Free-Field Localization Performance With a Head-Tracked Virtual Auditory Display

Griffin D. Romigh, *Member, IEEE*, Douglas S. Brungart, *Member, IEEE*, and Brian D. Simpson

***Abstract*—Virtual auditory displays are systems that use signal processing techniques to manipulate the apparent spatial locations of sounds when they are presented to listeners over headphones. When the virtual audio display is limited to the presentation of stationary sounds at a finite number of source locations, it is possible to produce virtual sounds that are essentially indistinguishable from sounds presented by real loudspeakers in the free field. However, when the display is required to reproduce sound sources at arbitrary locations and respond in real-time to the head motions of the listener, it becomes much more difficult to maintain localization performance that is equivalent to the free field. The purpose of this paper is to present the results of a study that used a virtual synthesis technique to produce head-tracked virtual sounds that were comparable in terms of localization performance with real sound sources. The technique made use of an *in-situ* measurement and reproduction technique that made it possible to switch between the head-related transfer function measurement and the psychoacoustic validation without removing the headset from the listener. The results demonstrate the feasibility of using head-tracked virtual auditory displays to generate both short and long virtual sounds with localization performance comparable to what can be achieved in the free field.**

***Index Terms*—Head-related transfer functions (HRTFs), localization.**

I. Introduction

To preserve accurate spatial perception, a virtual system must recreate acoustic cues inherent in natural free-field listening. Of particular importance are those represented in a subject's head-related transfer function (HRTF), the set of filters derived to match the acoustic transformation imparted on a sound source as it interacts with a listener's head, shoulders, and outer ears [1], [2]. These HRTFs are complex functions of space, frequency and left-or-right ear, and they also must be measured separately for each individual listener to preserve localization performance [3]–[5]. Extreme care must also be taken to ensure that the transfer function from the headphone system used to reproduce the sound to the listener's eardrum is completely accounted for in the virtual reproduction [6].

Manuscript received July 15, 2014; revised December 05, 2014, March 11, 2015; accepted March 12, 2015. Date of publication April 09, 2015; date of current version July 14, 2015. This work was supported by the Air Force Office of Scientific Research (AFSOR) under Grant 08-RH-06-COR. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Lauri Savioja.

G. Romigh and B. Simpson are with the 711th Human Performance Wing, Air Force Research Laboratory, Dayton, OH 45435 USA (e-mail: griffin.romigh@us.af.mil).

D. Brungart is with Walter-Reed National Military Medical Center, Bethesda, MD 20889 USA (e-mail: douglas.s.brungart.civ@health.mil).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2015.2421874

Because of these complications, very few studies have demonstrated perceptual equivalence between the spatial locations of real (free-field) and virtual sources. Langendijk and Bronkhorst [7], and Zahorik *et al.* [8] showed that, with careful calibration and an in situ presentation, virtual sources could be generated that were indistinguishable from free-field sources in a discrimination task. Similarly, Hartman and Wittenberg [9], showed that careful calibration allowed for real-virtual equivalence of subjective externalization judgments. In terms of localization, Martin *et al.* [10] showed that, for a small number of subjects, free-field equivalent localization performance could be achieved with virtual sources. Despite the few successes, most studies have still shown differences in localization performance between free-field and virtual sources, even when individualized HRTF measurements are available [11], [12], [4].

In general, the largest difference in localization performance between free-field and individualized virtual sources seems to be an increase in the occurrence of front-back reversals, where a listener hears a sound source at the correct lateral angle but at the mirror image of its true location across the frontal plane. In natural listening situations, listeners are able to alleviate this type of error through small exploratory head movements that create differential changes in binaural cues that allow for disambiguation of front-back location [13]–[15]. In the absence of these dynamic head-motion cues, as occurs with brief stimuli under 300 ms and static, non-head-tracked, virtual audio systems, listeners must rely on subtle features contained within the sound source spectrum to make front-back location judgments.

In order to provide the appropriate head-motion cues, virtual audio systems must monitor the position of the listener's head in real-time, and be able to smoothly transition between sampled HRTF locations in response to the listener's head movements. Unfortunately, this feature introduces several sources of potential error that are not included in the full-fidelity *static* virtual audio systems like those in Langendijk and Bronkhorst [7] and Martin *et al.* [10].

Several authors have shown that degradation of localization performance can occur if the combined process of acquiring head position data, selecting the appropriate synthesis filters, and implementing those filters cannot be accomplished in sufficiently near real-time [16]–[18]. Therefore, in order to reduce processing time, most head-tracked virtual audio displays utilize truncated minimum-phase FIR representations of the HRTFs, along with a separate interaural time delay. Due to the minimum energy delay property of minimum phase filters [19], truncation of the corresponding time domain representation of these filters results in minimal distortion. While processing speed is less of a consideration than it used to be for real-time HRTF sys-

tems, the potential advantages of using shorter filters in terms of processing load and battery life are still quite relevant, and reduced processing load may be important for minimizing the overall total latency of the virtual audio system. Additionally, because traditional HRTF sets only contain measurements for a finite number of discrete locations and virtual sources may need to be rendered at any arbitrary location relative to the listener, head-tracked virtual audio systems must employ some type of interpolation and/or cross-fading technique, both of which inherently introduce some amount of error into the underlying spatial representation.

While a number of authors have described an implementation of head-tracked virtual audio (e.g., [20]–[24]) few have included a localization task to validate system performance. Of those that have provided localization results Wenzel [17] and Brungart *et al.* [18] did not include a comparison to free-field performance, while Bronkhorst [4] and Sandvad [16] both showed virtual performance that was significantly worse than free-field, even with the added benefit of dynamic head-motion cues. Djelani *et al.* [24] reported that the performance of their headtracked system was comparable to the free field; however, a direct comparison to free-field data was only available for a single subject, and, as reported, that subject showed atypical localization responses to elevated sources with all stimuli.

In order to conduct basic research on what auditory cues are used for localization, and how these auditory cues are interpreted perceptually in highly dynamic and interactive environments, there is a need to be able to generate interactive virtual sounds that produce localization performance comparable to the performance achieved in the free field. This will allow researchers to systematically degrade or otherwise modify the acoustic cues present in the individualized HRTF in order to examine the impact these changes have on overall localization performance. Thus, an important first step is to identify the fundamental limitations that might make it impossible to achieve free-field-equivalent localization performance in an interactive virtual display.

The first portion of the current investigation describes an experiment similar to the high-fidelity static HRTF experiments conducted by Langendijk and Bronkhorst [7] and by Martin *et al.* [10], which was designed to ensure that our HRTF measurement and headphone equalization procedures are sufficient to produce localization comparable to free-field performance in a static display where the full HRTF was reproduced without modification for a short sound source. The second effort investigates whether the lossy HRTF processing necessary to implement an interactive virtual auditory display significantly degrades localization performance relative to a Full-HRIR implementation. The following sections detail the relevant methods, results and conclusions related to these efforts.

II. HRTF Measurement and Implementation

A. Facility and Equipment

All of the measurement and experimental sessions were conducted in the Auditory Localization Facility (ALF), part of the Air Force Research Laboratory at Wright-Patterson Air Force Base in Dayton, OH. The ALF consists of a large, floating-floor anechoic chamber with 4-foot fiberglass wedges covering all six surfaces, providing a nominal anechoic cutoff frequency of 63 Hz. In the center of the facility is a large seven-foot-radius geodesic sphere with 277 loudspeakers (Bose Acoustimass) positioned at its vertices as indicated in Fig. 1. This results in a average inter-loudspeaker distance of around 14 degrees. At the center of each speaker is a cluster of four individually controllable LEDs.

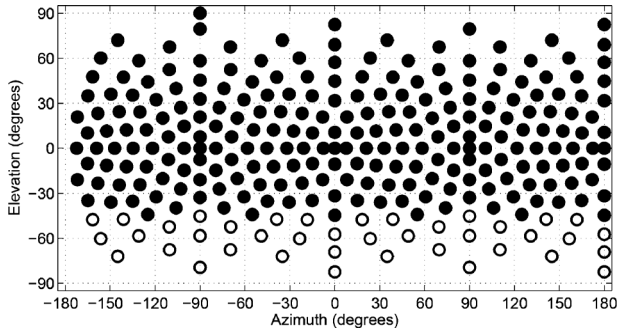

Fig. 1. Loudspeaker layout in the Auditory Localization Facility, Wright-Patterson Air Force Base, Ohio. Circles indicate the 277 loudspeaker locations used during the HRTF measurement. Closed circles represent the subset of 245 locations used during the localization task.

During both measurement and experimental sessions, the subjects stood on a 2-foot by 3-foot platform that is elevated so that the subject's interaural axis was aligned with the vertical center of the sphere. The ALF is equipped with an Intersense IS900, an ultrasonic six-degree-of-freedom tracking system that provides real-time location and orientation information about the position of the subject's head (provided by a sensor mounted to a headband) and the position and orientation of a hand-held response wand. The tracking system has a manufacturer specified performance of 0.75 mm in translation and less than 0.5° resolution in orientation.

All signal generation and processing was accomplished in Matlab (version 7.1) on a Dell Precision workstation outfitted with an RME RayDAT sound card with digital-optical I/O. The first 16 digital inputs and outputs of the sound card are connected to two Berhringer ADA8000 AD/DAs. The outputs are then passed to eight stereo Crown amplifiers, and any of the 16 amplified outputs can be dynamically routed to any one of the 277 loudspeakers via custom, computer-controlled switching hardware (Winntech).

B. Measurement Stimulus

Transfer function measurements were made using a series of linear sweeps as described in Müller and Massarani [25]. The sweeps were designed in the frequency domain to have a constant magnitude, G, and a group delay that increases linearly with frequency. This results in the discrete phase response described by (1), where N is the length of the discrete Fourier transform (DFT) for a single sweep and is chosen such that $N = 2^P$. For our system $P = 11$, giving $N = 2^{11} = 2048$ with a sampling rate of 44.1 kHz.

$$\phi[k] = \begin{cases} -\sum_{i=0}^{k} \frac{4\pi i}{N} & 0 \le k \le \frac{N}{2} \\ -\phi[N-k] & \frac{N}{2}+1 \le k \le N-1 \end{cases} \quad (1)$$

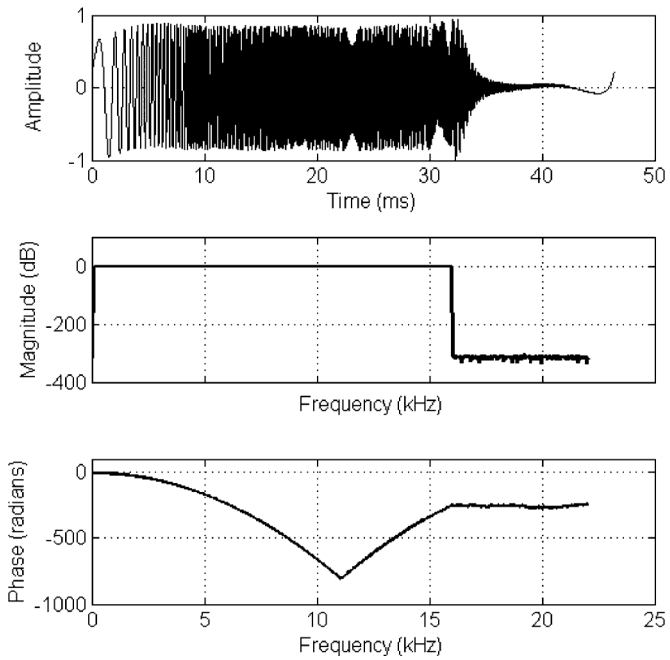

Fig. 2. An example of the linear sweep signal used to measure the response of the audio system, headphones, and the HRTF, shown as a time domain representation (Top), magnitude response (Middle), and phase response (Bottom).

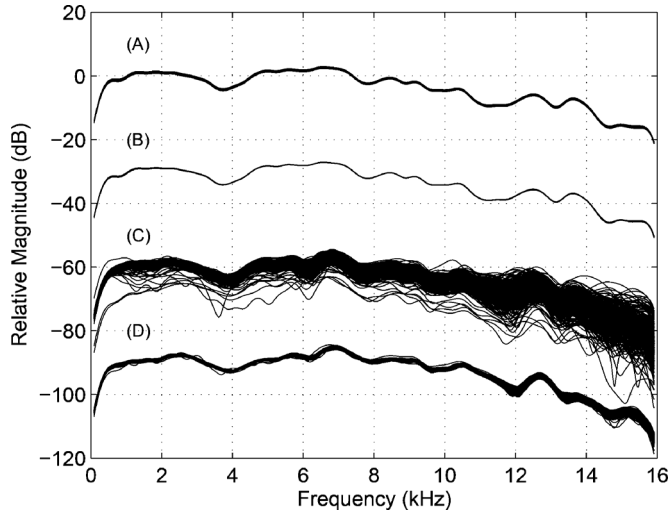

Fig. 3. Magnitude response curves for individual amplifiers and a reference speaker, before (A) and after (B) compensation. Magnitude response curves for individual speakers and a reference amplifier, before (C) and after (D) compensation. Response curves are offset by 30 dB for illustration purposes.

A time domain representation is then formed by taking the inverse DFT of the reconstructed discrete frequency response, $S[k]$, given in (2).

$$S[k] = Ge^{j\phi[k]} \tag{2}$$

Measurements were made with a test signal consisting of a train of seven sequtive consweeps, all of which had been band-pass filtered between 100 Hz and 16 kHz, and windowed with 20-ms onset and offset sine-squared ramps. A single, band-limited sweep is shown in both the time and frequency domain in Fig. 2.

After each sweep-train presentation the corresponding recording was first decomposed into individual sweeps by finding the first time index in the binaural recording for which the absolute value of the amplitude is greater than half of the maximum amplitude. Seven adjacent segments of length N were then removed from the recording and treated separately. The first and last segments were discarded to eliminate the effects of the onset and offset ramps and to allow the system to reach steady state prior to the actual measurements. The remaining five interior segments were used to make independent estimates of the resulting transfer function by dividing the DFT of the recorded segment by the DFT of the test signal. All five estimates were converted to impulse responses using an inverse DFT then averaged to obtain a single impulse response. The transfer functions were then windowed to remove any late reflections caused by the presence of the spherical loudspeaker array. The windowing was accomplished using 10-ms Hamming windows temporally-centered on the largest peak of each impulse response.

C. Loudspeaker Calibration

Before the start of the experiment, the response of the audio-signal pathway was calibrated using the sweep-train technique discussed above. The first part of the calibration procedure was designed to ensure that any differences in overall gain and frequency response across the left and right channels of the eight stereo amplifiers used to drive the loudspeakers in the ALF were compensated for the purposes of both the HRTF collection procedure and the presentation of free-field sounds in the facility. This part of the calibration procedure involves sequentially playing a sweep-train through each of the sixteen audio-amplifiers routed to a single reference loudspeaker. The loudspeaker signals are then recorded using an omni-directional probe-tube microphone (Etymotic ER-7C) mounted at the center of the sphere. The 16 lines shown in Fig. 3(A) depict the results of a typical amplifier calibration in the ALF. These lines are all very similar, which indicates that the gains and frequency responses of the sixteen amplifier channels were well matched in this case even without compensation. In order to minimize the amount of compensation required to match the amplifier responses, the amplifier with the frequency response closest to the median response is always selected as the reference, and frequency division is used to construct a 512-tap FIR compensation filter for each of the other fifteen amplifier channels that will best match that channel's frequency response to the reference channel. Fig. 3(B) depicts the magnitude response of all 16 amplifiers when remeasured using a sweep-train that had been pre-processed using the appropriate amplifier compensation filters and shows almost no residual variation between responses.

A technique similar to that described above is then used to develop compensation filters to match the responses of the 277 individual loudspeakers in the ALF. For each loudspeaker location, a sweep-train is pre-filtered with the appropriate amplifier compensation filter, played through the selected loudspeaker, recorded using the probe-tube microphone at the center of the sphere, and processed to derive a frequency response. The resulting loudspeaker that has the frequency response closest to the median response across all the loudspeakers is then treated as a reference loudspeaker, and a compensation filter is generated using frequency division to match all loudspeakers

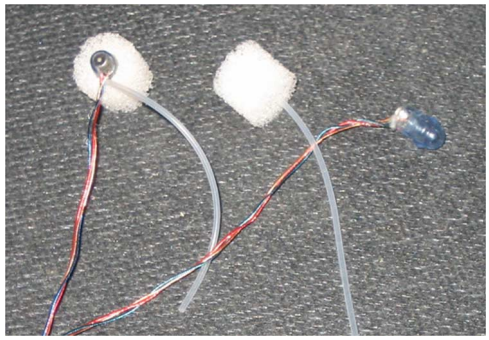

Fig. 4. Custom microphone assembly used for HRTF collection.

to the response of this reference loudspeaker. As a final step in the calibration procedure, the frequency response of each loudspeaker is remeasured after the compensation filter has been applied. Fig. 3(C) shows the individual magnitude responses of each loudspeaker in the ALF prior to compensation, and Fig. 3(D) shows the magnitude responses from all locations after application of the individualized speaker compensation filters. These results show that there is substantial variation in the raw speaker responses, but that most of this variation is eliminated by the application of the compensation filters. After compensation, the measured loudspeaker magnitude responses were within one decibel of the reference loudspeaker response when averaged across all loudspeaker locations for the frequency range from 100 Hz to 16 kHz, with a maximum difference for any location or frequency of 3 dB. All of the amplifier and loudspeaker compensation filters are stored electronically, and these filters are used to pre-process the test signals generated during the HRTF measurement to ensure that a consistent system response results from each measurement location. The frequency response curves for all loudspeakers after compensation were also saved electronically, and these curves were used to adjust the HRTFs for any residual differences in the frequency responses of the different loudspeakers as described in the following sections.

D. HRTF Measurement

Subjects initiated the beginning of the HRTF measurement procedure by aligning their head (via the head-slaved LED cursor) to a reference speaker and pressing a button on the hand-held tracker wand. Test stimuli were then presented sequentially from each of the 277 loudspeaker locations with approximately 250 ms between stimulus presentations. The subject's head position and orientation were monitored at all times, and if the subject's orientation changed by more than 3° during the presentation, that measurement was repeated. The subject's actual head position was used to calculate the head-relative position of each test signal presentation.

Binaural recordings were made during each test signal presentation using the blocked ear canal technique [2], utilizing a pair of custom-built in-ear microphone assemblies. The microphone assembly, shown in Fig. 4, consisted of a pair of miniature microphones (Knowles FG3329) embedded into small plastic caps from a Toslink cable. These microphones have a nominally flat frequency response between 100 Hz and 10 kHz and a manufacturer-specified noise floor of 30 dBA. The cap provided

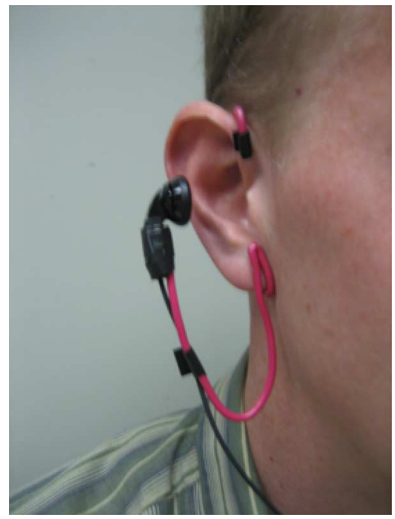

Fig. 5. Custom WireBud headphones used to enable *in situ* virtual presentation.

protection for the delicate solder connection on the rear of the microphone, while simultaneously enabling the microphone to be securely housed inside an oto-dam (Westone PROs), a small earplug used in the custom ear mold process. The oto-dams are disposable and available in several sizes allowing for a sanitary and customized fit for each subject. The dams also contain a pressure-release tube that can be used to remove the microphone assembly from the subject's ears without stressing the microphone wires.

E. Headphone Compensation

Each recording session also included the measurement of the headphone frequency response in order to generate compensation filters for a pair of *in situ* headphones. One of the goals of the study was to find a way to interleave real and virtual sound sources within the same block of trials. In order to achieve this goal, custom headphones were built based on the implementation of Langendijk and Bronkhorst [7] and Kulkarni and Colburn [26]. These headphones were designed to allow the presentation of broadband signals while having a minimal impact on the HRTF of the listener wearing them. These "WireBud" headphones, which are shown in Fig. 5, consist of a pair of Sennheiser ear-bud-style headphones (MX 470) mounted on a custom ear harness constructed from bent semi-rigid wire. The harness hooks easily over the top of the pinna, curves behind the ear, and gently squeezes the ear lobe, creating a stable distal positioning of the ear-bud, even during normal head movements. In both experiments *in situ* headphone correction filters were calculated as follows. The same chirp-based test signal was presented through the headphones and recorded binaurally after the HRTF measurements but before the microphones were removed from the listeners' ears. Compensation filters for each ear were then calculated by dividing 1 by the complex DFT then taking an inverse DFT. The inverse filters were then bandpass filtered between 500 Hz and 15 kHz, windowed with 1024 point (2.32 ms) Hamming windows centered on the peak of the impulse response. The 500 Hz low frequency cutoff was applied to eliminate a large resonance in the inverse filter due to poor low frequency response of the headphones. While this lower

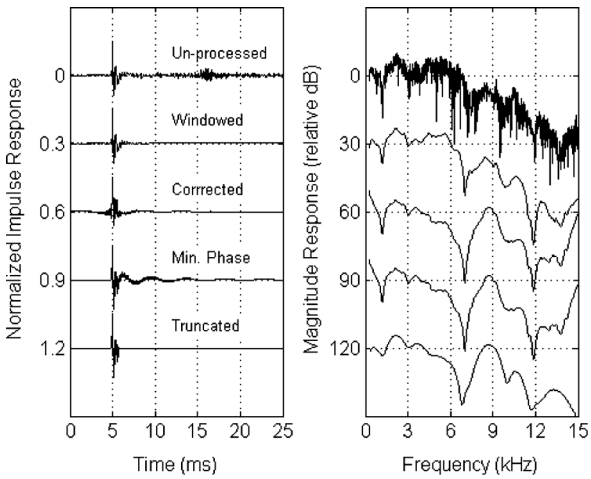

Fig. 6. Impulse response measurements at various stages of the HRTF collection and processing procedure (Left) along with the corresponding magnitude responses (Right). The first peak of each HRIR was aligned to the 5 ms time point for visual comparison purposes.

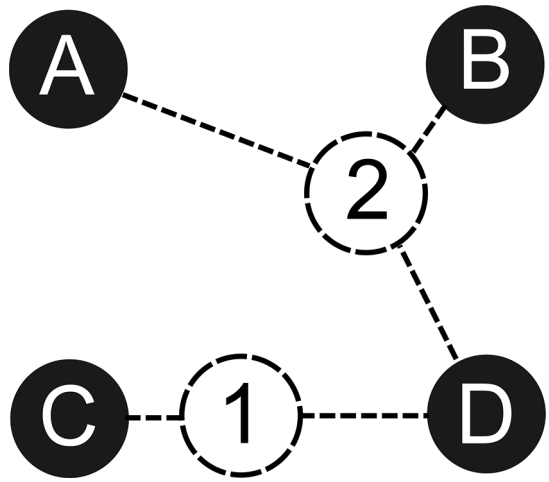

Fig. 7. Diagram showing variable N-nearest neighbor linear interpolation strategy. Solid circles represent measurement locations, empty dashed circles represent interpolation sites, and the dashed lines connect each interpolation site to its contributing measurement locations.

cutoff is atypical for traditional virtual auditory displays presenting real-world sounds, it is unlikely to differentially affect localization results across listening conditions.

F. HRTF Processing

Fig. 6 shows the results of each signal processing stage used to process the HRTFs for a typical left-ear HRTF located 45° to the right of the subject on the horizontal plane. Impulse responses (left panel) and their corresponding magnitude responses (right panel) are shown at each stage of the processing chain. The top signal of Fig. 6 depicts the raw measured head-related impulse response (HRIR). Note that this raw HRIR clearly shows the small amount of residual reverberation that exists in the Auditory Localization Facility as a result of sounds reflecting off the other speakers and the structure of the geodesic sphere. These reflections are at least 13 dB down from direct signal and arrive about 15 ms after the start of the direct impulse. Because they occur much later than the direct impulse, these reflections are easily removed by the application of a 10-ms Hamming window centered on the greatest peak in the HRIR. The magnitude response of this windowed HRIR, the second signal in Fig. 6, has a smooth frequency response that shows the spectral peaks and notches that result from spectral filtering by the head, torso, and pinna. The third trace from the top of Fig. 6 shows the signal after application of the inverse speaker and headphone filters. The frequency-response corrected HRTF is then converted to minimum phase (fourth signal in Fig. 6) and truncated to 256 samples using a rectangular window, resulting in the fully-processed HRTF shown at the bottom of Fig. 6.

For each HRTF location, an interaural time delay was also extracted using a method based on differences in group delay in the windowed HRIRs for the left and right ears (Fig. 6). For each ear, a linear least squares fit was applied to the low-frequency unwrapped phase response taken between 300 Hz and 1500 Hz. The slope of this line represents the negative of the average group delay over this frequency region. The ITD (in samples of delay) was then calculated directly from the derivative of this slope.

In order to make it possible to render the HRTFs at any desired location, the fully processed, truncated, minimum phase HRIRs and ITDs corresponding to each measured location were interpolated onto a grid with 5° resolution in both azimuth and elevation. This was done using a method based on linear nearest-neighbor interpolation, similar to that employed by Martin and McAnally [27]. The current study uses a slight variation on this method, as illustrated in Fig. 7. Here solid circles represent HRIR measurement locations, empty dashed circles represent interpolation sites, and the dashed lines connect each interpolation site to its contributing measurement locations. If an interpolation site lies directly in line between two measured locations (such as interpolation site 1) only those two HRIRs (C and D) are used in the interpolation. Otherwise, if an interpolation site does not lie directly between two measurement sites (such as interpolation site 2), the three closest measurement location are used (A,B,C). In both cases the interpolated HRIR is found by taking the weighted average of the contributing HRIRs. Each contributing HRIR's weight is inversely proportional to the geometric distance from the measurement site to the interpolation site (the length of the dashed line), and normalized so that all of the weights for a given interpolation site add to one. An identical process is used to interpolate the ITDs.

This interpolated grid of processed minimum-phase HRIRs and ITDs represents the level of processing required to render a virtual sound in headtracked VAD. However, as mentioned in the introduction, we also wanted to include an intermediate virtual audio condition that more closely matched the conditions used by [10] and [7] in previous experiments, which included a direct comparison between free-field and virtual sounds. These full-HRIRs were obtained simply by applying the headphone compensation filter directly to the full HRIR shown in Fig. 6. These Full-HRIRs are not suitable for interpolation between positions, so their use is limited to the synthesis of stationary sound sources at fixed locations relative to the listener. However, within those constraints, they would presumably generate

the greatest possible level of realism, reproducing almost exactly the same waveform in the listener's ear as the free-field sound source, including any room reflections or other artifacts that might be present in the facility.

G. Head-Tracked Virtual Presentation

In order to allow for continuous virtual presentation, the grid of interpolated HRTFs was loaded into SLAB (slab3d.6.1), an open source real-time software-based acoustic scene renderer developed by NASA. Details concerning SLAB's internal processing can be found in [28], [29]. In brief, the SLAB software handles the real-time signal processing required to continuously update the virtual sound source's position with respect to the listener's dynamically moving head. This is accomplished with a headtracker interface that provides access to changes in the listener's head position detected by the IS900 headtracker. At each time frame SLAB then selects the closest HRTFs and ITDs based on the current head position and linearly interpolates them to approximate the current appropriate values at the current head-relative location. SLAB also includes a temporal crossfading process to ease rough transitions between neighboring time frames.

The SLAB engine was controlled via a custom-built command server with a MATLAB interface. Once initiated, SLAB reads the sound appearing at a designated input of the sound card, processes the sound using the loaded HRTF, and delivers the binaural signal to the output of the sound card. Internal SLAB audio processing has a frame update rate of approximately 120 Hz, while head position data was updated via the server at approximately 50 Hz providing an end-to-end system latency of approximately 50 ms. It is important to note that this latency is less than the critical latencies discussed in [16]–[18], [30].

III. Experiment I: Direct Comparison of Real and Virtual Sounds

A. Methods

In the initial validation experiment, our goal was to develop a localization procedure that would allow us to interleave real and virtual sources and do a direct comparison of how well listeners were able to localize and distinguish between each type of sound. In order to achieve this, we conducted both the HRTF measurements and the perceptual experiment with the WireBud headphones present on the listeners ears. The perceptual evaluations conducted in Experiment 1 consisted of a combined localization and subjective evaluation task.

1) Subjects: Seven subjects participated in the experiment (4 males, 3 females) all between the ages of 18 and 25. All seven subjects had audiometric thresholds within 15 dB HL between 125 Hz and 8 kHz. All subjects had participated in previous free-field localization experiments using the same response technique, but had not participated in any virtual localization tasks. No additional training was given to the subjects before the start of the experiment.

2) Stimuli: All stimuli were created from independent samples of random white Gaussian noise, which was bandpass filtered between 500 Hz and 14 kHz. On a given trial the stimulus was presented randomly by one of three presentation types (Free-Field, Full HRIR, Head-Tracked). The free-field presentation consisted of filtering the noise stimulus with the speaker correction filter corresponding to the target location and presenting the stimulus through an ALF loudspeaker. The 'Full HRIR' presentation consisted of filtering the noise stimuli with a full 2048-tap headphone-corrected HRIR corresponding to the target location and presenting the stimulus through headphones. The 'Head-Tracked' presentation consisted of filtering the noise stimulus using the SLAB software to generate a virtual source at the target location. Therefore, 'Full HRIR' represents the best available virtual presentation with no additional processing and no head-tracking, while 'Head-Tracked' constitutes a truncated minimum-phase implementation with head-tracking. For both the free-field and head-tracked conditions, the stimulus duration was set to either 250 ms (Burst) or 10 seconds (Continuous). Because the full-HRIR condition did not employ head-tracking, stimulus duration was limited to the 250-ms burst condition to eliminate the possibility of head-movement during presentation. The presentation order of all conditions was randomized within a block, and the stimulus in each trial was presented at a level that was randomly selected to be somewhere between 55 and 65 dB SPL. Target locations were randomly chosen from a set of 245 ALF loudspeaker locations. This set of locations was selected to eliminate locations below -45° in elevation and to ensure that all the virtual sound presentations occurred in locations that could correspond to the locations of actual physical speakers. The experimental conditions were selected randomly for each trial, with a total of 60 trials per subject in each of the short duration conditions, and 30 trials per subject in each of the long-duration conditions.

During the first experiment the WireBuds were positioned prior to the HRTF collection so that any acoustic effects caused by their presence would be captured in the HRTF recordings. In the second experiment, the WireBuds were positioned on the subject after the HRTF recordings were complete, with care to avoid disturbing the in-ear microphones.

3) Procedure: Prior to the start of each experimental session, the position of the headband on the subject's head was adjusted by 1) instructing the subject to face the speaker in the front of the sphere (i.e., 0° Az, 0° El); 2) turning on an LED cursor that was continuously updated to illuminate the speaker located directly in front of the measured location of the headtracking sensor on the headband; and 3) having the subject adjust the location of the headband until the LED cursor illuminated the front speaker. Once in place, the headtracker was never taken off or adjusted for the entire collection-presentation session. In each half-hour experimental session, a new set of HRTFs was measured for the current subject. The entire *in situ* measurement procedure, including the headphone correction measurement, took approximately five minutes to complete, immediately followed by three, 40-trial experimental blocks. Each trial consisted of a single stimulus presentation that required both a localization judgment and a subjective evaluation response. Subjects initiated each trial by centering their head using the head-slaved LED cursor and pulling a trigger on a hand-held wand. After the stimulus was presented, the subject made a localization judgment by moving an LED cursor, slaved to the tracked hand-held wand, to the location of a speaker and pressed the trigger. Subjects were then asked to indicate whether they thought the sound

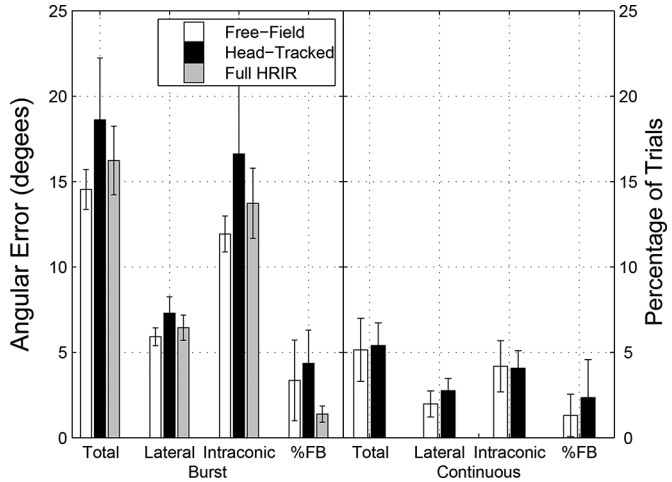

Fig. 8. Localization results for Experiment I with short duration, "Burst" stimuli (left) and long duration "Continuous" stimuli (right). Results are averaged over all subjects in terms of the total angular error and its lateral and intraconic components as well as the percentage of trials which resulted in a front-back reversal. All error bars represent 95% confidence intervals.

had originated from one of the actual speakers or from a virtual source by selecting an appropriate button on the wand. All subjects were familiar with the localization task and no instructions were given discouraging or promoting exploratory movements. Once a decision was made, subjects received feedback about the correct location of the target, but not whether it was a real or virtual source.

B. Results

The results show generally good agreement between free-field stimuli and both types of virtual stimuli. The left panel of Fig. 8 shows the localization results for the short duration "burst" condition. There was a general trend for total angular error to be best in the free-field condition (15°), slightly worse in the full-HRIR condition (16°), and worst in the head-tracked condition (18°), which represents a statistically significant difference (as indicated by the results of a one-factor within-subjects ANOVA, $\mathrm{F}(2,12) = 4.05$, $\mathrm{p} = .045$). Total angular error is calculated here as the great circle error between the target and response locations. Error differences seemed to be concentrated in the up-down and front-back dimensions rather than the left-right dimension. This is shown in the second two sets of bars in each panel of Fig. 8, which show the localization error broken down into the errors in the lateral dimension and the errors within a cone of confusion (intraconic). These errors are calculated by transforming the target and response directions into the interaural-polar coordinate system and taking the absolute difference between the target and response for each component. The last set of bars in Fig. 8 show the percentage of trials in which a front-back reversal occurred (where a front-back reversal is defined as a trial in which the target and response locations were in different front-back hemispheres, and neither location was within 15 degrees of the frontal plane). The percentage of front-back reversals was relatively low across all conditions ($< 5\%$), and there were no statistically significant differences across the conditions (one-factor within-subjects ANOVA, $\mathrm{F}(2,12) = 2.017$, $\mathrm{p} = .176$). This suggests that the differences in overall localization performance across the conditions cannot be fully explained by additional front-back confusions in the virtual listening conditions.

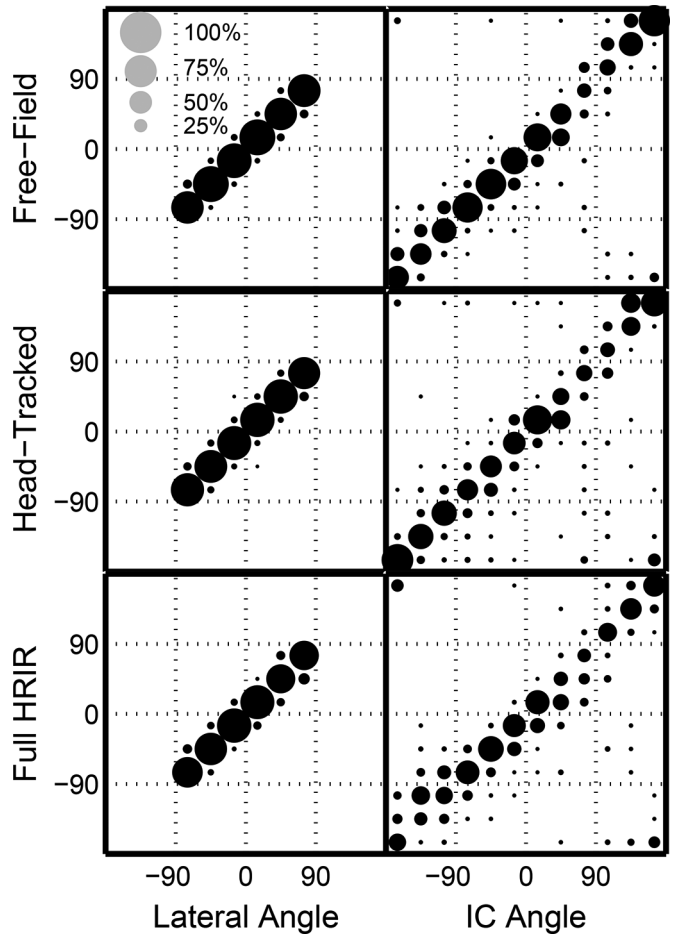

Fig. 9. Response histograms for each experimental condition (rows). The size of the filled circles represents the percentage of responses which occurred in the corresponding bin for each target bin. The left column shows the lateral angle of the response for each target lateral angle, all locations were binned into 30° sections. The right column shows the same data in terms of target and response intraconic angle.

The right panel of Fig. 8 shows performance in the "continuous" sound conditions. Continuous stimuli were left on for a maximum of 10 seconds or until the subject made a response. Subjects typically responded within the first 3 seconds of the stimulus, so the 10-second duration limit was not an issue. In this condition, there were no significant differences between the overall angular localization errors observed with the head-tracked virtual sounds and the free-field sounds (one-factor within-subjects ANOVA, $\mathrm{F}(1,12) = 0.424$, $\mathrm{p} = 0.539$). Both resulted in very small average angular errors on the order of 5–6°.

Fig. 9 shows the distribution of localization responses for each of the presentation types and the burst duration. Response locations are plotted as a function of target location in terms of their lateral and intraconic (IC) angles. As can be seen, there is a very good correspondence between target and response direction independent of presentation type or target location. In general performance seems to be best in front of the listener and worst near the frontal plane.

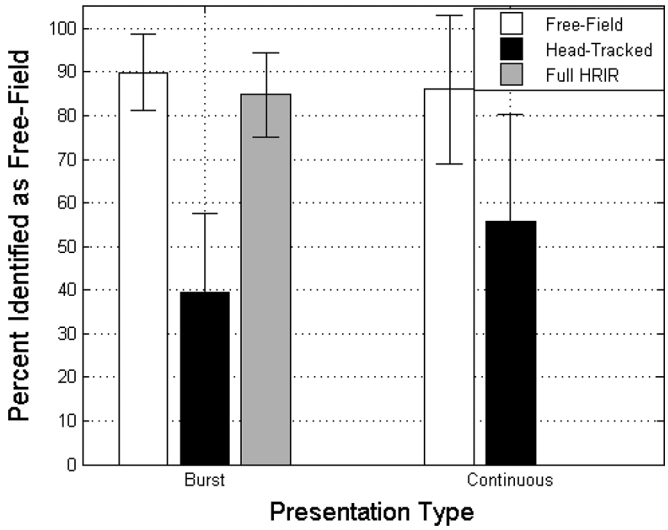

Fig. 10. Average subjective evaluation score for Experiment I taken over all subjects. Score indicates the percentage of trials in which the subjects thought the source was from a loudspeaker. All error bars represent 95% confidence intervals.

Fig. 10 depicts the results of the subjective evaluation. This figure shows the percentage of trials in which the subject indicated that the source appeared to come from an actual loudspeaker, as opposed to a virtual sound generated over headphones. Only a small difference was found between the virtual Full-HRIR and free-field conditions, suggesting that the virtual rendering in the full-HRIR condition was perceptually realistic and these virtual sounds were almost indistinguishable from the real sound sources. The head-tracked virtual sounds did not achieve the same level of realism as the full-HRIR sounds, even when long duration stimuli made exploratory head movements possible; the results show that listeners mistakenly identified the head-tracked virtual sounds as coming from a real loudspeaker in almost half the trials, including the continuous trials.

C. Discussion

The results of Experiment I show that there was essentially no difference in localization performance or in perceived realism between actual sounds generated by physical loudspeakers in the ALF facility and virtual sounds rendered using the full-HRIR synthesis technique. This result is consistent with the results of both Langendijk and Bronkhorst [7] and Martin *et al.* [10], and it confirms that the rapid HRTF measurement procedure used in the ALF is capable of accurately measuring individual HRTFs that may be used to synthesize high-fidelity virtual sounds through a set of custom-designed headphones.

Furthermore, the results of our head-tracked virtual condition demonstrate that careful processing can be used to generate short-duration minimum-phase HRTFs that are suitable for interpolation and real-time rendering in an immersive head-tracked virtual audio display with little reduction in localization performance. This suggests that the temporal windowing, conversion to minimum phase, and truncation of the HRTF illustrated in Fig. 6 was successful in converting the HRTF into a form suitable for linear interpolation without substantially degrading the cues necessary for accurate localization.

It is also necessary to discuss the possibility that the equivalent performance in the Full-HRIR and head-tracked conditions could reflect an offsetting effect between a degradation in performance caused by the minimum phase processing and an *improvement* in performance caused by the elimination of reflections that may have degraded performance in the Full-HRIR (and free-field) conditions. This possibility was evaluated in a pilot experiment that examined the effect of presenting a 'quasi-anechoic' stimulus in which a temporal window was applied to the HRTF to eliminate the late reflections shown in the unprocessed HRTF in Fig. 6 without modifying the early portion of the HRTF. There was no measurable difference in localization performance between the full HRTF and the quasi-anechoic HRTF. Thus, although Laitinen *et al.* [22] showed that small changes in update rate and tracker bias can have a similar effect on subjective quality ratings of head-tracked virtual audio, these changes are likely to negatively affect localization performance as well. Therefore it is more likely that the differences in the subjective experiment shown in Fig. 10 are related to the fact that the head-tracked virtual sounds did not contain the late reflections that were present in both the free-field and full HRIR stimuli. Since these sounds were all interleaved randomly, this caused the head-tracked sounds to have a distinctly different coloration than the other two types of sounds, and consequently, the subjective evaluations for these sounds indicated that they were much less likely to be identified as coming from a real loudspeaker than a true free-field sound for both the short and long duration stimuli.

Another possible concern with the results of Experiment I is that both the real and virtual localization scores may have been negatively impacted by the distortions introduced by the presence of the Wire-Bud headphones during HRTF collection and free-field sound playback. In order to address this issue, a second experiment was conducted where the HRTFs were measured without the headphones in place, and the headphones were placed on the head and their compensation filters were measured after the HRTF collection but prior to the removal of the in-ear microphones. This made it possible to examine real and virtual localization performance without any possible contamination of the HRTF from the presence of the Wire-Bud headphones, but made it necessary to collect free-field and virtual localization data in different blocks of trials.

IV. Experiment II: Comparison of Real and Virtual Sounds in Separate Blocks

A. Methods

The second experiment was designed to address the possibility that the physical presence of the WireBuds during HRTF collection could affect localization cues both for free-field stimuli and virtual stimuli. The experimental procedure was a replication of the first experiment with two modifications. First, the WireBud headphones were not worn during the HRTF measurement process but rather were donned immediately afterwards in order to allow for an *in-situ* headphone compensation before virtual conditions. Secondly, the free-field conditions were collected in a separate set of experimental sessions to avoid the need to remove the headphones between

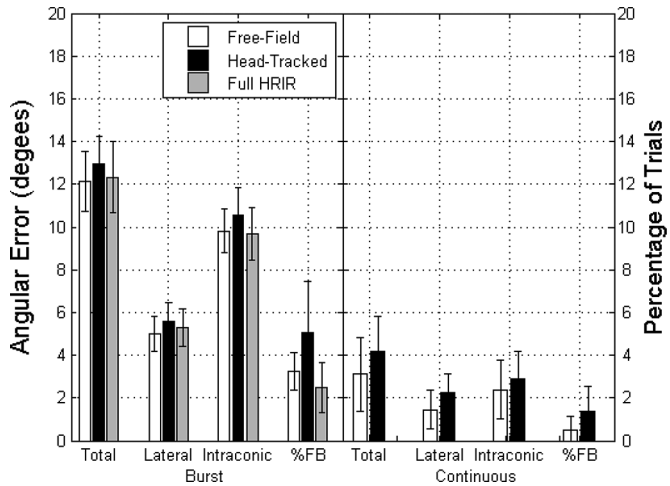

Fig. 11. Localization results for Experiment II with short duration, "Burst" stimuli (left), and long duration, "Continuous" stimuli (right). Results are averaged over all subjects in terms of the total angular error and its lateral and intraconic components as well as the percentage of trials that resulted in a front-back reversal. All error bars represent 95% confidence intervals.

conditions. The experiment utilized the same set of seven subjects (4 males, 3 females) from Experiment I, as well as the same distribution of stimulus conditions (60 trials per subject in each short duration stimulus condition, 30 trials per subject in each long duration stimulus condition).

B. Results

Fig. 11 shows the localization results for the short duration burst stimuli in Experiment II. A comparison between these results and the results from Experiment I shown in Fig. 8 clearly shows that performance in the "burst" condition of Experiment II was superior to performance in the "burst" condition of Experiment I in all dimensions, with the overall angular error dropping from roughly 15° to roughly 12° in the free-field condition. While secondary procedural results like blocking on condition and experience-based learning could account for some of the effects, the performance difference was likely the result of spectral distortions that were caused by the presence of the WireBud headphones over the listener's ears during the HRTF measurements and free-field localization trials of Experiment I. Note, however, that performance in the "continuous" conditions of the experiment was not significantly different between Experiments I and II. This suggests a ceiling effect related to the fact that the listeners who had access to exploratory head motion cues were able to accurately localize the continuous sounds even when the localization cues were distorted by the presence of the WireBud headphones.

The results in Fig. 11 also show that, once the distortions related to the WireBud headphones were removed, the differences in performance between the free-field, head-tracked, and full-HRIR conditions of the experiment were reduced to the point where they were essentially negligible. Although the data show that the mean errors tended to be slightly larger in the head-tracked condition than in the free-field condition, the differences in angular error were typically less than one degree and they were not statistically significant (1-factor within-subjects ANOVA, $F(2, 12) = 0.620$, $p = .557$).

V. Discussion

In comparison to other studies that have examined either free-field or virtual localization, the localization errors measured in this study were quite low. The benchmark for determining the quality of any virtual sound presentation is a comparison to performance in the free field in a comparable task. Thus, it makes sense to first examine performance in the free-field condition in the context of other studies that have used similar methods to measure free-field localization performance. In this regard, both the short and long stimulus conditions of Experiment II compare favorably with other studies in the literature. The free-field localization performance achieved in that experiment was comparable to, and often better than, the localization performance reported in other studies that have used broadband sources distributed across the entire range of possible azimuth and elevation sources locations. For short duration stimuli, the 12° average angular error seen in Experiment II effectively matches average angular error reported by Martin *et al.* [10]. When the results of both studies are analyzed to resolve all front-back reversals, Experiment II had an overall angular error of 11° compared to the 10° error reported in the Martin *et al.* study. However, the Martin *et al.* study used only a very small number (three) very well-trained subjects, and its results are substantially lower than most other reports of overall angular error in the literature, including the 16° RMS angular error reported by Middlebrooks [31], the 14° mean angular errors reported by Wightman and Kistler [11], and the 15.6° mean angular error reported in a recent meta-study that combined data from more than 82,000 free-field localization trials collected across 161 listeners in five different laboratories using a variety of different response procedures [32]. In part, we believe the lower localization errors reported here may have been related to our response technique, which required listeners to identify the apparent location of the sound by moving an LED cursor to the discrete loudspeaker location that most closely matched that perceived location.

Fewer studies have looked at free-field localization errors for continuous sound sources. Bronkhorst [4] used a "closed-loop" head-pointing response for localization of continuous broadband sources, and found an average angular error of approximately 9°. This is slightly higher than the 5–6° average angular error obtained for long-duration stimuli in our study. Similarly, Sandvad [16] reported average azimuth and elevation errors of 7° and 9°, respectively. While it is difficult to directly compare absolute localization performance across different experiments, the fact that the free-field performance seen in our study is comparable with other published results for both short and long duration stimuli gives us confidence that comparisons with virtual performance will not be limited by floor effects due to poor free-field performance.

In terms of virtual localization performance, our results are lower than most reported measures of localization performance in the literature. The best virtual localization performance results we are aware of are those from the study by Martin *et al.* [10], which also found no statistically significant differences in performance in terms of localization performance was observed between free-field sources and virtual sources generated with full-HRIR digital filters. However, as mentioned before, that study was limited to static virtual sounds generated by non-interpolated full-length HRTFs.

Localization errors reported for other virtual audio displays that have used individualized HRTFs and interactive head-tracking have been substantially larger. For example, Pec *et al.* [33] reported localization errors for both static and dynamic sources that were twice as large as those measured for a free-field source. Djelani *et al.* [24] investigated both short duration and long duration head-tracked stimuli using a number of different response techniques, however none of their average localization errors were less than 19° for short duration stimuli and 10° for long-duration stimuli. Similarly, Wenzel *et al.* [17] reported localization errors for long-duration (8 s) virtual stimuli of approximately 25°. These average localization errors are substantially larger than both the free-field and virtual head-tracked stimuli seen in the current study (approximately 4° for long duration stimuli and 12° for short duration stimuli).

Recently, Majdak *et al.* [34] used a virtual audio display condition that incorporated a head-mounted visual display that projected a cross-hair on the visual scene at the location the listener was indicating with a hand-held response wand (similar to the LED cursor used in this experiment). Localization performance in this virtual display system was validated by comparing it directly to the virtual localization results reported by Middlebrooks [31] and showing that, in the best condition, virtual performance was comparable to that achieved in the Middlebrooks experiment (for example, quadrant errors occurred in 7.8% of trials versus 7.7% in the Middlebrooks study). However, this comparison failed to note that subjects in the Middlebrooks study performed substantially better in the free-field (where there were only 4.4% quadrant errors) than they did with the virtual audio stimuli. The comparison also did not account for the fact that the Majdak study used substantially longer stimuli than the Middlebrooks experiment (500 ms vs 250 ms) that could have provided additional benefit. Thus, while virtual localization performance in the Majdak study was quite respectable, we would argue that it is likely that it did not approach the level that would likely be achieved if the same subjects were asked to perform the same localization study in the free field.

When taken together with the results of these earlier experiments that have measured localization performance with real and virtual sources, we believe that the results of the current experiment provide strong validation for the use of the *in-situ* HRTF measurement and rendering technique described here as a tool for evaluating the effects that small changes in HRTF have on virtual localization performance. Localization performance in the head-tracked condition of Experiment II was essentially identical to performance in the free-field condition, where performance was as good or better than any published data showing the best possible localization performance human listeners can achieve in the free field. This indicates that performance in both the real and virtual conditions was limited by the performance of human sound localization, rather than by any artificial limits resulting from errors in the response technique. Thus, while we have no way of knowing if the level of care we have taken in measuring, processing, and rendering the HRTFs used in this experiment is truly *necessary* to achieve near-equivalent real-virtual performance, we are fairly confident that it is *sufficient* to obtain reasonably accurate localization performance in a virtual display. Thus, we believe our methods can serve as a guide for investigators who are interested in setting up a high-fidelity virtual audio display for research applications.

VI. Conclusion

The results of this study provide evidence that head-tracked virtual audio systems can provide real-virtual equivalent localization performance despite the limitations of practical implementation such as system latency, HRTF interpolation errors, and filter truncation errors. These experiments show that, if enough care is taken, it is possible to generate short-duration, minimum phase HRTFs that effectively capture the relevant spatial cues used for sound localization and lead to localization performance that is essentially equivalent to the free-field, even with brief stimuli, which make the use of dynamic head-motion cues impossible. This result is important for conducting studies to examine the relative importance of different acoustic cues contained in the HRTF, because it allows these studies to start with a simplified minimum-phase baseline HRTF, which is relatively easy to parametrically adjust, rather than a much more complicated full-HRIR, which includes a lot of spectral detail and phase distortions that may be difficult to systematically modify in a parametric way. The ability to achieve free-field-equivalent localization performance with a simplified HRTF is important, because it ensures that the experimenter will be able to successfully detect the impact of HRTF modifications that result in *any* degradation in localization performance, and not just those distortions that make localization performance worse than a baseline that is already degraded due to fundamental inaccuracies.

Another interesting aspect from the results of these experiments is that it was able to show equivalent real-virtual localization performance in a visual context where the potential free-field sound source locations were clearly visible to the listener. The ALF is a unique facility, because it uses a configuration that has loudspeakers distributed over the entire surface of a sphere. In contrast, virtually all prior localization studies have been conducted in facilities that use a hoop or ring to generate free-field sounds. The direct consequence of this difference is that all prior experiments that have compared real and virtual sound source localization in the same facility have had to do so in a context where the actual sound sources were hidden from the view of the listener, either by conducting the experiment in darkness or by placing an acoustically transparent screen between the listener and the loudspeakers. Visual context is likely to play an important role in the perception of the spatial properties of sounds, and in order to be successful a virtual audio display must be able to convincingly simulate sounds originating from sound source locations that are visible to the listener as well as those that are obscured from view. The results of this experiment show that this is definitely possible, and that in cases where a listener can listen to a continuous sound source and make as many exploratory head movements as desired, it is possible to produce a virtual sound source that is frequently confused with the corresponding free-field source.

A final important point that needs to be emphasized is that the relatively high level of virtual localization performance achieved in the current experiment relative to previous investigations is likely due to the *in-situ* HRTF measurement and

presentation procedure, a technique previously used only for rigorous real-virtual discrimination tasks like Langendijk and Bronkhorst [7]. Because the ALF allows us to implement this procedure fast enough and reliably enough to allow a new HRTF to be measured in each experimental session, this procedure makes it possible to create headphone compensation filters for a specific placement of the headphones used to render the virtual sounds and then to run a localization experiment using those headphones without ever removing the headphones from the listener's head. This eliminates the possibility that HRTFs used to render sounds in a virtual localization experiment might be contaminated by inconsistencies in the placement of the headphones. In environments where it is not feasible to use the *in-situ* measurement method, and it is necessary to remove and replace the headphones between the time when the HRTF is measured and the time the virtual source is presented, one might expect somewhat less accurate localization of virtual sounds. Even in this case, however, it would be reasonable to expect the HRTF measurement and processing technique outlined here to be sufficient to obtain the best possible level of performance within the constraints caused by variability in the frequency response of the headphones.

ACKNOWLEDGMENT

The authors would like to thank Billy Swayne for his careful assistance during the initial testing and data collection process as well as the Air Force Office of Scientific Research (AFSOR).

REFERENCES

[1] S. Mehrgardt and V. Mellert, "Transformation of the external human ear," *J. Acoust. Soc. Amer.*, vol. 61, pp. 1567–1576, 1977.

[2] H. Møller, M. Sørensen, D. Hammershøi, and C. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, vol. 43, pp. 300–321, 1995.

[3] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 93, pp. 111–123, 1993.

[4] A. W. Bronkhorst, "Localization of real and virtual sound sources," *J. Acoust. Soc. Amer.*, vol. 98, pp. 2542–2553, 1995.

[5] H. Møller, M. Sørensen, C. B. Jenison, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *J. Aud. Eng. Soc.*, vol. 44, pp. 451–469, 1996.

[6] D. Hammershøi and H. Møller, "Methods for binaural recording and reproduction," *Acta Acust. United With Acust.*, vol. 88, no. 3, pp. 303–311, 2002.

[7] E. H. A. Langendijk and A. W. Bronkhorst, "Fidelity of three-dimensional-sound reproduction using a virtual auditory display," *J. Acoust. Soc. Amer.*, vol. 107, no. 1, pp. 528–537, Jan. 2000.

[8] P. Zahorik, F. L. Wightman, and D. J. Kistler, "On the discriminability of virtual and real sound sources," in *Proc. IEEE ASSP Workshop Applicat. Signal Process. Audio Acoust.*, 1995, pp. 76–79.

[9] W. M. Hartman and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Amer.*, vol. 99, pp. 3678–3688, 1996.

[10] R. Martin, K. McAnally, and M. Senova, "Free-field equivalent localization of virtual audio," *J. Aud. Eng. Soc.*, vol. 49, pp. 14–22, 2001.

[11] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Amer.*, vol. 85, pp. 868–878, 1989.

[12] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. Amer.*, vol. 106, pp. 1493–1510, 1999.

[13] F. L. Wightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Amer.*, vol. 105, pp. 2841–2853, 1999.

[14] H. Wallach, "The role of head movements and vestibular and visual cues in sound localization.," *J. Exper. Psychol.*, vol. 27, p. 339, 1940.

[15] I. Pollack and M. Rose, "Effect of head movement on the localization of sounds in the equatorial plane," *Percept. Psychophys.*, vol. 2, pp. 591–596, 1967.

[16] J. Sandvad, "Dynamic aspects of auditory virtual environments," in *Proc. Audio Eng. Soc. Conv. 100*, May 1996 [Online]. Available: http://www.aes.org/elib/browse.cfm?elib=7547

[17] E. M. Wenzel, "Effect of increasing system latency on localization of virtual sounds," in *Proc. Audio Eng. Soc. Conf.: 16th Int. Conf.: Spatial Sound Reprod.*, 1999, Audio Eng. Soc..

[18] D. S. Brungart, B. D. Simpson, R. L. McKinley, A. J. Kordik, R. C. Dallman, and D. A. Ovenshire, "The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources," in *Proc. ICAD 04—10th Meeting Int. Conf. Auditory Display*, Sydney, Australia, Jul. 6–9, 2004.

[19] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, A. V. Oppenheim, Ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[20] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, and H. Nironen, "Techniques and applications of wearable augmented reality audio," in *Proc. Audio Eng. Soc. Conv. 114*, Mar. 2003 [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12495

[21] M. Noisternig, T. Sontacchi, T. Musil, and R. Höldrich, "A 3D ambisonic based binaural sound reproduction system," in *Proc. AES 24th Int. Conf. Multichannel Audio*, 2003.

[22] M.-V. Laitinen, T. Pihlajamäki, S. Lösler, and V. Pulkki, "Influence of resolution of head tracking in synthesis of binaural audio," in *Proc. Audio Eng. Soc. Conv. 132*, Apr. 2012 [Online]. Available: http://www.aes.org/elib/browse.cfm?elib=16261

[23] R. S. Bolia, W. R. D'Angelo, and R. L. McKinley, "Aurally aided visual search in three-dimensional space," *Human Factors: J. Human Factors Ergon. Soc.*, vol. 41, no. 4, pp. 664–669, 1999.

[24] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert, "An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization," *Acta Acust. United With Acust.*, vol. 86, no. 6, pp. 1046–1053, 2000.

[25] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, pp. P.443–471, 2001.

[26] A. Kulkarni and H. S. Colburn, "Role of spectral detail in sound-source localization," *Nature*, vol. 396, pp. 747–749, 1998.

[27] R. Martin and K. McAnally, "Interpolation of head-related transfer functions," Air Operations Division Defence Science and Technology Organisation, Tech. Rep., 2007.

[28] J. D. Miller and E. M. Wenzel, "Recent developements in SLAB: A software-based system for interactive spatial sound synthesis," in *Proc. Int. Conf. Auditory Display*, Kyoto, Japan, 2002.

[29] J. D. Miller, 2013 [Online]. Available: http://slab3d.sonisphere.com/ [Online]. Available: http://humansystems.arc.nasa.gov/SLAB/

[30] S. Yairi, Y. Iwaya, and Y. Suzuki, "Influence of large system latency of virtual auditory display on behavior of head movement in sound localization task," *Acta Acust.*, vol. 94, pp. 1016–1023, 2008.

[31] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," *J. Acoust. Soc. Amer.*, vol. 92, pp. 2607–2624, 1992.

[32] V. Best, D. Brungart, S. Carlile, C. Jin, E. A. Macpherson, R. L. Martin, K. McAnally, A. T. Sabin, and B. D. Simpson, "A meta analysis of localization errors made in the anechoic free-field," in *Principles and Applications of Spatial Hearing*. Hackensack, NJ, USA: World Scientific, 2011, pp. 14–23.

[33] M. Pec, M. Bujacz, and P. Strumiłło, "Personalized head related transfer function measurement and verification through sound localization resolution," in *Proc. 15th Eur. Signal Process. Conf.*, 2007, pp. 2326–2330.

[34] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attent., Percept., Psychophys.*, vol. 72, no. 2, pp. 454–469, 2010.

Griffin D. Romigh received the B.S. degree in biomedical engineering from Wright State University in 2009, an M.S. in 2011, and Ph.D. in 2012, both from Carnegie Mellon University in electrical and computer engineering. He is currently an Electrical Engineer and Program Manager for the Battlespace Acoustics Branch at the Air Force Research Laboratory, Wright-Patterson Air Force Base, OH.

His research focuses on the application of signal processing and machine learning techniques to better understand human spatial auditory and speech perception, particularly how to efficiently model and estimate individualized head-related transfer functions. He also conducts applied research on the application of spatial audio and language technology to improve communication and situation awareness in complex military environments. Dr. Romigh is a member of IEEE, Acoustical Society of America, and former U.S. Department of Defense SMART Scholar.

Douglas S. Brungart received the B.A. degree from Wright State University in 1993, an S.M. in 1994, and Ph.D. in 1998, both from the Massachusetts Institute of Technology in electrical engineering all in computer science and electrical engineering. He is currently the Chief Scientist at the Army Audiology and Speech Center at Walter Reed NMMC and Director of Research for the U.S. Dept. or Defense Hearing Center of Excellence.

His research addresses aspects of basic and applied research in the areas of spatial hearing, hearing impairment, speech perception, and hearing protection. He is most well-known for his work on informational masking in multi-talker speech displays and his work on near-field spatial hearing. Dr. Brungart is a member of IEEE and a Fellow of the Acoustical Society of America.

Brian D. Simpson received his A.B. in psychology from Washington University in 1995, and his M.S. and Ph.D. in psychology with an emphasis on Human Factors from Wright State University in 2002 and 2011, respectively. He is currently a Research Psychologist and the Technical Advisor for the Battlespace Acoustics Branch at the Air Force Research Laboratory, Wright-Patterson Air Force Base, OH.

Dr. Simpson's research has focused on the investigation of peripheral and central processes that mediate speech perception and spatial hearing in multi-source acoustic environments, spatial auditory attention, and the development of auditory displays to support performance in complex task environments. He is a member of the Acoustical Society of America and the Human Factors and Ergonomics Society.