The Effect of Head-Related Transfer Function Measurement Methodology on Localization Performance in Spatial Audio Interfaces

Justin A. MacDonald and **Phuong K. Tran**, Army Research Laboratory, Aberdeen Proving Ground, Maryland

Objective: Four head-related transfer function (HRTF) data sets were compared to determine the effect of HRTF measurement methodology on the localization of spatialized auditory stimuli. **Background:** Spatial audio interfaces typically require HRTF data sets to generate the spatialized auditory stimuli. HRTF measurement is accomplished using a variety of techniques that can require several nearly arbitrary decisions about methodology. The effects of these choices upon the resulting spatial audio interface are unclear. **Method:** Sixteen participants completed a sound localization task that included real-world, broadband stimuli spatialized at eight locations on the horizontal plane. Four different HRTF data sets were utilized to spatialize the stimuli: two publicly available HRTF data sets and two data sets obtained using different in-house measurement systems. All HRTFs were obtained from the Knowles Electronics Mannequin for Acoustics Research. **Results:** Unsigned localization error and proportion of front/back reversals did not differ significantly across HRTF data sets. Poorest accuracy was observed in locations near the medial (front/back) axis of the listener, mainly because of the relatively large proportions of reversals at these locations. **Conclusion:** This study suggests that the particular generalized HRTF data set chosen for spatialization is of minimal importance to the localizability of the resulting stimuli. **Application:** This result will inform the design of many spatial audio interfaces that are based upon generalized HRTFs, including wayfaring devices, communication systems, and virtual reality systems.

INTRODUCTION

Spatial (or 3-D) audio interfaces allow for the simulation of free-field sounds over a pair of headphones. Potential applications for this technology have been a consistent focus in the human factors literature for nearly 20 years. Interface designers often use the auditory pathway to offload information from the visual pathway to reduce cognitive workload and increase user performance. Spatial audio is particularly suited for the transmission of location information, as listeners are able to localize spatialized sounds quickly and accurately (Wightman & Kistler, 1989). Location information can be conveyed by taking advantage of the innate sound localization capabilities of the user. This strategy has been shown to improve the speed and accuracy of visual searches (Gunn et al., 2005; Perrott, Cisneros, McKinley, & D'Angelo, 1996; Perrott, Saberi, Brown, & Strybel, 1990; Rudmann & Strybel, 1999).

Spatial audio can also be used as an effective aid to navigation (Walker & Lindsay, 2006), allowing visually impaired users to utilize the auditory pathway to complete indoor and outdoor navigation tasks. Spatial audio is also an excellent tool to take advantage of the cocktail party effect (Cherry, 1953), allowing the user to attend to one or more speech streams by separating them in virtual space. Spatial audio has been shown to increase the intelligibility of speech in both divided and selective attention tasks (Abouchacra, Breitenbach, Mermagen, & Letowski, 2001; Bolia, Nelson, & Morley, 2001; Doll & Hanna, 1995; MacDonald, Balakrishnan, Orosz, & Karplus, 2002; Ricard & Meirs, 1994).

Address correspondence to Justin A. MacDonald, New Mexico State University, P.O. Box 30001/MSC 3452, Las Cruces, NM 88003; jmacd@nmsu.edu. ***Human Factors,*** Vol. 50, No. 2, April 2008, pp. 256–263. DOI 10.1518/001872008X250700.

Spatial audio interfaces function by mimicking the changes undergone by a sound wave as it travels through the environment to the ears of the listener. Most of these changes are a result of the filtering effects of listener's head and torso upon the incoming sound wave. These effects vary systematically with the location of the sound source and are unique to the individual listener. The listener's head serves as a barrier that introduces observable differences between the sounds that arrive at each of the ears. Such "binaural" effects include the interaural time difference (a difference in the time-of-arrival at the ears) and the interaural level difference (a difference in the intensity and frequency content at the ears). Other external structures, including the torso and outer ears, add additional "monaural" effects that further alter the sound before it arrives at the opening of the ear canal.

Although both binaural and monaural effects vary with the azimuth of the sound source (Musicant & Butler, 1984), binaural cues demonstrate the largest variance and are the dominant cue in azimuth estimation (Bernstein, 2004; Searle, Braida, Davis, & Colburn, 1976). Binaural cues do not provide a unique azimuth estimate, however. If the domain of potential locations is restricted to the horizontal plane, for example, any observed difference in time of arrival between the ears restricts the domain of possible source azimuths to two locations in the front and rear hemispheres (Blauert, 1997).

Although monaural cues can help to disambiguate the binaural information, head and torso cues are often subtle and easily overlooked. For this reason, human listeners are prone to front/back confusions in which the perceived and actual locations of the sound source are on opposite sides of the interaural axis. Listeners are considerably worse at estimating the elevation of a sound source, presumably because they are forced to rely exclusively on monaural cues (Makous & Middlebrooks, 1990; Perrott & Saberi, 1990).

Head-related transfer functions (HRTFs) are digital filters that approximate the effects of the listener's head and torso on an incoming sound wave. These filters are used to artificially introduce the binaural and monaural effects into a sound, allowing for a simulated free-field environment over headphones. Given that HRTFs vary considerably across individuals, spatial audio simulations are most realistic when generated using the HRTF of the specific user of the interface (Wenzel, Arruda, Kistler, & Wightman, 1993). Presumably, listeners are very familiar with the effects of their head and torso upon incoming sounds and make use of this information to estimate sound source location. A spatial audio interface based on individualized HRTFs is likely to be impractical for widespread deployment, however.

HRTF measurement is usually time-consuming and costly, requiring a large array of speakers, several microphones, and specialized software. As a consequence, generalized (generic) HRTFs are preferred for most spatial audio applications, especially those with a large number of users. Unfortunately, using a generalized HRTF to spatialize auditory stimuli introduces location cues that are different from those normally encountered by the listener, thereby reducing the realism of the simulation (Wenzel et al., 1993). Users of spatial audio interfaces based on generalized HRTFs exhibit greater errors in both azimuth and elevation estimates and an increased proportion of front/back confusions. Despite these disadvantages, practical considerations ensure that generalized HRTFs will remain a popular choice for spatial audio interface design.

Designers of spatial audio interfaces are confronted with a wide variety of HRTF data sets and measurement methodologies (e.g., Algazi, Duda, Thompson, & Avendano, 2001; Gardner & Martin, 1995; Zhou, Green, & Middlebrooks, 1992; Zotkin, Duraiswami, Grassi, & Gumerov, 2004). Determining which method or data set to use is a largely arbitrary decision. Estimating an HRTF is a complex process that can be accomplished using a wide variety of techniques and equipment, and the effects of measurement methodology upon the resulting spatial audio interface are unknown. Toward this end we compared the sound localization performance afforded by spatial audio interfaces based upon four different HRTF data sets.

We included two data sets that are available in the public domain: the Massachusetts Institute of Technology (MIT) HRTF database (Gardner & Martin, 1995) and the Center for Image Processing and Integrated Computing (CIPIC) HRTF database (Algazi et al., 2001). Given the difficulty of measuring HRTFs, the majority of spatial audio interface designers are likely to use databases that are available for public download. The MIT and CIPIC HRTF databases were chosen for two reasons: First, they are well known and easily obtained and, therefore, are likely to be encountered by spatial audio interface designers; second, the databases include HRTFs of the Knowles Electronics Mannequin for Acoustics Research (KEMAR; Knowles Electronics). The KEMAR serves as a "standard listener" that allows for direct comparisons between HRTF data sets.

HRTFs from the KEMAR were obtained using two other measurement systems: the HeadZap system by AuSim, Inc., and the Army Research Laboratory's (ARL's) in-house system. The HeadZap system is available for purchase directly from AuSim and was included because it is one of the few available commercial systems meant to measure HRTFs. ARL's recently developed in-house system was included so that its measurements could be compared with those of more well-established systems.

The four HRTF data sets differ in both the equipment and computational techniques used to measure the HRTF of the KEMAR. The four sets of KEMAR HRTFs were compared in a localization task using spatialized stimuli presented over headphones to determine if these considerable differences have any effect upon the resulting spatial audio interfaces. Sound stimuli were filtered through each HRTF to produce spatialized sounds. The test criterion was the localizability of the spatialized sounds when presented over headphones.

The goal of this study was to determine whether each of the four HRTF sets led to similar localization performance. We hypothesized that differences in HRTF measurement methodology would not lead to an associated change in localization performance. The rationale for predicting null results was twofold: First, each HRTF measurement system was designed to measure the same head and torso effects, and therefore correspondence among them is a reasonable expectation. Second, we needed to validate ARL's recently constructed measurement system by comparing its HRTF estimates with those from other systems. Validation requires similar results across systems – hence, the hypothesis of null results.

METHOD

HRTF Data Sets

The publicly available MIT HRTF database (for a more detailed description, see Gardner & Martin, 1995) includes 710 HRTFs (actually, head-related impulse responses) measured at 14 elevations from –40° to +90° and at equally spaced azimuths within each elevation (5° intervals on the horizontal plane). Gardner and Martin (1995) made the assumption that the head of the KEMAR was perfectly symmetrical and, therefore, that HRTFs need be collected for only one of the hemispheres (either left or right). This assumption allowed them to mount mismatched pinnae on the KEMAR so that the HRTFs associated with both pinna types could be collected simultaneously.

The KEMAR was mounted on a rotating turntable and positioned 1.4 m from a Realistic Optimus Pro 7 loudspeaker. Etymotic ER-11 microphones were mounted at the "eardrum" of the KEMAR to record the 16,383-sample maximum-length sequence (MLS; see Rife & Vanderkooy, 1989) output from the loudspeaker. Input to the microphones was sent through an Etymotic ER-11 preamplifier before being sent to a computer for HRTF estimation. The resulting impulse responses were truncated to 256 samples to remove room reflections and filtered to eliminate the effects of the loudspeaker's nonflat frequency response. Microphone-related frequency effects were not removed from the impulse response estimates.

The CIPIC HRTF database (see Algazi et al., 2001) includes HRTFs measured from 45 humans and the KEMAR. HRTFs were measured using a modified version of the Snapshot system (manufactured by Crystal River Engineering, now part of AuSim, Inc.). KEMAR measurements were taken at 5° intervals on the horizontal plane. The KEMAR was positioned at the center of a movable, 1-m-radius hoop with Bose Acoustimass loudspeakers mounted on it. The ear canals of the KEMAR were blocked, and Etymotic ER-7C microphones were mounted at the entrance to the ear canal. HRTFs were estimated using the Golay code method (Golay, 1961), filtered to remove the frequency effects of the microphone and loudspeakers, and truncated to 200 samples to remove room reflections.

ARL's in-house HRTF measurement system was used to generate the third KEMAR HRTF data set. This system includes a loudspeaker attached to a RoboArm 360 robotic arm (manufactured by Tucker-Davis Technologies [TDT]) that can be positioned at any point between –50° and +60° elevation to 1° precision. The KEMAR was positioned at the center of the robotic arm's range of movement during measurement while the computer-controlled robotic arm moved the loudspeaker to the location of interest. Stimulus presentation and recording of the output signals were controlled through TDT System II/III signal processing hardware. The MLS method was used to measure the HRTF.

A 16,383-sample MLS stimulus was generated in Matlab, sent through the TDT System II hardware, and played through a GF0876 loudspeaker (CUI, Inc.) attached to the end of the robotic arm. The distance from the loudspeaker to the center of the head was 1 m. Input to each of the ears was recorded using a pair of EM-125 miniature electret microphones (Primo Microphones, Inc.) mounted in foam inserts in the ear canal of the KEMAR. The input to each of the microphones was sent to a preamplifier and microphone amplifier before being sent to a computer. HRTFs were computed by convolving the raw impulse responses with the inverse of the system impulse response measured at the location of the center of the head without the presence of the KEMAR. HRTFs were truncated to 256 samples to remove room reflections from the measurements.

The HeadZap system is a complete HRTF measurement system consisting of hardware and software developed by AuSim, Inc. HeadZap uses a proprietary loudspeaker attached to a stationary mount. Proprietary microphones are attached to foam inserts and inserted into the ear canal. HRTF measurement requires the movement of the participant. Markers are located at 15° intervals, and the participant faces each of the markers in turn to measure the HRTF.

The loudspeaker can be attached at several preset positions on the mount to allow for measurements at ±54°, ±36°, ±18°, and 0° elevation and is located 1 m from the center of the participant's head. The loudspeaker is repositioned when the measurements for a new elevation are needed. Finer resolution of the HRTF can be interpolated by linear or exponential fitting through the software provided by AuSim. HeadZap measures impulse responses using the Golay code method (Golay, 1961). HRTFs are filtered to remove the frequency distortion introduced by the system components, and room reflections are removed from the HRTF by truncating them to 256 samples.

Participants

Sixteen civilian participants (5 men, 11 women) were compensated at the rate of $20 per hr ($40 total) for their participation in this study. Participants ranged in age from 19 to 27 years (mean: 23.3 years). Participants were tested for normal hearing, which was defined as thresholds less than or equal to 20 dB HL at the octave frequencies between 250 and 4000 Hz. All participants passed the hearing screening and reported no previous experience with spatial audio displays.

Apparatus

The experiment was conducted on a laptop computer with a 14.1-inch (35.8-cm) screen at 1,024 × 768 resolution. Responses were made using a two-button mouse. Stimuli were presented through a pair of AKG K240DF headphones.

Stimuli

Ten stimuli ranging in duration from 350 to 500 ms were used in this experiment. The stimuli were sounds that should be familiar to the listener, such as a car horn, a frog croak, a speech stimulus, and breaking glass. All stimuli were filtered to account for the nonflat frequency response of the headphones. Each stimulus was convolved with four different sets of HRTFs to produce spatialized sounds.

Stimuli were spatialized at eight virtual locations around the head: 0°, 30°, 90°, 150°, 180°, –120°, –90°, and –60°, in which 0° is directly in front and +90° is to the right of the listener. All stimuli were spatialized on the horizontal plane. A pilot version of this study included locations spaced at 30° intervals on the horizontal plane. Roughly half of these locations are redundant because of the left/right symmetry of the head of the KEMAR and were therefore not included in the full-scale study. A total of 320 sounds (10 sounds × 4 HRTF sets × 8 virtual locations) were used in the experiment, with levels of approximately 80 dB(A) measured at the output of the headphones.

Procedure

The experiment began with a hearing test. Participants who successfully passed the hearing test were seated in front of the computer to begin the study. Experiment trials began after a short instruction period. For each trial the participant clicked a button on the screen using the computer mouse. After a 1-s interval a sound was presented at one of eight locations in virtual space. The task of the participant was to indicate the apparent location of the sound by clicking on the appropriate point on a circle on the computer screen. Participants were allowed to respond at any time after stimulus offset. Participants then clicked a button on the screen to proceed to the next trial. Feedback was not provided. Breaks were taken as needed between trials, and the experiment was entirely self-paced. Participants completed a total of 640 trials, composed of 2 trials for each of the 320 spatialized stimuli. The order of all stimuli was completely randomized for each participant.

RESULTS

Mean absolute errors for each location and HRTF set are shown in Figure 1. Locations in Figure 1 are arranged in order from directly in front of the listener to directly behind the listener. As each of the 10 stimuli led to similar localization performance, data for each of the 10 stimuli were combined for all analyses. Mean absolute error is defined as

$$\frac{1}{n}\sum|\theta_e|,$$

in which n is the number of trials in each condition and θ_e is the angular distance between the estimated and actual locations in degrees. Individual localization performance was qualitatively very similar to the mean performance illustrated in Figure 1. Performance at 0° (directly in front of the listener) was somewhat variable, however: 4 of the 16 participants localized the majority of the 0° sounds in the rear hemisphere. This behavior occurred regardless of the HRTF set used to spatialize the sounds.

The bars around each mean represent the 95% confidence interval (CI) associated with that mean. Overlapping CIs indicate no significant difference between means. Most importantly, the HRTF sets led to similar localization performance: all four CIs overlapped at each location, indicating no significant difference in the localization performance afforded by each of the HRTFs. Not surprisingly, a strong effect of location was observed: Sounds were localized more accurately when they were located off of the medial axis.

In addition, the HRTF set used to spatialize the sounds had no significant effect on the number of front/back reversals at a given location (see Figure 2). The overall mean reversal rate ranged from 45% to 48% across HRTF data sets. By definition, a reversal occurred when the perceived and actual locations of the sound were on opposite sides of

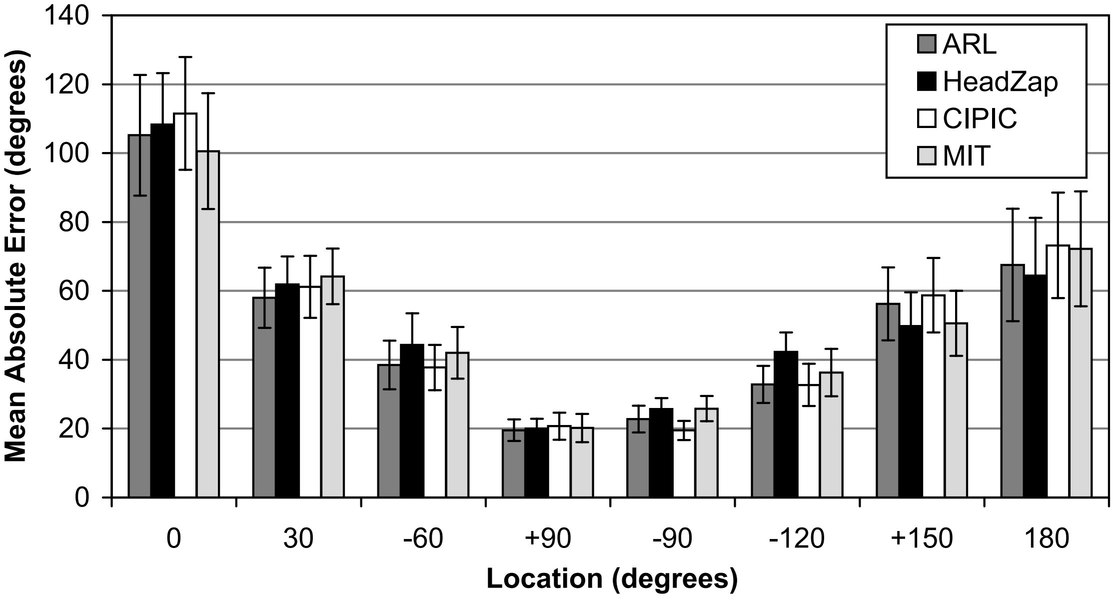

Figure 1. Mean absolute error by virtual location arranged in order from directly in front of the listener to directly behind the listener. 0° corresponds to the point directly in front of the listener on the horizontal plane, and negative and positive angles are located in the left and right hemispheres, respectively. The error bars represent the 95% confidence intervals around each mean. (ARL = Army Research Laboratory in-house system, CIPIC = Center for Image Processing and Integrated Computing database, MIT = Massachusetts Institute of Technology database.)

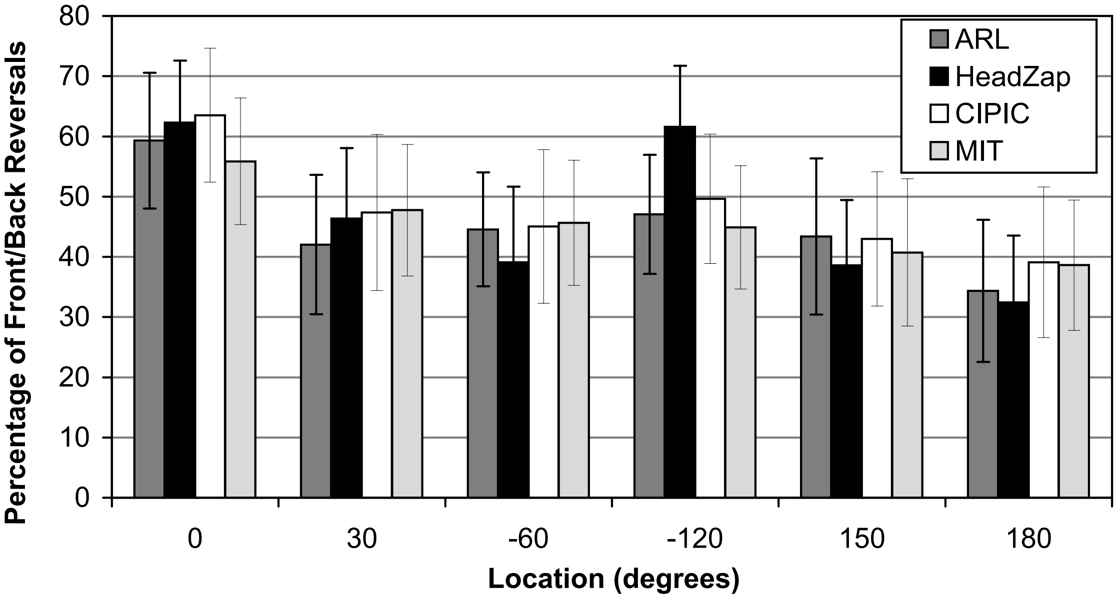

Figure 2. Percentage of front-to-back or back-to-front reversals by location. By definition a reversal occurred when the estimated and actual locations were on opposite sides of the interaural axis. (ARL = Army Research Laboratory in-house system, CIPIC = Center for Image Processing and Integrated Computing database, MIT = Massachusetts Institute of Technology database.)

the interaural axis. However, a small effect of location on reversals was observed: sounds located at 0° led to more reversals than did those located at 180° for three of the four HRTF sets. Because of the large number of front/back reversals, most generalized HRTF-based studies report data that are "corrected" by reflecting front/back confusions across the interaural axis, thereby reducing the localization error associated with trials in which a reversal occurred. The corrected mean localization error in this study ranged from 22° to 24° across HRTF data sets.

DISCUSSION

A hypothesis predicting null results requires a careful choice of statistical treatment and an equally careful interpretation of experiment results. The use of traditional null-hypothesis significance testing in this case is questionable (Loftus, 1996). Instead we chose to utilize CIs to describe our data. Indeed, there is considerable support for this choice in the literature (Aberson, 2002; American Psychological Association, 2001; Fisher & Belle, 1998; Grant, 1962). Loftus (1996) argued that CIs are particularly appropriate when the researcher is arguing for acceptance of the null hypothesis.

Null findings require the consideration of statistical power before they can be interpreted properly. Interpreting null results as evidence for a trivial effect size requires the elimination of low statistical power as a possible explanation. CIs provide an easily understood indication of statistical power: Narrow intervals indicate high power. Of course, if the true effect size is small, then the statistical power of any experiment is likely to be less than ideal: Power is proportional to the effect size. A poorly designed experiment attempting to uncover a large effect could quite easily be more powerful than a well-designed experiment looking for a small effect. Given that size of the effect under consideration (the effect of HRTF measurement methodology on localization performance, in this case) is not known a priori, power can be estimated by determining the ability of the experiment to uncover effects of known magnitudes.

In the case of our experiment, the pattern of localization errors and reversals observed in our study can be compared with the results of similar localization studies using HRTF-filtered stimuli. The corrected error rate of 23° and reversal rate of 45% observed in this experiment are quite comparable to results reported in previous studies using generalized HRTFs. Begault and Wenzel (1993) reported a mean corrected localization error of 28° in azimuth and a front/back confusion rate of 29%.

Endsley and Rosiles (1995) reported corrected errors between 16° and 24° in azimuth, depending upon the type of stimulus that was presented.

Wenzel et al. (1993) reported corrected azimuth localization errors ranging from 22° to 31° across elevations and a front/back confusion rate of 31%. Reversals occurred most often in locations near the medial axis, and more front-to-back reversals than back-to-front reversals were observed. Chen (2003) reported mean corrected localization errors between 19° and 24° across a variety of short-duration stimuli located on the horizontal plane and front/back confusion rates between 22% and 40%. Brungart and Simpson (2001) reported a front/back reversal rate of 44% using KEMAR's HRTF.

In addition, the effect of virtual source location on localization performance in the current study matches previously reported results quite closely. Maximal errors were observed for locations on or near the medial axis, whereas errors were minimized for locations along the interaural axis. This pattern was observed in several other studies that have examined error across source location using spatial audio displays (Begault & Wenzel, 1993; Wenzel et al., 1993; Wightman & Kistler, 1989). These studies also reported a higher proportion of front-to-back reversals than back-to-front reversals. This pattern was replicated in the current study, although the CIs for the 0° and 180° locations overlapped in the HeadZap condition.

Given the correspondence between the results of the current study and those of similar studies, it is reasonable to conclude that this study had power sufficient to uncover any nontrivial effect of measurement methodology on localization performance. Considering that no effect was found, therefore, we conclude that the true effect size was likely too small to be considered important. The HRTF sets can be considered to be functionally equivalent in terms of their localization performance, at least within the range of azimuths considered in the study.

Of course, differences between the HRTF data sets may be much more pronounced if other comparison metrics are used. Localization accuracy is not an ideal comparison metric for all possible spatial audio applications. For example, designers of spatial audio communication interfaces should consider a metric based upon speech intelligibility rather than localization accuracy. In addition, it is possible that differences in localization performance across measurement systems would have been uncovered had we employed participants with more experience using spatial audio displays. Experienced participants would likely exhibit decreased response variability, and a corresponding increase in statistical power would result.

Furthermore, it is possible that major differences between HRTF data sets are to be found only in their spatialization of the vertical dimension. We chose not to gather elevation data in this experiment, given the relatively poor acuity in elevation judgments (Perrott & Saberi, 1990). Wightman and Kistler (1989) reported larger interparticipant variance in elevation judgments relative to azimuth judgments when individualized HRTFs were used, suggesting that any comparison of HRTF data sets based upon elevation estimates would suffer from relatively low statistical power. The situation is even worse when generalized HRTFs are used: Wenzel et al. (1993) reported that the performance degradation attributable to the use of generalized HRTFs is most often observed in elevation estimates. Therefore, if our experiment had varied the elevation rather than the azimuth of the sound sources, it would likely have suffered from low statistical power unless the sample size had been increased to a prohibitively large number.

Given that the power of such an experiment is likely to be low and that the addition of nonzero elevations would increase the number of trials by an order of magnitude, we felt that a full comparison of measurement systems across azimuth and elevation was beyond the scope of this initial study. A full-scale comparison examining localization performance in both azimuth and elevation is planned for the near future.

It is clear from an examination of these results that the generalized HRTFs used in this study do not allow for highly accurate sound localization. Unfortunately, relatively poor localization accuracy and frequent reversals are the norm for spatial audio interfaces utilizing generalized HRTFs. What effect poor localization performance has upon the utility of the spatial audio interface most likely depends upon the application. Presumably, high levels of localization blur will reduce the effectiveness of systems such as wayfaring devices, which must transmit precise location information. Other types of systems, such as spatial audio-based communication devices, would likely be less affected, however.

The most important result from this small study is that the ARL, HeadZap, MIT, and CIPIC estimates of the KEMAR HRTF led to functionally similar spatial audio simulations despite the considerable differences in measurement methodology. Considered in terms of the localization performance they provide, the four sets of generalized HRTFs were nearly indistinguishable from one another. This is good news for the spatial audio interface designer: This result suggests that generalized HRTF data sets obtained using considerably different methods can be used interchangeably.

REFERENCES

Aberson, C. (2002). Interpreting null results: Improving presentation and conclusions with confidence intervals. *Journal of Articles in Support of the Null Hypothesis, 1,* 36–42.

Abouchacra, K. S., Breitenbach, J., Mermagen, T., & Letowski, T. (2001). Binaural helmet: Improving speech recognition in noise with spatialized sound. *Human Factors, 43,* 584–594.

Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. (2001). The CIPIC HRTF database. In *Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics* (pp. 99–102). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Begault, D. R., & Wenzel, E. M. (1993). Headphone localization of speech. *Human Factors, 35,* 361–376.

Bernstein, L. R. (2004). Sensitivity to interaural intensitive disparities: Listeners' use of potential cues. *Journal of the Acoustical Society of America, 115,* 3156–3160.

Blauert, J. (1997). *Spatial hearing.* Cambridge, MA: MIT Press.

Bolia, R. S., Nelson, W. T., & Morley, R. M. (2001). Asymmetric performance in the cocktail party effect: Implications for the design of spatial audio displays. *Human Factors, 43,* 208–216.

Brungart, D. S., & Simpson, B. D. (2001). Auditory localization of nearby sources in a virtual audio display. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics* (pp. 107–110). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Chen, F. (2003). Localization of 3-D sound presented through headphone – Duration of sound presentation and localization accuracy. *Journal of the Audio Engineering Society, 51,* 1163–1171.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America, 25,* 975–979.

Doll, T. J., & Hanna, T. E. (1995). Spatial and spectral release from masking in three-dimensional auditory displays. *Human Factors, 37,* 341–355.

Endsley, M. R., & Rosiles, S. A. (1995). Auditory localization for spatial orientation. *Journal of Vestibular Research, 5,* 473–485.

Fisher, L. D., & Belle, B. D. (1998). *Biostatistics: A methodology for the health sciences.* Chichester, UK: Wiley.

Gardner, W. G., & Martin, K. D. (1995). HRTF measurements of a KEMAR. *Journal of the Acoustical Society of America, 97,* 3907–3908.

Golay, M. J. E. (1961). Complementary series. *IRE Transactions on Information Theory, 7,* 82–87.

Grant, D. (1962). Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review, 60,* 54–61.

Gunn, D. V., Warm, J. S., Nelson, W. T., Bolia, R. S., Schumsky, D. A., & Corcoran, K. J. (2005). Target acquisition with UAVs: Vigilance displays and advanced cuing interfaces. *Human Factors, 47,* 488–497.

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 5,* 161–171.

MacDonald, J. A., Balakrishnan, J. D., Orosz, M. D., & Karplus, W. P. (2002). The intelligibility of speech in a virtual 3D environment. *Human Factors, 44,* 272–286.

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America, 87,* 2188–2200.

Musicant, A. D., & Butler, R. A. (1984). The psychophysical bases of monaural localization. *Hearing Research, 14,* 185–190.

Perrott, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1996). Aurally aided visual search under virtual and free-field listening conditions. *Human Factors, 38,* 702–715.

Perrott, D. R., & Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *Journal of the Acoustical Society of America, 87,* 1728–1731.

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Perception and Psychophysics, 48,* 214–226.

Ricard, G. L., & Meirs, S. L. (1994). Intelligibility and localization of speech from virtual directions. *Human Factors, 36,* 120–128.

Rife, D. D., & Vanderkooy, J. (1989). Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society, 37,* 419–444.

Rudmann, D. S., & Strybel, T. Z. (1999). Auditory spatial facilitation of visual search performance: Effect of cue precision and distractor density. *Human Factors, 41,* 146–160.

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

Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors, 48,* 265–278.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America, 94,* 111–123.

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

Zhou, B., Green, D. M., & Middlebrooks, J. C. (1992). Characterization of external ear impulse responses using Golay codes. *Journal of the Acoustical Society of America, 92,* 1169–1171.

Zotkin, D. N., Duraiswami, R., Grassi, E., & Gumerov, N. A. (2004). *Fast head related transfer function measurement via reciprocity* (Computer Science and UMIACS Tech. Rep. CS-4620). College Park: University of Maryland.

Justin A. MacDonald is currently an assistant professor at New Mexico State University in Las Cruces, New Mexico. At the time this work was conducted, he was a research psychologist at the U.S. Army Research Laboratory in Aberdeen Proving Ground, Maryland. He received his Ph.D. in psychology from Purdue University, West Lafayette, in 2003.

Phuong K. Tran is an electrical engineer at the U.S. Army Research Laboratory in Aberdeen Proving Ground, Maryland. She received her M.S. in engineering science from Pennsylvania State University, University Park, in 1994.

Date received: January 3, 2007
Date accepted: October 24, 2007