaural judgments in the high-level condition suggested near normal localization (Fig. 4), the reduction in level had a dramatic effect. While this listener's judgments to stimuli on the open side are about the same at the two levels, the responses to stimuli on the occluded side are completely different at the lower level: Azimuth is strongly biased toward the open side, and elevation is nearly eliminated (clustered around 0°). We interpret this result as suggesting that for some listeners the plug and muff typically used to monauralize listeners may not be completely effective in preventing stimulation of the occluded ear. This in turn would allow the listener to use some interaural cues, most likely low-frequency ITDs. Of course, if the stimulus had not contained low frequencies, as was the case in the experiment reported by Slattery and Middlebrooks (1994), the consequences of inadequate interaural attenuation would probably have been quite different.

C. Discussion

The results of this experiment led us to two conclusions which we feel are important. One is that interpretation of the results of experiments in which listeners are monauralized by using an ear plug and muff must take into account the amount and frequency dependency of the attenuation produced by the plug and muff. In the case of localization studies, inadequate attenuation forces investigators to present stimuli at very low levels. At these low levels, the accessibility of spectral cues may depend critically on stimulus spectral content, spectral variability, and sensitivity of the listener at high frequencies. A second conclusion is that monaurally presented virtual sources are not localizable. Whatever differences exist between free-field and virtual sources seem to be magnified in the monaural condition. While even at low levels there is some hint of localizability for monaural free-field sources, a monaural virtual source cannot be localized.

III. EXPERIMENT 2: INFLUENCE OF SPECTRAL UNCERTAINTY ON THE SALIENCE OF MONAURAL SPECTRAL CUES

The spectrum of a sound at each eardrum is the product of the pinna filtering and the spectrum of the sound source itself. The only way the two components of the product can be deconvolved, to extract the spectral cue produced by pinna filtering, is through knowledge of the spectrum of the source. The clearest evidence of the importance of prior knowledge of the stimulus spectrum comes from research on

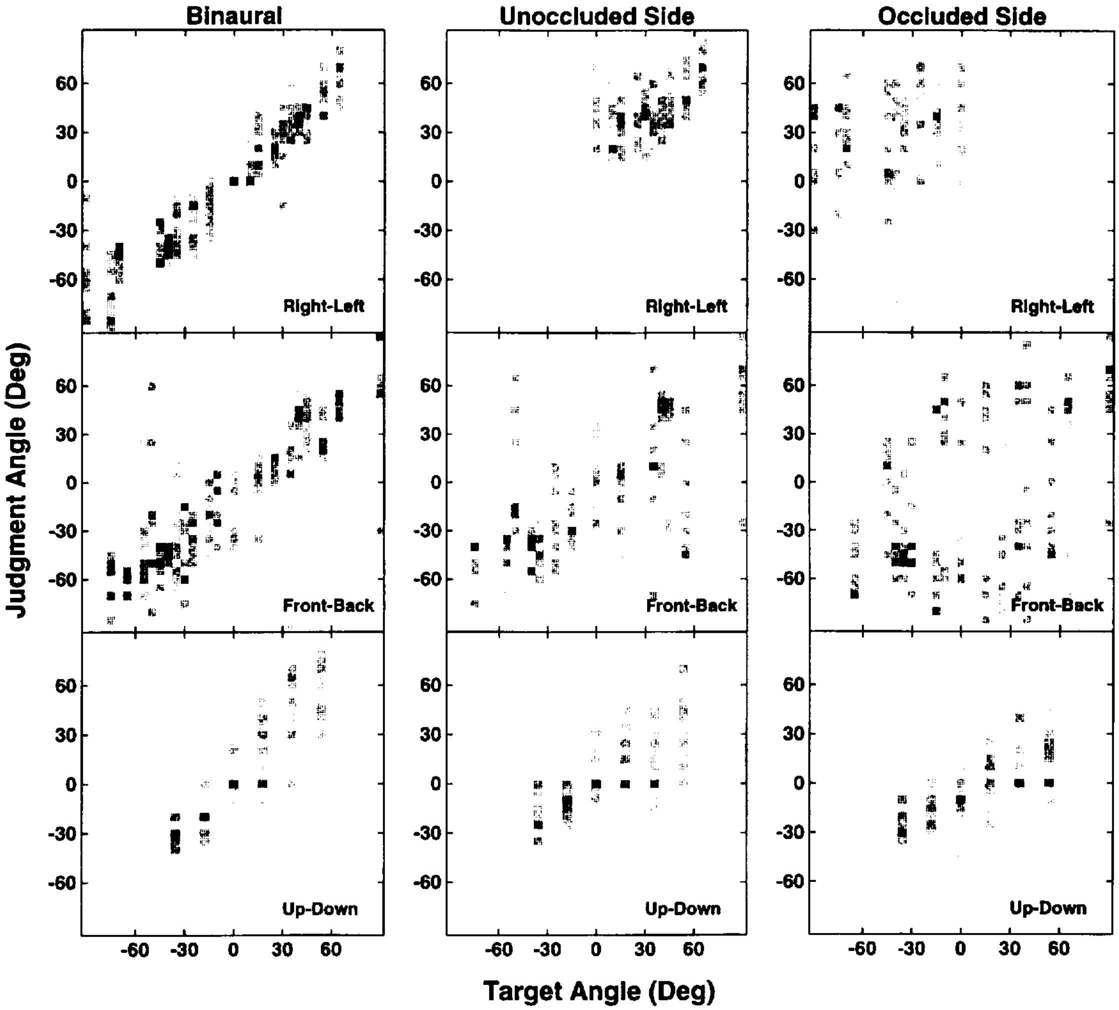

FIG. 7. Same as Fig. 6, but data are from listener SIK.

the apparent positions of narrow-band sounds (e.g., Blauert, 1969; Middlebrooks, 1992). Results from such studies consistently show that the apparent position of a narrow-band sound, especially its apparent elevation, is determined primarily by its center frequency and not by its actual position. The apparent position is one at which the pinna filter has a prominent peak at that frequency (Middlebrooks, 1992). These results imply that listeners know the characteristics of their own pinna filters and that they assume the spectrum of an incoming stimulus is relatively flat.

Many of the studies that demonstrate the importance of monaural spectral cues have used stimuli with spectra which were both relatively smooth over a broad frequency range and unchanging from presentation to presentation. It is possible that these conditions are optimal for extraction of monaural spectral cues, since the stimulus spectrum can be considered "known" to the listener and since it has no prominent spectral peaks or valleys. In more realistic conditions listeners encounter numerous stimuli which have nonflat spectra and must deal with considerable uncertainty about the stimulus spectrum. Both of these factors could interfere with the use of monaural spectral cues.

There has been very little research on the role of listeners' prior knowledge of or expectations about stimulus spectral characteristics. One study, reported by Hebrank and Wright (1974), showed that localization of flat-spectrum median plane sources was significantly degraded when random peaks and valleys were introduced into the sound spectra. The conclusion was that the uncertainty of the stimulus spectrum from trial-to-trial prevented extraction of monaural spectral cues.

A. Method

In this experiment the role of *a priori* knowledge of stimulus characteristics was studied by comparing listener's judgments of the apparent positions of real free-field sources with flat or randomly scrambled spectra in both monaural and binaural listening conditions (thus, four conditions in all). The essential features of the stimuli and experimental procedure were as described above. The stimulus level in this experiment was the same as the low level in experiment 1 (40 dB SPL). In the scrambled-spectrum conditions the range of randomization of critical band levels was 40 dB (it was 20 dB in experiment 1). Six listeners participated in this experiment. One listener, SIK, also participated in experiment 1.

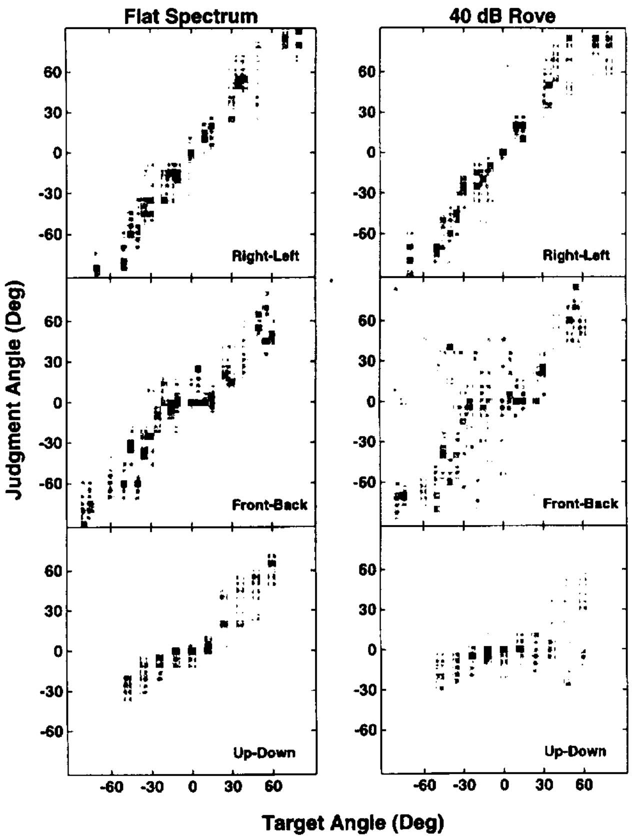

FIG. 8. Scatter plots of data from a typical listener in the flat- and scrambled-spectrum binaural conditions.

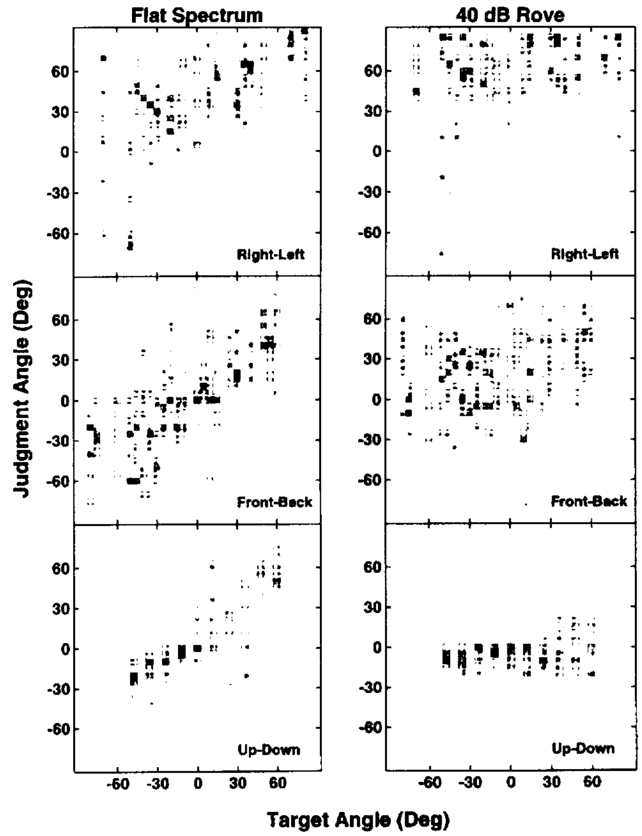

FIG. 9. Same as Fig. 8, but data are from the same listener in the flat- and scrambled-spectrum monaural conditions.

B. Results

Figure 8 shows the judgments from a typical listener in the flat and scrambled binaural conditions. Note that for this listener the effect of scrambling is to increase the number of front–back confusions (off-diagonal judgments in the "front-back" panels) and to degrade the perception of apparent elevation. Both of these effects are indicative of the way monaural spectral cues are being used. Scrambling the stimulus spectrum would presumably reduce the effectiveness of monaural spectral cues, since listeners would be unable to learn the spectral characteristics of the stimuli. The effects of scrambling are greater for some listeners than for others and thus may reflect the extent to which each listener relies on monaural spectral cues.

Figure 9 shows the judgments from the same listener in the flat and scrambled monaural conditions. In the flat condition, note that the most dramatic effect of the monauralization is in the right–left component of the judgment. It is clear that nearly all the stimuli were perceived to be on the right side (positive right–left *judgment* angles). Moreover, even when the source itself was on the right side, there was little correspondence between the target angle and the judgment angle. The consistent lateralization of monaural stimuli to the side of the unplugged ear agrees with many previous findings (Musicant and Butler, 1980; Hebrank and Wright, 1974; Butler *et al.*, 1990; Oldfield and Parker, 1986; Blauert, 1983; Butler, 1975) and probably reflects the perceptual salience of the large ILD caused by plugging one ear. The front–back and up–down components of the judgments were affected less by the monauralization of the listener, confirming the importance of monaural spectral cues for front–back and up–down perception. This result is also consistent with other monaural localization data in the literature (e.g., Oldfield and Parker, 1986; Butler *et al.*, 1990; Slattery and Middlebrooks, 1994). In the scrambled-spectrum condition both front–back and up–down perception is severely degraded, presumably because the monaural spectral cues have been rendered ineffective by the scrambling. Scrambling the spectrum over a smaller range (e.g., 20 dB, as in experiment 1) produces less severe disruption of free-field monaural localization, leaving the up–down components of the judgments only slightly degraded (cf. Figs. 6 and 7).

C. Discussion

The results of this experiment suggest that spectral uncertainty interferes with a listener's ability to extract monaural spectral cues. Thus, the results of experiments which present flat-spectrum stimuli, which are near optimal for extraction of monaural spectral cues, may not be generalizable to more typical listening conditions which include some uncertainty about the stimulus spectrum. This is especially important when considering experiments such as monaural localization experiments in which apparent position judgments may be unusually dependent on monaural spectral cues. Whether or not the 20-dB spectral scrambling we chose for experiment 1 more accurately represents natural spectral uncertainty is unknown. Unfortunately, it is difficult to estimate the extent of spectral uncertainty in everyday sounds. Since most everyday sounds are time variant, both analysis of their spectra and determination of which components determine their apparent position are complex problems with no obvious solutions.

IV. EXPERIMENT 3: LOCALIZATION WITH INTERAURAL LEVEL IMBALANCE

The results of experiment 1 suggest that small energy levels in a nominally occluded ear can have a dramatic impact on the apparent position judgments of some listeners. Because it is difficult to know precisely the attenuation characteristics of a plug and muff, and because those characteristics almost certainly are different for each listener, it is not possible to determine the extent to which our results and the results of previous studies of monaural localization might be affected. For this reason, and to aid a more complete understanding of the various factors which affect sound localization in natural listening situations, we carried out an experiment on the effects of various degrees of interaural intensity imbalance on sound localization.

A. Method

The only feasible way to conduct a localization experiment in which interaural level difference are varied is with the virtual source technique. Since the signals delivered to the two ears via headphones are essentially independent (see Fig. 1), independent control of the overall level at the two ears is straightforward. In this experiment, scrambled-spectrum (20-dB range) virtual sources were presented with the average overall level in the right ear set at approximately 70-dB SPL. In separate conditions the signal being delivered to the left ear was attenuated by 10, 20, 30, or 40 dB. In order to avoid problems of response bias, trials involving unilaterally attenuated virtual sources were interlaced (as in experiments 1) with trials involving "normal" virtual sources. The seven listeners who participated in this experiment had been tested in experiment 1, so data from both normal and monaural virtual source conditions were available for comparison.

B. Results

While the details in the patterns of responses were different for each of the seven listeners, the general trend was the same for all. Therefore, only the data from one listener will be shown here. Figure 10 shows the judgments from this listener in the binaural and monaural virtual source conditions (data from experiment 1). Note that the binaural data from this listener are normal, and that the monaural data suggest a complete elimination of normal localization. Regardless of nominal target position, the apparent positions of all stimuli are concentrated at 90° azimuth (90° left–right and 0° front–back) and 0° elevation, directly opposite the stimulated ear. This is the typical pattern of judgments in the monaural virtual source condition, as discussed above in connection with experiment 1. Figure 11 shows the data from the same listener in the conditions in which the signal to one ear was attenuated. Note a unilateral level imbalance of as much as 40 dB results in a pattern of responses that is

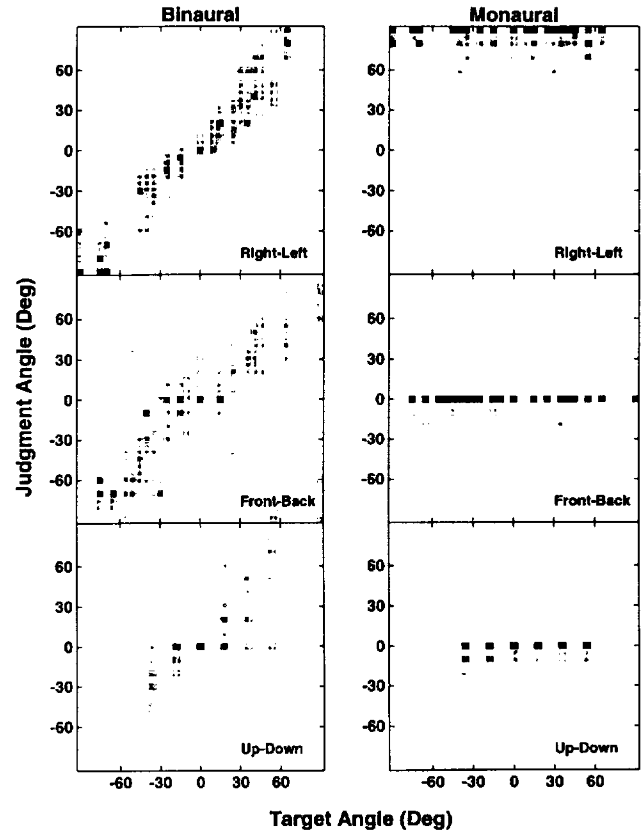

FIG. 10. Scatter plots of data from a representative listener in the binaural and monaural virtual source conditions of experiment 1.

clearly different from that obtained in the monaural condition, especially in the right–left components of the judgments. An imbalance of 10 dB produces a pattern of responses that is nearly "normal" (i.e., like that obtained with no unilateral attenuation).

The sensitivity to interaural level imbalance varied somewhat from listener to listener. For some, an imbalance of 40 dB was equivalent to the monaural condition, but for most it was not. Also, for some an imbalance of 10 dB had very little impact (as for the listener whose data are shown in Fig. 11), and for others the effect was more obvious. Finally, sensitivity to interaural level imbalance seems to be inversely correlated with the ability to extract interaural cues in the free-field monaural condition. Our data on this point are limited since not all listeners who participated in experiment 3 were also tested in experiment 1. However, a qualitative analysis of the available data suggests that those listeners whose localization judgments (especially the left–right components) were most accurate in the high-level free-field monaural condition of experiment 1 were among those whose judgments were least affected by interaural intensity imbalance. To the extent that monauralization with an earplug and muff is equivalent to creating an interaural level imbalance, such a correspondence would be expected.

C. Discussion

The results of this experiment are somewhat unexpected. Lateralization experiments, which involve presentation of

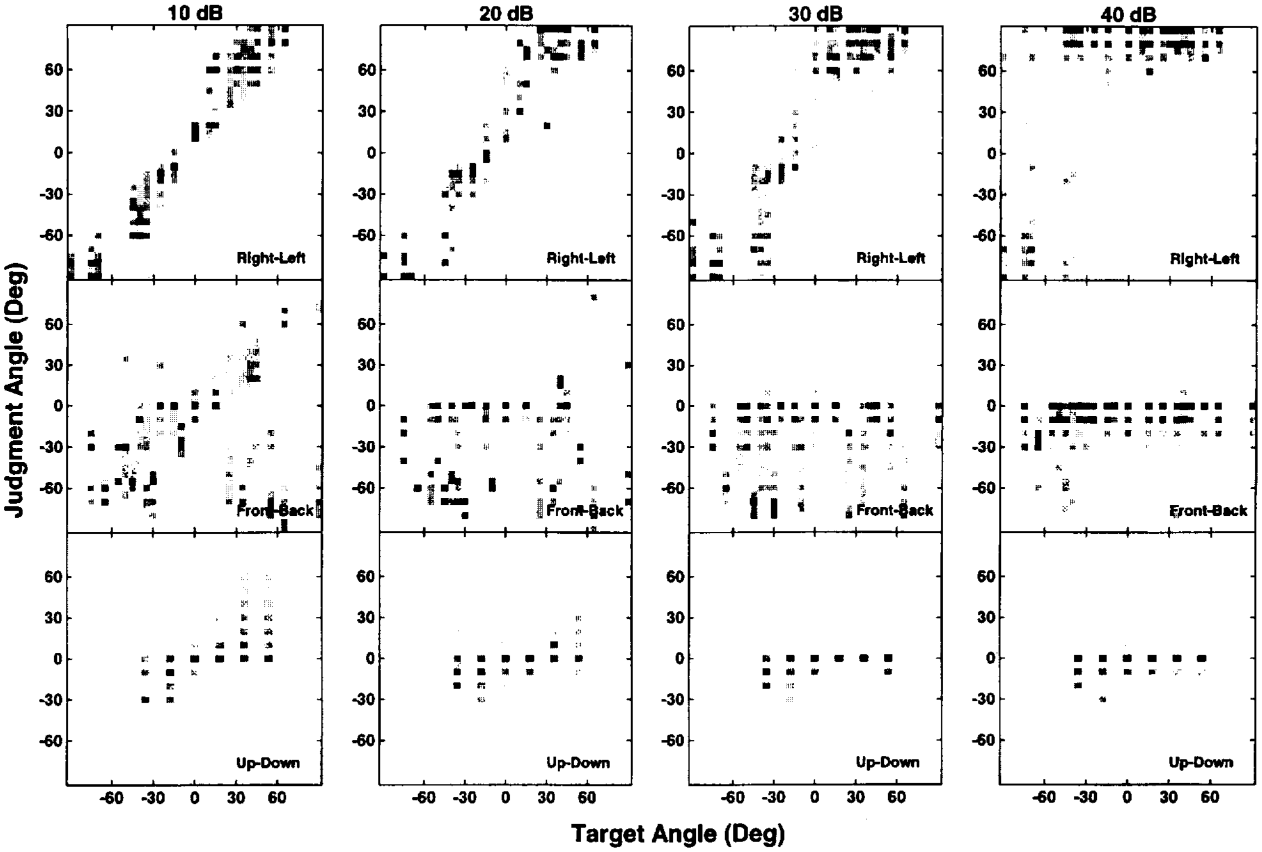

FIG. 11. Scatter plots of data from the same listener whose data are shown in Fig. 10 showing the effects of progressive attenuation of the signal at one ear. The amount of attenuation is indicated at the top of each panel.

nonspatialized stimuli (i.e., stimuli devoid of the spectral cues provided by pinna filtering), suggest that an interaural imbalance of 10–15 dB is sufficient to cause complete lateralization of the sound image, a shift of its apparent position all the way to one side (e.g., Yost and Hafter, 1987). In this experiment, with all the naturally occurring localization cues present, a 10-dB imbalance has a generally small effect on the apparent position of the sound image. However, this result is consistent with previous data on the apparent positions of virtual sources presented with conflicting localization cues (Wightman and Kistler, 1992). Those data suggest that with low frequencies present in the stimulus, as in the current experiment, the ITD cue was dominant, and the ILD and spectral cues were essentially ignored. If the effect of a 10–15 dB interaural level imbalance on the ITD cue is negligible, the results of the present experiment are less surprising.

The most important result from this experiment is the observation that even very low signal levels delivered to the attenuated ear can have a measurable influence on judgments of the apparent positions of virtual sources. This finding not only broadens our understanding of how the various localization cues are extracted and processed, but also complicates the interpretation of the results of many free-field monaural localization experiments.

V. GENERAL DISCUSSION AND CONCLUSIONS

Sound localization is a perceptual process that involves integration of several different types of information: auditory, visual, and cognitive, at least. The auditory substrate of sound localization derives from what we call the acoustical cues: ITD, ILD, and the spectral cues. Much is known about the auditory system's sensitivity to these cues and about how at least the interaural difference cues are extracted from acoustical stimuli. However, the extent to which, in any given situation, each contributes to the perception of the apparent position of a sound source is not well understood.

The monaural localization paradigm is thought to represent a situation in which the contributions of the spectral cues is emphasized. However, because monaural listening actually provides conflicting and unnatural cues to sound source position, one cannot be certain that a listener's judgments of apparent sound source position will reflect only the influence of spectral cues. Thus, interpretation of the results of monaural localization experiments strictly in terms relating to the use of spectral cues is not straightforward.

The three experiments reported here focus on various aspects of the monaural localization paradigm with the aims of clarifying the results of such experiments and increasing our general understanding of the processing of spectral cues. Experiments 1 and 3 deal with the consequences of the incomplete monauralization achieved by plugging a listener's ear. The results suggest that even very low stimulus levels in the occluded ear provide access to interaural cues for some listeners. Studies with the virtual source technique, which offers considerably improved monauralization, suggest that localization is essentially eliminated in monaural listening.

This result conflicts with the results of all monaural localization experiments (including our own) conducted in the free field, even those which used stimulus levels low enough to assure no stimulation of the occluded ear. The free-field experiments suggest that at least some residual localization, particularly in the up–down dimension, is maintained in monaural listening conditions. We will return to a discussion of this discrepancy shortly.

Experiment 2 is concerned with the role of prior knowledge of the stimulus spectrum. Extraction of reliable monaural spectral cues requires knowledge of the source spectrum. However, there is ample evidence from experiments involving narrow-band stimuli (e.g., Middlebrooks, 1992; Rogers and Butler, 1992) that, even without explicit information about source characteristics, the spectral shaping provided by an individual's own pinnae influences apparent position judgments. This suggests that listeners make certain assumptions about the source spectrum in order to extract the monaural spectral cue. The results of experiment 2, in which source spectrum is directly manipulated, suggest that extraction of monaural spectral cues is a process that, as expected, can be disrupted by uncertainty in the source spectrum.

There remains a curious discrepancy between the free-field results, which suggest that vertical localization is only moderately degraded by monauralization, and the virtual source results, which suggest that localization is effectively eliminated by monaural listening. There are two differences between free-field and virtual source conditions which we feel could be the source of this discrepancy. One is the fact that interlacing of monaural and binaural stimuli was done only in the virtual source conditions. It is possible that without the frequent exposure to normal binaural localization cues provided by the interlacing, listeners attended more to the available spectral cues. However, an informal pilot experiment which involved localization of monaural virtual sources without interlacing convinces us that this explanation is not correct. There were no differences between performance with and without the interlacing. The other major difference between listening in free field and listening to virtual sources lies in the consequences of small head movements. With the static, non-head-coupled virtual sources used here, head movements cause no change in the stimulus reaching the eardrum. In free-field conditions the eardrum stimulus is constantly changing since all listeners move their heads slightly when listening to the stimuli, even though they are asked to hold their heads still. We have monitored this movement with a magnetic head tracker and find that for some listeners the standard deviation of head azimuth during a stimulus presentation (which consists of four bursts of noise) is as large as 2°. Head movements of this magnitude, while small, could easily provide useable cues in the form of changes in the spectrum of the stimulus at the eardrum. At 6 kHz, for example, the frequency response of the outer ear changes at the rate of at least 0.25 dB/deg on the horizontal plane (Middlebrooks *et al.*, 1989). Since a 0.25-dB difference between spectra at high frequencies is detectable (Leshowitz, 1971), we conclude that very small head movements could produce detectable spectral changes, which could influence apparent position judgments in the free-field condition. That head movements cause no change in the stimulus in virtual source conditions is unnatural, and the effects of this, if any, on apparent position judgments are unknown.

In summary, while there can be no doubt about the importance of monaural spectral cues for sound localization, the monaural localization paradigm may not be the best means for studying their role. Problems of implementation and problems of interpretation greatly complicate the endeavor and argue for finding alternatives.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Marianne Arruda and Ramona Agrawal in the technical phases of the work, and the very helpful comments of Ewan Macpherson on an earlier version of this paper. Financial support was provided by NASA (Cooperative Agreement No. NCC2-542) and the NIH-NIDCD (DC00116).

Angell, J. R., and Fite, W. (**1901**). "The monaural localization of sound," Psychol. Rev. **8**, 225–246.

Belendiuk, K., and Butler, R. A. (**1977**). "Spectral cues which influence monaural localization in the horizontal plane," Percept. Psychophys. **22,** 353–358.

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

Blauert, J. (**1983**). *Spatial Hearing The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA), pp. 1–427.

Butler, R. A. (**1975**). "The influence of the external and middle ear on auditory discriminations," in *Handbook of Sensory Physiology,* edited by W. D. Keidel and W. D. Neff (Springer-Verlag, Berlin), pp. 247–260.

Butler, R. A. (**1986**). "The bandwidth effect on monaural and binaural localization," Hear. Res. **21**, 67–73.

Butler, R. A., and Flannery, R. (**1980**). "The spatial attributes of stimulus frequency and their role in monaural localization of sound in the horizontal plane," Percept. Psychophys. **28,** 449–457.

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

Flannery, R., and Butler, R. A. (**1981**). "Spectral cues provided by the pinna for monaural localization in the horizontal plane," Percept. Psychophys. **29,** 438–444.

Hebrank, J. H., and Wright, D. (**1974**). "Are two ears necessary for localization of sound sources on the median plane?," J. Acoust. Soc. Am. **56,** 935–938.

Hood, J. D. (**1962**). "Bone conduction: A review of the present position with especial reference to the contributions of Dr. Georg. Von Békésy," J. Acoust. Soc. Am. **34,** 1325–1332.

Kistler, D. J., and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91,** 1637–1647.

Leshowitz, B. (**1971**). "Measurement of the two-click threshold," J. Acoust. Soc. Am. **49,** 462–466.

Makous, J. C., and Middlebrooks, J. C. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87,** 2188–2200.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92,** 2607–2624.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86,** 89–108.

Mills, A. W. (**1972**). "Auditory localization," in *Foundations of Modern Auditory Theory–Volume II,* edited by J. V. Tobias (Academic, New York), pp. 301–345.

Musicant, A. D., and Butler, R. A. (**1980**). "Monaural localization: An analysis of practice effects," Percept. Psychophys. **28,** 236–240.

Musicant, A. D., and Butler, R. A. (**1984**). "The psychophysical basis of monaural localization," Hear. Res. **14**, 185–190.

Musicant, A. D., and Butler, R. A. (**1985**). "Influence of monaural spectral cues on binaural localization," J. Acoust. Soc. Am. **77,** 202–208.

Oldfield, S. R., and Parker, S. P. A. (**1986**). "Acuity of sound localization: A topography of auditory space. III. Monaural hearing conditions," Perception **15**, 67–81.

Rogers, M. E., and Butler, R. A. (**1992**). "The linkage between stimulus frequency and covert peak areas as it relates to monaural localization," Percept. Psychophys. **52,** 536–546.

Slattery III, W. H., and Middlebrooks, J. C. (**1994**). "Monaural sound localization: acute versus chronic unilateral impairment," Hear. Res. **75**, 38–46.

Strutt, J. W. (**1907**). "On our perception of sound direction," Philos. Mag. **13,** 214–232.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening II: psychophysical validation," J. Acoust. Soc. Am. **85,** 868–878.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening I: stimulus synthesis," J. Acoust. Soc. Am. **85,** 858–867.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91,** 1648–1661.

Yost, W. A., and Hafter, E. R. (**1987**). "Lateralization," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 49–84.