Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005

THE DETECTABILITY OF HEADTRACKER LATENCY IN VIRTUAL AUDIO DISPLAYS

Douglas S. Brungart
Brian D. Simpson

Air Force Research Laboratory
2610 Seventh Street
WPAFB, OH 45433
douglas.brungart@wpafb.af.mil
brian.simpson@wpafb.af.mil

Alexander J. Kordik

General Dynamics
Dayton, OH
alex.kordik@wpafb.af.mil

ABSTRACT

A critical engineering parameter in the design of interactive virtual audio displays is the maximum amount of latency that can be tolerated between the movement of the listener's head and the corresponding change in the spatial audio signal presented to the listener's ears. In this study, subjects using a virtual audio display were asked to detect the difference between a control stimulus that had the lowest possible latency value for the display system (11.7 ms) and a test stimulus that had an artificially increased headtracker latency ranging from 36 to 203 ms. In a standard listening configuration with only a single virtual sound source, the results show that typical listeners are unable to reliably detect the presence of headtracker latencies smaller than 80 ms, and that even the best listeners are unable to detect changes smaller than 60 ms. However, the addition of low-latency reference tone at the same location of the target signal decreases the minimum threshold for latency detection by about 25 ms. This result suggests that augmented reality systems may require headtracker latencies smaller than 30 ms to ensure the delays are undetectable to all users in all listening environments.

1. INTRODUCTION

Applications that require the fast and accurate localization of virtual sounds typically require the use of head-coupled auditory displays that update the spatial sound field to compensate for listener head movements. Real-time head-coupled displays provide dynamic localization cues that are absolutely critical to the resolution of front-back confusion errors in the horizontal plane [1, 2]. They also aid in the perception of elevation for low-frequency sounds [3] and increase the realism and immersiveness of virtual audio simulations [4]. In practice, however, the utility of dynamic head-coupled cues may be impaired by the delay that necessarily occurs between the movement of the listener's head and the resulting change in the signals presented to the listener's ears. Headtracking delays can come from a number of sources, including the latency of the actual tracking device, the communications delay between that device and the audio display, the time required to select the appropriate head-related transfer function (HRTF) and switch to that HRTF, the processing time required for the HRTF filtering, and any output buffering that occurs between the digital filtering of the sound and its eventual presentation to the listener over headphones [5].

Depending on their duration, headtracking delays can have two kinds of detrimental effects on the performance of a virtual display. The first is a decrease in localization performance, which can take the form of a decrease in localization accuracy, an increase in localization time, or a combination of the two. Previous experiments that have examined the effect of headtracker latency on localization accuracy have produced somewhat mixed results. Some researchers have reported that headtracker latencies as large as 150 ms [6] or even 500 ms [7, 8] have relatively little impact on the localization of virtual sounds. Other researchers have reported significant increases in localization error and response time for headtracker latencies as small as 93 ms [9]. Recent research in our own laboratory [10] has shown that headtracker latencies greater than 73 ms tend to decrease localization accuracy for brief stimulus presentations (from 125 ms to 2 s in duration) and increase localization response time for long stimulus presentations (greater than 500 ms in duration).

The second detrimental effect that headtracker latency can have on the performance of a virtual audio display is the loss of realism (and possibly the increase in fatigue or annoyance) that occurs when the delay is long enough to become noticeable to the user. At this point, there are surprisingly few data available to indicate what the smallest detectable headtracker delay is for human listeners. Indeed, the only study we are aware of that has indirectly asked listeners to detect the presence of headtracker latency is a study by Wenzel [8] that asked listeners to rate the latency of a virtual display on an arbitrary 25-point scale. Wenzel's results showed that listeners attributed significant amounts of latency to the system when the headtracker delay was 250 ms or greater, but not when the delay was 100 ms or smaller. They also showed that listeners only rated systems with 500 ms of latency (the largest latency values tested) to about half of the 25 point value labeled as "maximum latency." These results suggest that listeners are perceptually insensitive to headtracker latency, and that they might in some cases not even notice the presence of headtracker delay values that are large enough to impair their localization accuracy. However, Wenzel's experiment did not explicitly require listeners to identify the presence or absence of latency, so it is possible that her results reflect the perceived unimportance of headtracker latency rather than the inability of the listeners to detect its presence. In this paper, we present the results of a series of experiments that explicitly required listeners to detect the presence or absence of headtracker latency under a variety of listening conditions.

2. VIRTUAL AUDIO DISPLAY SYSTEM

The experiments were conducted with the General Dynamics 3D Virtual Audio Localization System (3DVALS) II audio display system, a custom-designed virtual audio display that combines two commercially available DSP processing boards (Texas Instruments TMS320C6211 Evaluation Boards) with a PC104 Pentium control computer and a custom-built backplane with twelve 24-bit A/D converters and two stereo 24-bit D/A converters. The basic processing path within the system is that the headtracker data arrives at one of the two DSP boards where it is used to look up the indices of the appropriate HRTF filters. Then these indices are passed to the second board where they are used to update the HRTFs used to process the input signal. This separation of the I/O and filtering functions of the display allows the HRTF filters to be updated very quickly with almost no buffering delays between the changing of the filter and the updating of the output signal.

For the purposes of this experiment, the 3DVALS system was set into 2D mode, where it uses headtracker information (collected from an Intersense IS-300 headtracker) to switch among 360 possible 126-point head-related transfer function (HRTF) filters, one for each 1° in azimuth in the horizontal plane. The filters used in this experiment were linear-phase FIR filters created at a 48 kHz sampling rate from HRTF measurements that were made every one degree in azimuth at a distance of 0.5 m from the center of the head of a Knowles Acoustic Manikin for Auditory Research (KEMAR) [11]. The processed stereo signals were then presented to the listener via stereo headphones (Beyerdynamic DT-990). For the purposes of this experiment, the software of the 3DVALS was modified to make it possible to artificially increase the latency of the headtracker by buffering the location information sent by the tracker in a first-in first-out (FIFO) queue. This allowed the total end-to-end headtracker latency of the system (measured from the onset of physical head movement to the corresponding change in the sound field of the system) to be set anywhere from a minimum of 11.7 ms (when there was no additional buffering the the FIFO queue) to a maximum of more than 250 ms, with a standard deviation of less than 1.5 ms [See Brungart et al. [12] for a complete description of the procedure used to measure the latency of the system].

3. EXPERIMENT 1: DETECTION OF HEADTRACKER LATENCY FOR A SINGLE SOUND SOURCE

3.1. Participants

Nine paid volunteer listeners, four male and five female, participated in the experiment. All had normal hearing ($<$ 15 dB HL from 500 Hz to 8 kHz), and their ages ranged from 21-24 years. All of the listeners had participated in previous experiments involving both real and virtual localization.

3.2. Procedure

The experiment was conducted with listeners located in a sound-treated listening room. A CRT was set up outside of a window in the sound room to allow the listeners to receive information during the experiment. Prior to the start of each trial of the experiment, the listener was asked to turn to face directly at this CRT and press the response switch. This response was used to "boresight" the headtracker by assigning that location to 0° azimuth. Then the first trial of the experiment was initiated by presenting a broadband noise signal directly in front of the listener (0° azimuth). During this stimulus presentation, the 3DVALS was either to produce the smallest possible headtracker delay value (11.7 ms) or to produce one of seven different artificially-inflated headtracker delay values (36 ms, 49 ms, 60 ms, 84 ms, 108 ms, 155 ms, or 203 ms). The listeners were then given up to twenty seconds to make exploratory head movements and use the left or right mouse button to indicate whether the stimulus was presented without headtracker latency (i.e. was presented at the minimum possible delay value of 11.7 ms) or whether it was presented with artificially increased headtracker latency. They were then given auditory feedback to indicate whether they had responded correctly or incorrectly and asked to face forward and press the mouse button to boresight the headtracker prior to the next trial.

The data were collected in blocks of 100 trials, with a different block of trials for each of the seven different delay values tested. Each listener participated in a total of four different blocks of trials at each delay value, in random order. Thus, each of the nine listeners participated in a total of 2800 trials in Experiment 1.

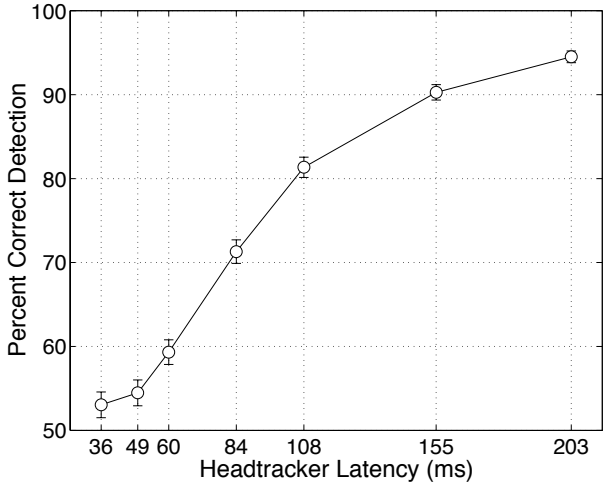

Figure 1: *Latency detection performance in Experiment 1, where the listeners had to distinguish between a reference signal with 11.7 ms of latency and a test signal with 36-203 ms of latency. The data have been averaged across all nine subjects, and the error bars represent the 95% confidence intervals around each data point.*

3.3. Results

Figure 1 shows mean overall performance in Experiment 1 as a function of the total headtracker delay in the test stimulus presentations in each block. The results indicate a systematic increase in detection performance from near-chance performance (50%) when the latency of the test signal was 36 ms to near-perfect identification when the latency was 203 ms. If the detection threshold for headtracker latency is set at 70% correct detections, these data imply that listeners can reliably detect headtracker latencies that are greater than 82 ms. This is smaller than the 100-250 ms detection threshold reported in the magnitude estimation study by Wenzel [8], but comparable to the approximately 73 ms delay value that started to degrade localization performance in our earlier experiment that examined localization accuracy with an experimental

setup similar to the one used in this study [12].

4. EXPERIMENT 2: DETECTION OF HEADTRACKER LATENCY WITH A LOW-LATENCY REFERENCE TONE

.

In Experiment 1, listeners were asked to detect headtracker latency in a stimulus that contained only a single target noise source. The results showed that listeners were unable to reliably detect headtracker latencies greater than 80 ms, which is a relatively large amount of latency in comparison to what can readily be achieved by most modern virtual audio display systems [5]. However, performance in the single-interval task used in Experiment 1 might have been somewhat impaired by the fact that listeners had to remember what the no-latency reference condition sounded like in order to make their responses within each trial of the experiment. This suggests the possibility that the listeners might have been more sensitive to headtracker latency in the target signal if they had been able to compare it to a simultaneously-presented reference sound that had no apparent headtracker latency. While these kinds of listening situations rarely occur in conventional virtual audio display applications, they could easily occur in "augmented reality" applications that use acoustically transparent headphones to superimpose virtual sound sources onto the natural auditory environment. In Experiment 2, listeners were asked to repeat the latency detection task of Experiment 1, but with a second low-latency reference sound source added to the stimulus.

4.1. Methods

The methods used in Experiment 2 were very similar to the ones used in Experiment 1. The major difference was the addition of a low-latency harmonic reference tone to the stimulus presented in each trial. This tone consisted of a five-frequency harmonic complex (500, 1000, 1500, 2000 and 1500 Hz) that was presented at the same location as the target noise source (0^o azimuth) with the minimum possible latency value available in the system (11.7 ms). The tone was scaled to have the same overall RMS power as the noise source. As in Experiment 1, each of the nine listeners participated in four 100-trial blocks for each of the seven headtracker latency values tested in the experiment.

4.2. Results

Figure 2 compares performance in the reference tone conditions of Experiment 2 to performance in the no-reference conditions of Experiment 1. These results show that the presence of the low-latency reference tone made the listeners substantially more sensitive to the presence of headtracker delay at all latency values less than 203 ms. The 70% latency detection threshold was lowered by 23 ms (from 84 ms to 61 ms) by the addition of the reference tone.

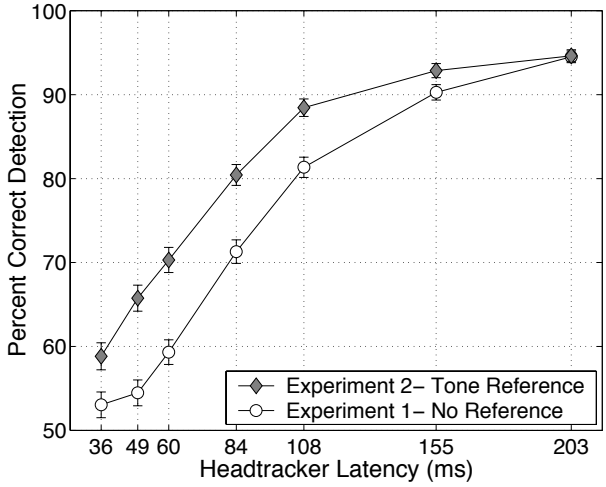

Figure 2: *Latency detection performance in Experiment 2, which replicated the latency detection task from Experiment 1 but added a reference tone that was co-located with the target noise stimulus but always had the lowest possible latency value (11.7 ms). For comparison, the data from Experiment 1 have also been replotted in the figure. The data have been averaged across all nine subjects, and the error bars represent the 95% confidence intervals around each data point.*

5. EXPERIMENT 3: SPATIAL SEPARATION OF REFERENCE TONE

The results from Experiment 2 show that listeners are substantially more sensitive to the presence of headtracker delay when they are able to compare the target stimulus to a spatially co-located reference tone with little or no latency. However, the results do not indicate how the detection advantages provided by the addition of a reference tone might vary with the location of that reference tone relative to the target signal. On one hand, one might argue that spatial separation between the target and reference signals might degrade performance because it would require the listener to divide attention across two locations in order to compare the relative latencies of the two sounds. On the other hand, one could argue that spatial separation might *improve* performance because it would provide a release from masking that would make the properties of the two signals easier to distinguish. Experiment 3 was conducted to examine the effects of spatial separation between the target and reference tones on the detection of headtracker latency in the target signal.

5.1. Methods

The experimental methods used in Experiment 3 were very similar to the ones used in Experiment 2, but with two major differences. The first major difference was that the headtracker delay value was always set to one of two values (60 ms or 83 ms). The second major difference was that the location of the reference tone was varied across 12 different locations ($0^\circ, \pm 5^\circ, \pm 15^\circ, \pm 30^\circ, \pm 60^\circ, \pm 90^\circ$, and 180° in azimuth), with a different fixed reference location in each block of trials. As in the earlier experiments, the target noise stimulus was always located directly in front of the listener (0°). The data collection was divided into blocks of 48 trials, with each block containing 44 trials at a single angular separation value plus four additional control trials with no reference tone. Each listener participated in a single block of 48 trials at each of the 12 possible reference tone locations at each of the two delay values, for a total of 2304 trials for each of the nine listeners in the experiment

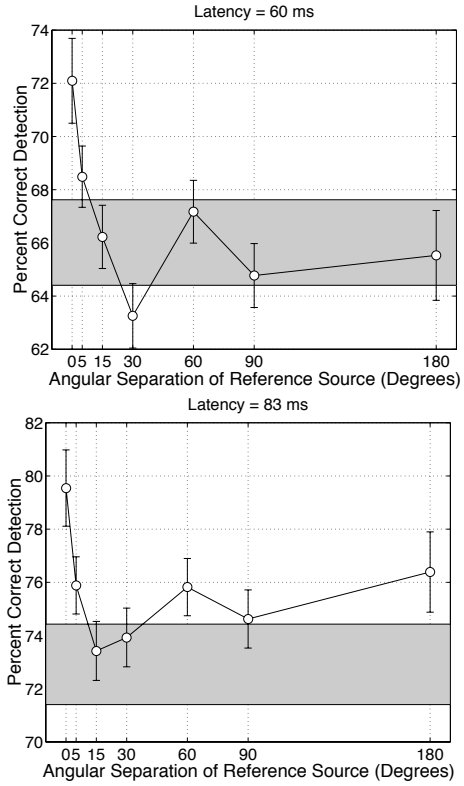

Figure 3: *Latency detection performance in Experiment 3, which replicated the latency detection task for two of the latency values from Experiment 2 but introduced a spatial separation between the low-latency reference tone and the target tone, which was always initially located directly in front of the listener. The error bars show +/- one standard error around each data point, and the shaded region in each panel shows mean performance +/- one standard error in the control conditions where no reference tone was present.*

5.2. Results

The top and bottom panels of Figure 3 show performance in the latency detection task as a function of the absolute angular separation between the target and reference tone for the two headtracker delay values tested in the experiment. For comparison, the shaded regions in each panel show mean performance in the control trials where no reference tone was presented. In the conditions with 0° of angular separation, the results clearly confirm the main result of Experiment 2 - that headtracker latency detection is enhanced by the addition of a low-latency reference tone that is spatially co-located with the target stimulus. However, these advantages appear to be greatly diminished by the introduction of even a small spatial separation (5°) between the target and reference tones. In practical terms, this suggests that the enhanced latency detection that might occur in an augmented reality situation that combines real and virtual sound sources is effectively limited to those cases where the real and virtual sound sources are located in the same direction. While this outcome would only rarely occur by accident, it could occur quite frequently in augmented reality systems that intentionally superimpose virtual sound sources onto the locations of real-world objects that might also be generating natural sounds.

6. INTER-SUBJECT DIFFERENCES

To this point, we have discussed the results of the experiments in terms of their mean values averaged across all the listeners. This allows us to estimate the smallest headtracker latency value the average listener would be able to reliably detect in a virtual audio display. However, a more useful statistic for the display designer is how large the delay could be in order to be undetectable to most, or all, users of the system. These answers can only be obtained by a careful analysis of the performance of the individual listeners in the experiment. The top panel of Figure 4 shows the 70% latency detection thresholds from Experiments 1 and 2 for each of the nine listeners in those experiments. These results have been extracted from linear interpolation of the performance curves of each listener. The results show that the best listeners were able to reliably detect headtracker latency values as small as 60 ms with no reference tone and as small as *36 ms* with a low-latency reference. Furthermore, it is likely that these numbers are slightly inflated due to the inability of the 3DVALS system to produce a true zero-latency signal in the reference trials. Thus, the best listener in Experiment 2 was really making a comparison between 36 ms and 11.7 ms, and not a comparison between 36 ms and 0 ms. From these data, one might infer that headtracker latencies will be essentially undetectable if they are kept below 50 ms in conventional virtual audio display systems and below 30 ms in augmented reality systems. Fortunately, even this lower bound is readily achievable with current virtual display technology [5].

The top panel of Figure 4 also shows that the advantages of having access to a low-latency reference tone are quite robust across different listeners - eight of the nine subjects in the study exhibited substantial reductions in their latency thresholds with the addition of the reference tone.

One striking aspect of the individual subject data in Figure 4 is the wide range of performance across the different listeners in the experiment. This raises the question of whether the performance difference between the best and worst listeners was due to differences in underlying psychoacoustic sensitivity or whether it was related to differences in the listening strategies of the different participants in the study. Some insights into this question are provided by the bottom three panels of Figure 4, which show three broad metrics of listener behaviour that were recorded during the course of the study. The second panel shows a rough estimate of the maximum head rotation rate achieved by each listener during the study. This estimate was determined by calculating the largest absolute change in head angle that occurred in any fixed-interval time frame (roughly 16 ms) during the course of each individual trial. The bars in the second panel of Figure 4 show the average of the maximum rotation rate across all the trials for that listener in Experiments 1 and 2. The third panel shows an estimate of the total absolute arc of head motion in each trial, which was calculated by summing the absolute changes in angle that occurred in each 16 ms frame of the trial. Again, the data in the figure show the average values across all the trials from Experiments 1 and 2. Finally, the last panel of Figure 4 shows the median total trial time for each listener. The most striking feature of these data is that the three best listeners (1-3) tended to make substantially more rapid head movements than the other listeners in the study. Thus, not surprisingly, there seems to be a direct correlation between head rotation speed and sensitivity to headtracker latency. It is worth

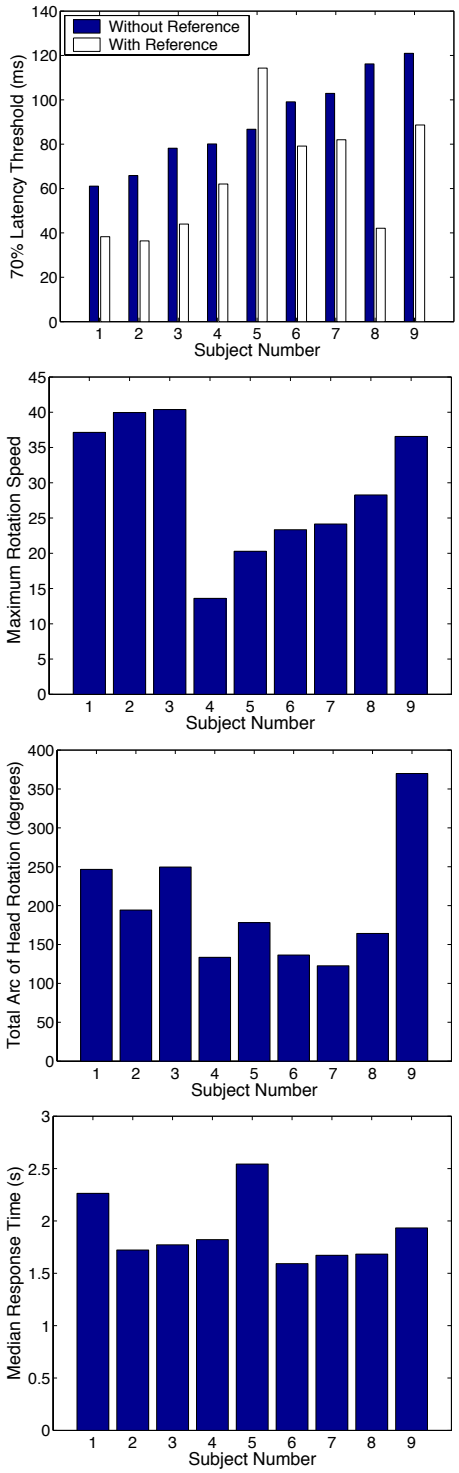

Figure 4: *Comparison of individual subject performance in Experiments 1 and 2. The top panel shows the 70% latency detection thresholds for each subject in each of the two experiments. The next panel shows a rough estimate of the maximum head rotation speeds achieved by each subject in the two experiments. The third panel shows the average total arc of head movement within each trial. And the fourth panel shows the median total trial time for each subject. See text for details.*

noting, however, that there are exceptions to this rule: Subject 4 performed reasonably well in the latency detection task but made relatively slow head movements, and Subject 9 performed very poorly but made very rapid head movements.

7. SUMMARY AND CONCLUSIONS

A critical engineering parameter in the design of interactive virtual audio displays is the maximum amount of latency that can be tolerated between the movement of the listener's head and the corresponding change in the spatial audio signal presented to the listener's ears. Previous experiments have shown that auditory localization performance is relatively unimpaired by headtracker latencies that do not exceed approximately 75 ms [8, 9, 12]. In this experiment, we have shown that the average listener is unable to reliably detect headtracker latencies smaller than about 80 ms in standard virtual audio display applications with a single virtual audio source. This suggests that headtracker latency starts to become noticeable at about the same point where it begins to impair performance in localization tasks. However, the results of Experiment 2 indicate that latency detection thresholds are reduced by approximately 25 ms when listeners are provided with a low-latency reference signal that is co-located with the virtual sound source, a result that suggests that augmented reality systems may require tighter headtracker latency tolerances than conventional audio display systems. Also, the individual subject results suggest that some listeners can detect latencies as small as 60 ms for isolated auditory stimuli and as small as 38 ms when a low-latency reference tone is also present in the stimulus. These results suggest that headtracker latencies will be undetectable to almost all listeners under almost all listening conditions when they do not exceed 30 ms, a level that is already readily achievable with current audio display technology. However, we should caution that further research is necessary to conclusively determine that latencies smaller than 30 ms have no effect on listener performance; they are not consciously detectable, but there is a slim possibility that they could cause increased irritation or fatigue over long periods of use.

8. REFERENCES

[1] H. Wallach, "The role of head movements and vestibular and visual cues in sound localization," *Journal of Experimental Psychology*, vol. 27, pp. 339–368, 1940.

[2] F.L. Wightman and D.J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *Journal of the Acoustical Society of America*, vol. 105, pp. 2841–2853, 1999.

[3] S. Perrett and W. Noble, "The effect of head rotations on vertical plane localization," *Journal of the Acoustical Society of America*, vol. 102, pp. 2325–2332, 1997.

[4] E.M. Wenzel, "Localization in virtual acoustic displays," *Presence*, vol. 1, pp. 80–107, 1991.

[5] J.D. Miller, M.R. Anderson, E.M. Wenzel, and B.U. McClain, "Latency measurement of a real-time virtual acoustic environment rendering system," in *Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, July 6-9*, 2003, pp. 111–114.

[6] A.W. Bronkhorst, "Localization of real and virtual sound sources," *Journal of the Acoustical Society of America*, vol. 98, pp. 2553–2553, 1995.

[7] E.M. Wenzel, "The impact of system latency on dynamic performance in virtual acoustic environments," in *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America, Seattle, WA, June, 1998*, 1998, pp. 2405–2406.

[8] E.M. Wenzel, "Effect of increasing system latency on localization of virtual sounds with short and long duration.," in *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001*, 2001, pp. 185–190.

[9] J. Sandvad, "Dynamic aspects of auditory virtual environments," *100th Convention of the Audio Engineering Society, Copenhagen, DK*, p. Preprint 4226, 1996.

[10] D.S. Brungart and B.D. Simpson, "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," *Journal of the Acoustical Society of America*, vol. 115, pp. 301–310, 2004.

[11] D.S. Brungart and W.M. Rabinowitz, "Auditory localization of nearby sources. i: Head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 106, pp. 1465–1479, 1999.

[12] D.S. Brungart, B.D. Simpson, R.L. McKinley, A.J. Kordik, R.D. Dallman, and D.A. Ovenshire, "The interaction between head-tracker latency, source duration, and response time in the localization of virtual sounds," in *Proceedings of the 2004 International Conference on Auditory Display, Syndey, Australia, July 6-9, 2004*, 2004.