Discrimination and identification of azimuth using spectral shape[a)]

Daniel E. Shub[b)]
Speech and Hearing Bioscience and Technology Program, Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215

Suzanne P. Carr
Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215

Yunmi Kong
Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215 and Princeton University, Princeton, New Jersey 08544

H. Steven Colburn
Hearing Research Center, Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215 and Speech and Hearing Bioscience and Technology Program, Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Monaural measurements of minimum audible angle (MAA) (discrimination between two locations) and absolute identification (AI) of azimuthal locations in the frontal horizontal plane are reported. All experiments used roving-level fixed-spectral-shape stimuli processed with nonindividualized head-related transfer functions (HRTFs) to simulate the source locations. Listeners were instructed to maximize percent correct, and correct-answer feedback was provided after every trial. Measurements are reported for normal-hearing subjects, who listened with only one ear, and effectively monaural subjects, who had substantial unilateral hearing impairments (i.e., hearing losses greater than 60 dB) and listened with their normal ears. Both populations behaved similarly; the monaural experience of the unilaterally impaired listeners was not beneficial for these monaural localization tasks. Performance in the AI experiments was similar with both 7 and 13 source locations. The average root-mean-squared deviation between the virtual source location and the reported location was 35°, the average slopes of the best fitting line was 0.82, and the average bias was 2°. The best monaural MAAs were less than 5°. The MAAs were consistent with a theoretical analysis of the HRTFs, which suggests that monaural azimuthal discrimination is related to spectral-shape discrimination. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2981634]

I. INTRODUCTION

Research on azimuthal localization, especially for sources located in the frontal portion of the horizontal plane, has traditionally focused on binaural performance where the most important information is carried by the interaural differences in time delay (ITD) and level (ILD). When interaural comparisons are not possible (e.g., due to a severe unilateral hearing loss), subjects must rely on spectral-shape or overall-level information, which is primarily determined by the head shadow and pinna effects (Shaw, 1974; Shaw and Vaillancourt, 1985). The limits of localization performance based on interaural comparisons are well known [e.g., see Blauert (1997)], but the ability to use the information in the spectral shape is less well characterized. Studies of monaural localization that utilize a roving-level stimulus to reduce the overall-level information include those of Slattery and Middlebrooks (1994) and Wightman and Kistler (1997). More recently, there have also been a number of monaural localization studies with cochlear implant (CI) subjects [e.g., see van Hoesel *et al.* (2002), van Hoesel and Tyler (2003), Litovsky *et al.* (2004), Nopp *et al.* (2004), and Poon (2006)].

The motivation for this study comes from Blauert (1982) where he noted that there is a difference between reporting the perceived location of the source and using all the information to determine the actual location of the source. This distinction could be particularly important for monaural localization in which the normal ITD and ILD cues are missing

---

or in conflict with the spectral-shape cues. In this monaural-listening case, the ITD is undefined and the ILD is consistent with an extreme lateralization toward the presentation side. In fact, most listeners report that a monaural stimulus is perceived at the presentation ear (Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997). In spite of this perception, the spectrum of the received sound provides information about the location of a source. In previous monaural localization studies, subjects were generally requested to "report the apparent location" of the sound as opposed to reporting the location that they believed was correct, and no trial-by-trial feedback was given (Litovsky *et al.*, 2004; Yu *et al.*, 2006; Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997; Lessard *et al.*, 1998). The goal of the current study is to measure monaural localization performance when the subjects are instructed to use all the available information to identify which virtual location was presented and when correct-answer feedback is given.

The monaural stimuli used in these studies are unnatural for normal-hearing (NH) listeners who normally listen with two ears, and one might speculate that subjects with large unilateral losses (ULs) would perform better. When instructed to report the apparent location, chronically unilateral listeners, who normally localize with input to a single ear, performed better than subjects with NH (Slattery and Middlebrooks, 1994) and bilateral CIs (Poon, 2006). The benefits of prolonged monaural experience for monaural localization based on the spectral profile are not known; it is possible that the ability to perform a complex profile analysis is either equally useful for both binaural and monaural listeners or rapidly learned.

The motivation for this experiment was not to learn where subjects perceived monaural stimuli but to see if the monaural spectral profiles associated with azimuthal locations provided subjects with sufficient information to select the correct azimuthal location. This experiment is relevant for the interpretation of performance by bilateral CI users who are asked to identify source directions using a single implant. In interpreting results from studies on the ability of CI users to localize sounds monaurally, the instructions of the experiment may play a confounding role, as in the difference between "relate your impression of where the sound came from" and "try to respond with the correct location." In particular, it is possible that there are sufficient spectral cues for a unilateral CI listener to map out a reliable spectral profile associated with azimuthal locations and "get the right answer" meanwhile perceiving the location of every source as coming from the side of the unilateral implant.

Keeping in mind CI users, the ability of NH listeners to perform monaural localization tasks with constraints similar to those used in CI localization studies was tested. Due to the limited availability of bilateral CI subjects, CI localization studies often do not provide extensive training. Further, CI processing often limits the high-frequency information (up to 8 kHz) and number of channels (approximately 20) available to the listeners. Finally, when listening unilaterally, bilateral CI listeners do not have access to any (even extremely attenuated) interaural difference cues. Therefore, in the current tasks, NH and unilaterally impaired listeners were given little

training, and the stimulus was a flat-spectrum tone complex with 20 logarithmically spaced components that had a maximum frequency of 8 kHz. To eliminate the usefulness of all interaural difference cues as well as to provide a standardized set of spectral differences, virtual auditory techniques based on nonindividualized head-related transfer functions (HRTFs) were used. The HRTFs that were used were recorded with an in the ear microphone and not with a behind the ear microphone that is common with CIs. The effects of microphone placement on the HRTFs are likely limited for stimuli with a maximum frequency of 8 kHz because much of the information about the sound source location arises from head shadowing effects and not from pinna filtering.

We utilized both discrimination and absolute identification (AI) of azimuthal-angle paradigms to assess if the spectral profile (or more accurately the aspects of the profile that the subjects utilize) depends on the azimuth in a monotonic manner. The experimental paradigm differs from traditional measurements of localization in four ways. The most obvious difference is that the stimuli were presented monaurally. Two other changes were made to help subjects learn the mapping between potential cues and source locations; the first was that feedback was provided after every trial, and the second was that a fixed reference stimulus, designed to serve as a perceptual anchor (Braida *et al.*, 1984), was presented on every trial. The final change was, as mentioned previously, to use a flat-spectrum tone complex, as opposed to the noise and speech stimuli typical of localization studies. An advantage of using a tone complex, as opposed to a stimulus with a continuous spectrum, is that there is no variability in the short-time spectral shape. Future studies will address the influences of spectral variability (both short time and long time) on monaural localization. With these changes, the current study provides data as to the ability to identify the location of a sound source when all the available information, not just the perceived location, is utilized.

II. EXPERIMENTAL METHODS

Two different experiments were conducted with the same subjects, stimuli, and experimental apparatus. Both experiments used virtual auditory techniques with nonindividualized HRTFs to simulate sound sources with different locations. In the first experiment, the minimum audible angle (MAA) was estimated to characterize listeners' abilities to distinguish between two locations. In the second experiment, AI performance was assessed with a virtual-speaker array (seven or thirteen locations) spanning the frontal hemi-field. In all experiments, correct-answer feedback was provided after each trial.

The MAA task, the 7-location AI task, and the 13-location AI task explore different aspects of monaural localization. The MAA task provides insight into whether or not there is sufficient information in the spectral profile to discriminate between two locations. In the AI tasks, memory requirements are greater with 13 locations than with 7 locations, particularly if there is no percept that varies continuously with azimuth. Thus, similarity in performance [e.g., the root-mean-squared (RMS) difference between the response

J. Acoust. Soc. Am., Vol. 124, No. 5, November 2008

Shub *et al.*: Monaural identification of azimuth   3133

location and source location] in these AI tasks could suggest that the subjects use a percept that varies in a systematic manner with source location.

A. Subjects

Six subjects, with ages between 19 and 37 years, participated in both experiments. The subjects had either NH, corresponding to thresholds below 20 dB HL at octave frequencies between 250 and 8000 Hz, or a UL. The UL subjects had normal hearing at one ear and a substantial impairment, thresholds greater than 60 dB HL, at the other ear such that much of their everyday listening was effectively monaural.

The etiologies of the hearing losses varied among the UL subjects. Subject UL1 had unilateral congenital microtia and atresia that resulted in a conductive loss on the left side. Subject UL2 had an idiopathic congenital sensory-neural impairment at her right ear. Subject UL3 had an acoustic neuroma (right side) removed approximately seven years prior to the testing. The stimuli were presented to the normal hearing ear of the UL subjects (the right ear for UL1 and the left ear for UL2 and UL3) and to the right ear of the NH subjects. By including both NH and UL subjects, the benefits, if any, of long-term "monaural" listening on monaural location identification, as measured in this experiment, can be assessed.

The experience of the subjects with psychoacoustic experiments and their familiarity with the questions being explored varied. Subjects NH1, NH2, and UL1 are authors and had prior exposure to the stimuli and paradigms used in this work. In particular, NH1 had extensive exposure to the stimuli and paradigms during pilot listening. Subject UL3 had participated in numerous psychophysical tasks but did not have previous exposure to the stimuli or paradigms used in this work; the subject was also aware of the purposes of the experiment. Subjects NH3 and UL2 were naïve and had never participated in a psychophysical experiment; they received an hourly wage for their participation.

B. Spatial processing

Virtual auditory techniques were used for the spatial processing to eliminate all binaural cues. Virtual locations were simulated by filtering the flat-spectrum tone complex with a HRTF recorded at the right ear of a Knowles Electronics Manikin for Acoustic Research (KEMAR).[1] For all subjects the right-ear HRTF was used (with appropriate ipsilateral and contralateral associations), even when the stimulus was presented to the left ear (UL2 and UL3). The HRTFs of the left and right ears are not identical due to slight asymmetries in KEMAR and the recording apparatus. By only using nonindividualized HRTFs obtained at the right ear, the same stimuli were used for all subjects. Since nonindividualized HRTFs were used in all cases, any effect of using the HRTFs recorded at the right ear of KEMAR when presenting stimuli at the left ear of the subjects should be small. It should also be noted that the use of nonindividualized HRTFs might negate some of the potential benefits of the UL listeners. Specifically, benefits from familiarity with the unique spectral differences associated with their individual-ized HRTFs are lost. However, benefits from experience using only spectral profiles for localization are preserved.

C. Stimuli

A tone complex with 20 components spaced logarithmically between 500 and 8000 Hz was used in both the MAA and AI tasks. The rise and fall times of the tone complex were 25 ms. The relative phases of each component were randomized on every presentation, and the relative levels of each component (before spatial processing) were identical (i.e., flat spectrum).

The stimuli were band limited to frequencies below 8 kHz for four reasons. First, for frequencies above 8 kHz, the dependence of the spectral profile on location is not monotonic. For example, at some of these higher frequencies, a small change in location toward the side ipsilateral to the recording unexpectedly increases the level. Second, individual differences in HRTFs are more pronounced at frequencies above 8 kHz. Third, future investigations may focus on monaural localization of speech, which has limited energy above 8 kHz. Fourth, CI processors often limit the amount of high-frequency information.

There were two differences in the stimuli used in the MAA and AI tasks. These differences arose because the experiments were designed separately even though they were ultimately run on a common group of subjects. The first difference relates to the duration, and the second relates to the level. The duration of the tone complex in the MAA task was 1000 ms, while in the AI task (both with 7 and 13 locations) the duration was 800 ms. In the MAA task, after the spatial processing, the overall level of the stimulus was roved over a 10 dB range (± 5 dB) around 60 dB SPL on every presentation. In the AI tasks, the overall level, after spatial processing, was first normalized (to reduce the loudness differences across locations) and then roved by ± 4 dB (an 8 dB range).

D. Equipment

During the experiments, subjects sat in a sound-treated room in front of a LCD computer monitor and responded through a graphical interface via a computer mouse. On each trial, "lights" displayed on the monitor denoted the current interval number. The cueing lights were arranged horizontally and therefore moved left to right. Further, in the AI task, a graphical display of the speaker array with the appropriate response locations was continuously displayed to the subjects on the monitor. This graphical display was simply a half circle with small squares representing the possible speaker locations and a small circle representing the position of the listener. The visual angles of the graphical representation did not match the true source locations since the virtual-speaker array spanned the frontal 180° of azimuth, and therefore many of the sources would be out of the angular range of the monitor. In addition, the subject's viewing position of the graphical representation was not controlled. The experiment was self-paced, and listening sessions lasted no more than 2 h with rest breaks. A computer and Tucker-Davis-

Technology System II hardware generated the experimental stimuli at a sampling rate of 50 kHz. Stimuli were presented over Sennheiser HD 265 headphones.

E. Minimum audible angle procedures

In the MAA experiment, subjects were presented a series of two-interval trials. The first interval was presented from the virtual reference location (fixed for each run of 90 trials), and the second interval was from the virtual comparison location (to either the left or right of the reference location). The left and right comparison locations were presented with equal probability, and the subjects were instructed to report whether the comparison location was to the left or to the right of the reference location. A level randomization (chosen independently from interval-to-interval and trial-to-trial) was applied to the stimulus in each interval. There was 200 ms of quiet between each interval. Trials were self-paced, and the subjects received correct-answer feedback after every trial.

The probability of a correct response was measured at three reference locations with four different comparison angles for each reference. For each reference location and comparison angle, 90 trials were conducted. For each reference location, all 360 trials, 90 for each of the four comparison angles, were completed before the next reference location was tested. The ordering of the comparison angles was always 30°, 15°, 10°, and 5°. The ordering of the reference location relative to the midline was the same for all subjects; specifically, the ordering was (1) 60° to the left of the midline, (2) at midline, and (3) 60° to the right of the midline. Therefore, the right-ear subjects started with the reference location on the side contralateral to the presentation ear, and the left-ear subjects (UL2 and UL3) started with the reference location on the side ipsilateral to the presentation ear.

In the analysis of the MAA task, the difference between the reference and comparison angles is used and not the difference between the left and right comparison angles. Although an ideal observer would ignore the reference presentation and decide if the second interval was the left or right comparison, many of the subjects reported basing their decisions on whether the comparison was to the left or right of the reference. This strategy is consistent with attempting to reduce memory demands by not making comparisons to a standard held in memory.

F. Absolute identification procedures

In the AI task, subjects were presented a series of three-interval trials. The first and third intervals cued the subjects to the extremes of the virtual-speaker array. Specifically, the first interval of each three-interval trial was presented from the extreme location (90° away from midline) on the side contralateral to the presentation ear, and the third interval was presented from the extreme location on the side ipsilateral to the presentation ear.[2] The second interval was presented from a random virtual location, and subjects were asked to identify the location of the source on the second interval. The level randomization applied to each interval was chosen independently. There was 800 ms of quiet between each interval. Trials were self-paced, and subjects re-

ceived feedback after every trial. The feedback informed them of the actual source number both through a textual message and by highlighting the correct location in the graphical representation of the speaker array. The subjects were told that the goal was for their responses to be as close as possible to the correct answers.

Data were collected for both the 7-virtual-speaker array and the 13-virtual-speaker array. With both arrays, the virtual speakers were equally spaced across the frontal hemifield. The separations between locations were 30° and 15° for the 7-location and 13-location arrays, respectively. For the 7-location task, data were collected in runs of 70 trials (ten presentations from each location), while for the 13-location task, data were collected in runs of 65 trials (five presentations from each location). All subjects completed all five runs of the 7-location task prior to the collection of the data for the 13-location task. In this 13-location task, all subjects completed five runs except UL2 who completed only three runs due to limited availability.

Since the first interval was always from the extreme location on the side contralateral to the presentation ear and the third interval was always from the extreme location on the side ipsilateral to the presentation ear, subjects received identical acoustical stimuli regardless of the presentation ear. However, there was an asymmetry between left- and right-ear listening; the virtual locations moved left to right for the right-ear listeners and right to left for the left-ear listeners (UL2 and UL3). Additionally, the interval markers (cueing lights) displayed on the monitor always moved left to right. Therefore, the interval markers and the sounds moved in the same direction for right-ear listeners, while for left-ear listeners, the interval markers and the sounds moved in opposite directions. Subject UL3 (a left-ear subject) spontaneously reported that the dissonance between the lights and the sounds was "annoying." The impact of these subtle differences between left- and right-ear listening is discussed further below.

The data from the AI tasks were collected in terms of source-response confusion matrices; each entry of these matrices contains the frequency of a particular response (i.e., speaker number) being reported conditioned on the location of the presented stimulus. Confusion matrices can be visualized as dot scatter plots where the diameter of each dot is proportional to the frequency of occurrence of the associated source-response pairing. The source-response confusion matrices can also be characterized by a number of statistical measures. The slope and intercept of the best fitting (least-mean-squared difference) straight line as well as the RMS differences and the mean-unsigned (MUS) differences between the response and source locations are calculated. The RMS and MUS differences are calculated for all source locations together as well as separately for source locations ipsilateral and contralateral to the presentation ear. For completeness, the RMS and MUS differences are also calculated for the center location since this location is not included in either the ipsilateral or the contralateral calculations.

G. Familiarization and training

Subjects were given only limited familiarization and training[3] similar in duration to the familiarization and training stages of studies involving CI listeners. Before each task began, the subjects were given a short demonstration of the stimuli. Immediately following the demonstration, subjects were generally given a short training block to familiarize them with the experimental paradigm. Since subjects NH1, NH2, UL1, and UL3 had experience in discrimination procedures, they were not given a training block in the MAA task (they were, however, given a training block before the 7-location and 13-location tasks). Immediately following the training block, data collection for the appropriate task was started. The MAA task was conducted first followed by the 7-location task and, finally, by the 13-location task.

The demonstration consisted of a stimulus being presented from each virtual location (7 locations in the MAA and 7-location tasks and 13 locations in the 13-location task) from right to left and then from left to right (independent of the ear at which the stimulus was being presented). Following this presentation, stimuli from the extreme left location, then from the middle location, and finally from the extreme right location were presented. During all of these presentations, the cartoon of the speaker array described above was displayed on the monitor, and the currently active location was highlighted.

The short training blocks consisted of a few trials to allow the subjects to familiarize themselves with the experimental procedures. In the MAA task, the practice block consisted of 90 trials with the first reference angle and a change in angle of 30°. In the 7-location AI task, the practice block consisted of two trials from each location, while in the 13-location task, the practice block consisted of one trial from each location.

III. ANALYSIS OF THE SPECTRAL PROFILE

The obvious physical cue, for monaural localization of a fixed-spectrum random-level source, is the spectral shape (magnitude as a function of frequency) of the HRTF since it varies significantly with azimuthal angle. Figure 1 shows the spectral shape of the HRTF for the locations used in the seven-location AI task. The levels of the low-frequency components (below 1000 Hz) change by only a few decibels as a source is moved from the extreme left to the extreme right (as would be expected from the minimal head shadow effect at low frequencies). The levels of the middle frequencies (1000–8000 Hz) show strong dependencies on location and generally increase as the source is moved from the side contralateral to the recording (i.e., the right) toward the side ipsilateral to the recording (i.e., the left). Since the stimuli used in this work do not include frequencies above 8000 Hz, the high-frequency regions of the spectral profiles are not shown.

A spectral analysis of the HRTFs was used to calculate the performance of the ideal observer in the MAA task. The analysis is based on the model of Durlach *et al.* (1986) in which the activity in separate frequency channels is assumed to be statistically independent and to be combined optimally
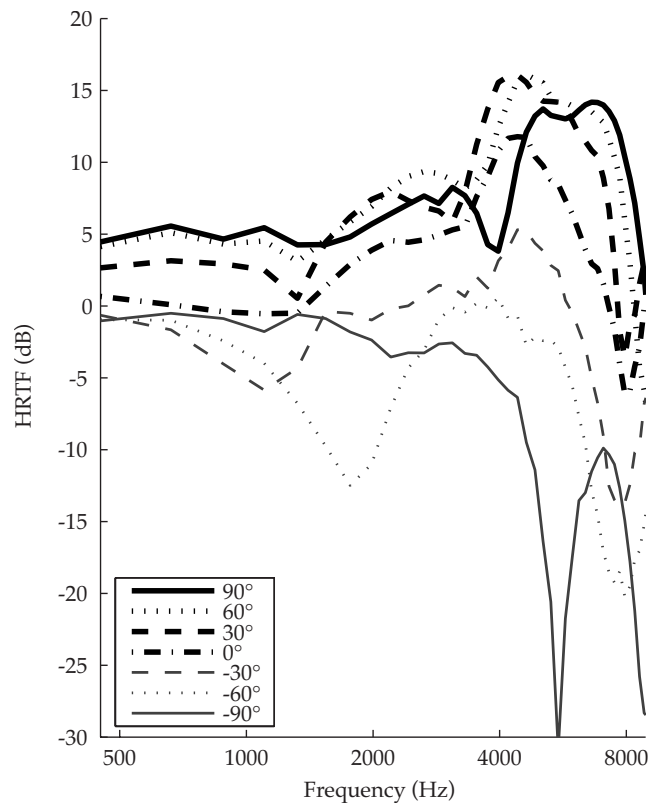

FIG. 1. Spectral profiles as a function of frequency for seven different locations in the frontal portion of the azimuthal plane. The spectral profiles are from right ear HRTF recordings of KEMAR. The same line styles are used for ipsilateral and contralateral locations; however, the linewidth for ipsilateral locations is wider than for contralateral locations. In the psychophysical experiments, the overall differences are reduced by normalization and level randomization, forcing the subjects to rely on differences in the spectral shapes.

such that the square of d' is equal to the sum of the squares of the d' values for the separate channels. The value of the predicted d' then is the product of the square root of the number of independent frequency bands and the RMS difference in the spectral levels for each band divided by the standard deviation of the internal noise. An internal noise with a standard deviation of 1 dB is consistent with thresholds from profile analysis studies involving relatively simple changes to tone complex stimuli [see Green 1988]. Farrar *et al.* (1987) estimated a larger standard deviation of the internal noise (between 4 and 13 dB) for the discrimination between two stimuli with continuous spectra and complicated spectral shapes.

In the current analysis, the standard deviation of the internal noise was chosen (*post hoc*) to be 8 dB, leading to predictions that roughly match the range of performance demonstrated by our subjects in the MAA experiments. In making the predictions we assume that each of the 20 components, which are logarithmically spaced between 500 and 8000 Hz, falls, into independent frequency bands. For comparison with the MAA results, d' is converted into the probability of a correct response by evaluating the standard normal cumulative probability distribution at one-half the predicted d'. The ideal observer ignores the reference and makes a decision based only on the comparison angle. The predictions do not take into account that many of the subjects

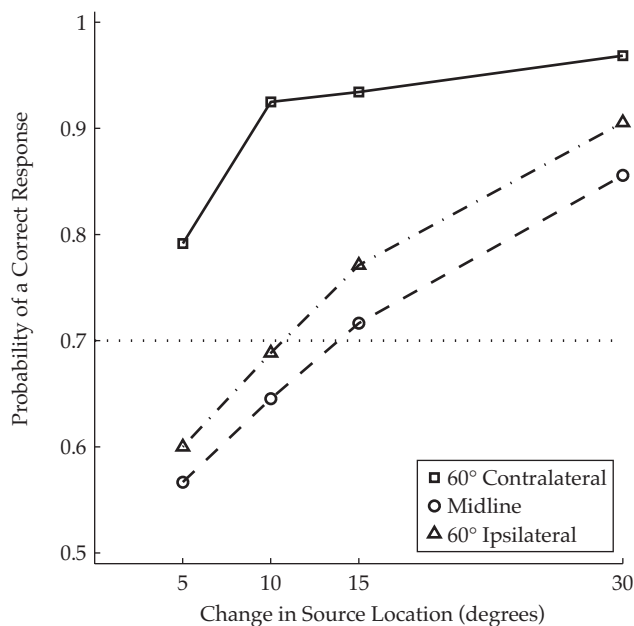Shub *et al.*: Monaural identification of azimuth

FIG. 2. Psychometric functions in the monaural MAA task for the ideal observer with an internal noise with a standard deviation of 8 dB. The solid, dashed, and dash-dotted psychometric functions correspond to source locations of 60° contralateral (square) to the presentation side, midline (circle), and 60° ipsilateral to the presentation side (triangle). The change in source location is the difference between the reference and left/right comparison locations, not the difference between the left and right comparison locations. The dotted line at 0.7 is a commonly used reference for threshold performance.
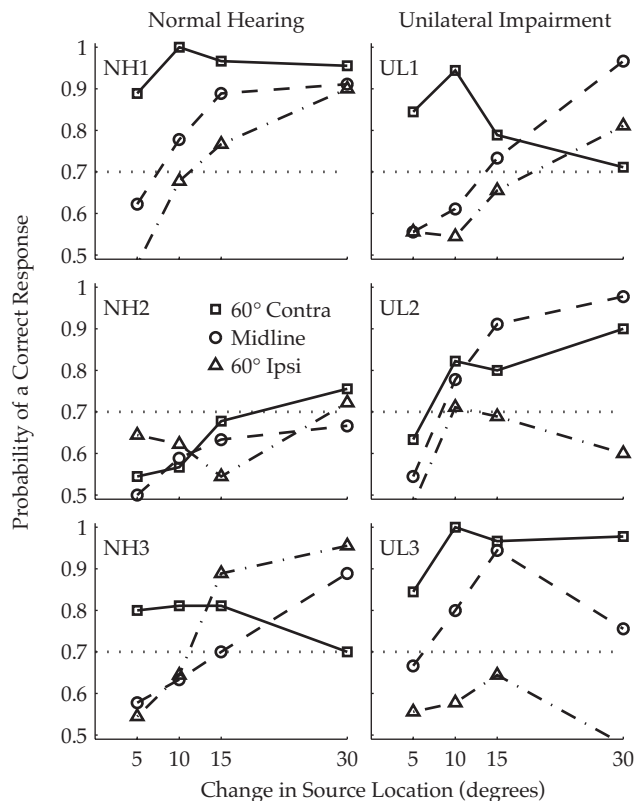

FIG. 3. Psychometric functions for the monaural MAA task for the six subjects. The panels in the left column are for the NH subjects, and the panels in the right column are for the unilateral-impairment subjects. The solid, dashed, and dash-dotted psychometric functions correspond to source locations of 60° contralateral (square) to the presentation side, midline (circle), and 60° ipsilateral to the presentation side (triangle). The presentation side was on the right for all subjects except UL2 and UL3. As in Fig. 2, the change in source location is the difference between the reference and left/right comparison locations, and the dotted line is a commonly used reference for threshold performance.

reported comparing the reference and comparison angles in the MAA task.

Figure 2 shows the predicted MAA psychometric functions (probability of a correct response as a function of the magnitude of the difference between the reference and comparison angles). As one might expect, the probability of a correct response generally increases with increasing spatial separation. Further, performance is greater for the 60° contralateral reference location than for the 60° ipsilateral reference location and is worst for the midline reference location. The probability of a correct response ranges from near chance with the smallest spatial separations at the midline reference location to better than 0.95 with the largest spatial separations at the 60° contralateral reference location.

IV. RESULTS

A. MAA results

Figure 3 shows the MAA psychometric functions obtained for the six subjects. Some of the psychometric functions are nonmonotonic even though the modeling suggests that increasing the spatial separation increases the information (see Fig. 2). Every subject, except NH2, was able to discriminate left from right for some of the tested reference angles and spatial separations. All subjects, except for NH2, were able to identify reliably (probability of correct greater than 0.7) if the comparison angle was 15° to the left or right of the reference angle for the 60° contralateral reference location (squares) and the midline location (circles). For large increments, subjects NH1, NH3, and UL1 could reliably dis-

criminate left from right for the 60° ipsilateral reference location (triangles), while NH2, UL2, and UL3 could not. From these MAA data, it is notable that the two populations (NH listeners and unilateral listeners) are overlapping in their abilities to perform these MAA tests with minimal training.

Learning effects are the likely cause of two aspects of the data. The non-monotonic shape of some of the measured psychometric functions may be an effect of learning because for each reference angle, all subjects were tested with the largest difference between the reference and comparison angles first and the smallest difference last. Learning effects may also be the reason that subjects UL2 and UL3 (the only left-ear subjects) were unable to achieve threshold performance with the 60° ipsilateral reference location that was tested first, but two of the three right-ear subjects (all except NH2), who were tested in the ipsilateral condition last, were able to obtain threshold performance. These effects require more testing on a larger group of subjects to be confident of the factors that are at play and to separate intersubject differences from learning effects.

Apart from subject NH2, the overall level of performance is roughly consistent with the analysis of the spectral profile (see Fig. 2) when the internal noise is assumed to have a standard deviation of 8 dB. One notable exception is

J. Acoust. Soc. Am., Vol. 124, No. 5, November 2008

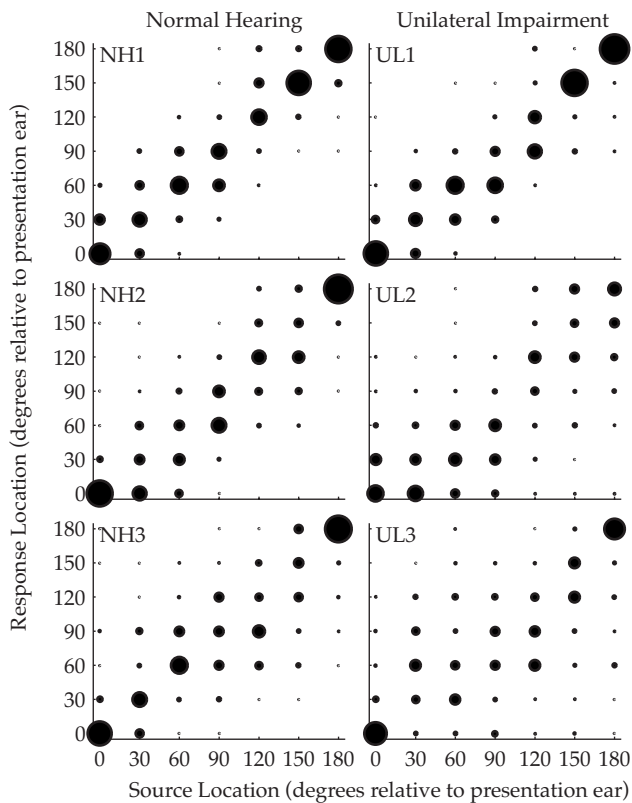Shub *et al.*: Monaural identification of azimuth 3137

FIG. 4. Dot scatter plots for the AI task with the 7-virtual-speaker array. The diameter of each dot is proportional to the frequency of the source/response pairing. The left column is for the NH subjects, and the right column is for the unilateral-impairment subjects. The source and response locations are relative to the ear at which the stimuli were presented (the right ear for all subjects except UL2 and UL3).

FIG. 5. Same as in Fig. 4, except for the 13-virtual-speaker array.

that the ideal observer performs worst with the midline reference location, while the subjects perform worst with the ipsilateral reference location. The relatively large amount of internal noise in the ideal receiver model needed to approximate the performance of our subjects may be due to the lack of training. It is possible that well trained subjects could perform substantially better on these discrimination tasks.

B. AI results

Figures 4 and 5 show scatter plots for the 7-location and 13-location AI tasks, respectively. As was the case in the MAA task, there were no consistent differences between the subjects with NH and those with unilateral hearing impairments. Overall, the performance in the AI task is quite good. For all subjects, the larger dots in the scatter plots generally lie near the major diagonal. Only a limited number of response locations differed from the source location by over 30°. Visually it seems clear that subjects are able to identify the location of a source within about 30°, consistent with the statistical results shown in Table I. For all the subjects in both tasks, the slopes of the best fitting lines are near unity. The maximum value of the slope is 0.97 and the minimum value is 0.61. The bias (intercept) is generally small, as indicated by the fact that all of the values of the intercepts were within one virtual speaker separation and the majority of the values (7 of 12) were within a quarter of a speaker separation. The measured value of the RMS differences ranged
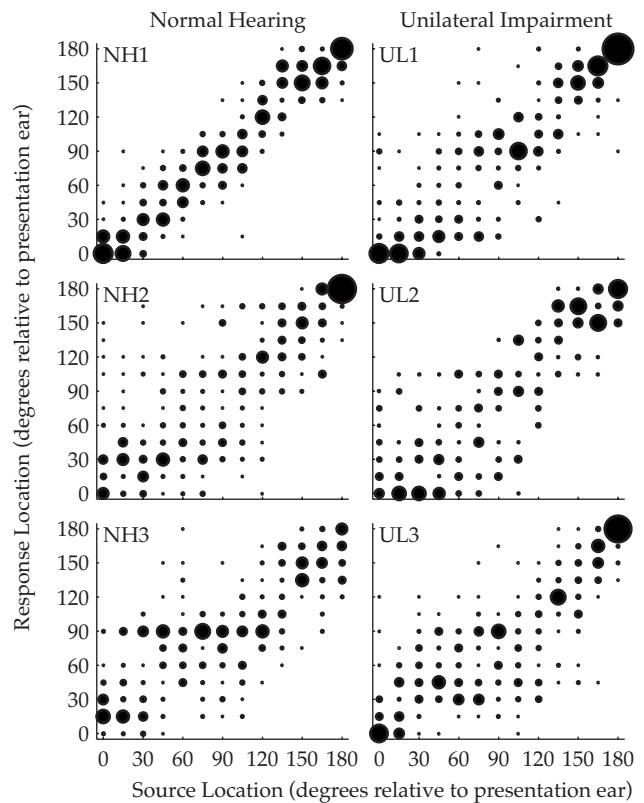
between 25.8° and 50.4° in the 7-location task and 19.1° and 39.3° in the 13-location task. The RMS difference weighs "outliers" (large mistakes) more heavily than the MUS difference. The MUS differences, however, had a similar range (16.0°–34.8°), suggesting that the effects of large errors on performance were small for both the 7- and 13-location tasks.

As in the MAA results, the data are consistent with the existence of some learning between the 7-location and 13-location tasks. Performance, as assessed by the slope, bias, RMS difference, and MUS difference, tends to be better in the 13-location task (tested second) relative to the 7-location task (tested first). In comparing these tasks, most slopes increased toward unity, the magnitudes of the biases decreased in every case, the RMS differences decreased on average and individually in two-thirds of the six cases, and the MUS difference decreased on average (but in only three of six cases). This trend toward a slight improvement suggests that some learning occurred in the course of testing. Whether this learning is related to the AI of location or to the experimental paradigm is unclear.

In contrast to the MAA case, better performance for sources located on the side contralateral to the presentation ear is not seen in the results for either the 7-location task or the 13-location task (see Table I). In these tasks, performance is similar for sources both ipsilateral and contralateral to the presentation ear. Performance for source locations of 0° and 180° is particularly interesting since those locations correspond to the ends of the range of stimuli and are the locations of the reference stimuli that were presented in the first (0°) and last (180°) intervals of each trial. The best perfor-

Shub *et al.*: Monaural identification of azimuth

TABLE I. Summary statistics for the 7-location and 13-location tasks.

			Normal hearing			Unilateral impairment		
			NH1	NH2	NH3	UL1	UL2	UL3
7 locations	Slope		0.91	0.91	0.78	0.90	0.72	0.61
	Bias (deg)		−2.31	6.77	2.22	4.97	−16.5	−8.74
	RMS difference (deg)	All locations	25.8	32.4	37.7	27.5	47.9	50.4
		Ipsilateral locations	26.3	33.4	37.2	26.5	38.3	45.5
		Center location	30.3	31.7	37.0	34.7	53.8	47.4
		Contralateral locations	23.6	31.7	38.5	25.7	54.1	55.7
	MUS difference (deg)	All locations	16.0	21.7	24.2	17.0	34.7	34.8
		Ipsilateral locations	17.4	22.0	21.0	16.6	27.6	31.8
		Center location	21.0	24.0	28.8	28.2	46.2	36.6
		Contralateral locations	13.0	20.6	25.8	13.6	38.0	37.2
13 locations	Slope		0.97	0.77	0.69	0.91	0.89	0.76
	Bias (deg)		0.55	−0.09	0.32	4.06	−4.92	−8.2
	RMS difference (deg)	All locations	19.1	39.3	35.6	32.0	35.0	36.4
		Ipsilateral locations	18.5	44.4	38.5	35.1	38.0	31.8
		Center location	20.6	43.3	32.6	37.2	44.2	37.5
		Contralateral locations	19.4	32.6	33.0	27.4	29.7	40.4
	MUS difference (deg)	All locations	13.6	28.2	27.2	22.5	26.3	25.8
		Ipsilateral locations	13.3	32.5	29.7	26.1	29.7	23.5
		Center location	15.0	38.4	22.8	28.8	34.0	26.4
		Contralateral locations	13.6	22.1	25.4	17.9	21.7	28.1

mance is consistently seen at these two extreme locations. The cueing intervals may have contributed to the similarity in performance for source locations ipsilateral and contralateral to the presentation ear.

V. DISCUSSION

The MAA results show that spectral-shape cues provide sufficient information for the discrimination of angles separated by 10° for many reference directions. This performance is consistent with that expected from spectral-shape discrimination as measured in previous studies (Farrar *et al.*, 1987). The AI results indicate that when feedback is provided so that subjects can use all available information to identify the virtual location, monaural localization performance can be quite reliable with RMS errors around 30°. In addition, it is notable that there were no consistent differences seen between chronically unilateral listeners and NH (bilateral) listeners, even though one might expect an advantage in these monaural experiments for listeners who never have interaural difference cues.

Previous studies found that the perceived location is not very reliable for location judgments with monaural stimuli (Litovsky *et al.*, 2004; Yu *et al.*, 2006; Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997; Lessard *et al.*, 1998). These previous studies generally focused on the perceived location as opposed to subjects' judgments about where the source is actually located and therefore did not provide feedback. The current study employed a paradigm,

including feedback, intended to encourage subjects to use any useful cue to optimize the identification of the location of a fixed-spectrum source. We consider the differences between our results and previous monaural localization experiments to reflect this distinction and the fact that the experiments with and without feedback (and with different instructions) measure different abilities.

This distinction between experiments with and without feedback is particularly important when the stimulus is monaural and is naturally perceived at the side of the stimulated ear. Providing feedback allows the subjects to use any information to identify the location of the sound source. Interestingly, the similarity in performance between the 7-location and 13-location AI tasks suggests that the subjects rely on a sound quality that varies systematically with source location. In fact, many of the subjects described attending to the timbre or sound quality of the stimulus for identifying the location of the sound source. Specifically, they reported that stimuli presented from locations contralateral to the presentation ear sounded more "muffled" than stimuli presented from source locations ipsilateral to the presentation ear.

Multiple factors influence performance so that comparisons between the current results and previous monaural localization studies must be done carefully. One clear difference is that unlike in Slattery and Middlebrooks (1994), the subjects in this study do not exhibit a strong bias (i.e., the intercepts of the best fitting lines are near zero). Another difference between the current results and previous monaural localization studies relates to differences in performance for

J. Acoust. Soc. Am., Vol. 124, No. 5, November 2008

Shub *et al.*: Monaural identification of azimuth   3139

sources ipsilateral and contralateral to the presentation ear. Slattery and Middlebrooks (1994) reported that performance was better for ipsilateral locations. Wightman and Kistler (1997) argued that the bias could account for the better performance on the ipsilateral side. Further, performance calculated from the optimum use of the spectral profile (see Fig. 2) suggests that performance should be better for contralateral locations. In the MAA task, performance was better for contralateral locations, but in the AI task, there were no substantial differences in performance for sources ipsilateral and contralateral to the presentation ear.

Another notable difference between the current results and previous monaural localization studies is the lack of an advantage associated with prolonged monaural experience (Poon, 2006; Slattery and Middlebrooks, 1994). The lack of a difference in performance between the NH and unilateral-impairment populations in this study is somewhat surprising. One possibility is that the ability to perform a complex profile analysis (not necessarily the analysis done by the unilaterally impaired in real world situations) is either equally useful (or not useful) for both binaural and monaural listeners or rapidly learned. It may also be that the benefits of the chosen paradigm (instructions, familiarization, cueing, feedback, and stimuli) are greater for normal-hearing subjects. Additionally, the use of nonindividualized HRTFs in a highly controlled laboratory task (e.g., fixed spectrum and virtual anechoic space) may have negated any advantages that the unilateral-impairment subjects have from prolonged monaural experience. The unilateral-impairment subjects may perform better then NH subjects in real world localization (familiar sounds, head movement, and visual input). Two of the subjects with unilateral impairments (UL1 and UL3), in fact, reported that they believed that the task would be easier with natural speech.

Understanding monaural localization abilities in more realistic situations (e.g., random-spectrum stimuli and reverberation) is particularly important for CI users. For CI listeners, the reported locations of fixed-spectrum sources in anechoic space is poorly correlated with the actual location (van Hoesel et al., 2002; van Hoesel and Tyler 2003; Litovsky et al., 2004; Nopp et al. 2004; Poon, 2006). In these studies, the improvement from near-chance performance with one implant to substantially better than chance performance with two implants is interpreted as evidence of a bilateral advantage. With the current paradigm and nonelectric hearing, however, subjects were able to perform reasonably well monaurally, about as well as listeners with bilateral CIs. The localization performance of NH individuals (using only one ear) and CI users (both unilateral and bilateral users) in more realistic environments needs to be measured. The extent to which this difference in monaural localization performance, between electric and nonelectric hearing, is a result of the CI processing (or impairments to the auditory system of CI listeners) or differences in the paradigm still needs to be investigated.

It appears that subtle changes in the paradigm can have measurable effects on performance. For example, the two subjects that listened with their left ears (UL2 and UL3) showed the most improvement between the 7-location (where the left-ear subjects performed worse than the right-ear subjects did) and 13-location tasks (where both performed the same). The paradigm for the left-ear listeners (specifically the incongruity between the cueing lights and the source locations) may have been less intuitive and therefore led to worse initial performance. This is consistent with the report by Hafter and Carrier (1970) in which a similar incongruity decreased the binaural release from masking. The concern about subtle changes in paradigm highlights the need for care when discussing monaural localization performance.

VI. SUMMARY

This work measured monaural performance in three localization tasks, all implemented with virtual stimuli and nonindividualized HRTFs. The three tasks were an azimuthal-angle discrimination experiment (MAA), a 7-location azimuthal-angle AI experiment, and a 13-location azimuthal-angle AI experiment. In general, conditions in the experiments were arranged to maximize performance. Most importantly, subjects were provided with correct-answer feedback and requested to try to get the correct answer using any useful cues. In addition, reference stimuli from the extremes of the range were provided in each trial of the identification experiments so that some context was maintained, reducing the memory demands on the subjects. Subjects in all experiments included both NH listeners and listeners with large unilateral impairments. Unlike previous monaural localization studies, which generally focused on the perceived location with no feedback provided, in the current experiments, subjects were asked to report where they thought the source was located rather than its perceived location, and feedback was provided. With this paradigm, the measured monaural AI performance was notably better than the near-chance localization performance reported in some previous experiments. Further, the response bias that is typically observed in such experiments was absent. Finally, advantages from long-term monaural experience were not observed; both the NH and unilateral-impairment listeners performed similarly on these tasks. The current results demonstrate that the paradigm can influence performance significantly, and considerable care is required in the interpretation of the ability to identify the location of a sound source when listening with a single ear.

ACKNOWLEDGMENTS

[1] The HRTFs used were obtained from the database of Algazi et al. (2001). There is no reason to suspect that results would have been different if KEMAR HRTFs from the Gardner and Martin (1995) database or the data reported in Shaw and Vaillancourt (1985) were used to process the stimuli.
[2] The cueing intervals provide no additional information and therefore do

not affect the ideal observer. The cues, however, can act as perceptual anchors and may greatly aid the subjects (Braida *et al.*, 1984).

Subject NH1 had substantial exposure to the stimuli and tasks during pilot listening, and subjects NH2 and UL1 had some exposure during pilot listening.

bibliography
Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics* (Mohonk Mountain House, New Paltz, NY), pp. 99–102.

Blauert, J. (**1982**). "Binaural localization," Scand. Audiol. Suppl. **15**, 7–26.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge).

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. **80**, 63–72.

Farrar, C. L., Reed, C. M., Ito, Y., Durlach, N. I., Delhorne, L. A., Zurek, P. M., and Braida, L. D. (**1987**). "Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises," J. Acoust. Soc. Am. **81**, 1085–1092.

Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York).

Hafter, E. R., and Carrier, S. C. (**1970**). "Masking-level differences obtained with a pulsed tonal masker," J. Acoust. Soc. Am. **47**, 1041–1047.

Lessard, N., Par, M., Lepore, F., and Lassonde, M. (**1998**). "Early-blind human subjects localize sound sources better than sighted subjects," Nature (London) **395**, 278–280.

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., and Yu, G. (**2004**). "Bilateral cochlear implants in adults and children," Arch. Otolaryngol. Head Neck Surg. **130**, 648–655.

Nopp, P., Schleich, P., and D'Haese, P. (**2004**). "Sound localization in bilateral users of med-el combi 40/40+cochlear implants," Ear Hear. **25**, 205–214.

Poon, B. B. (**2006**). "Sound localization and interaural time sensitivity with bilateral cochlear implants," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.

Shaw, E. A. G., and Vaillancourt, M. M. (**1985**). "Transformation of sound-pressure level from the free field to the eardrum presented in numerical form," J. Acoust. Soc. Am. **78**, 1120–1123.

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

van Hoesel, R., Ramsden, R., and Odriscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.

Yu, G., Litovsky, R. Y., Garadat, S., Vidal, C., Holmberg, R., and Zeng, F.-G. (**2006**). "Localization and precedence effect in normal hearing listeners using a vocoder cochlear implant simulation," Assoc. Res. Otolaryngol. Abstr. **29**, 660.

J. Acoust. Soc. Am., Vol. 124, No. 5, November 2008   Shub *et al.*: Monaural identification of azimuth   3141

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.243.253.102 On: Tue, 03 Dec 2013 17:55:20