

# 3D Audio Augmented Reality: Implementation and Experiments

V. Sundareswaran, Kenneth Wang, Steven Chen,  
Reinhold Behringer, Joshua McGee, Clement Tam  
Rockwell Scientific Company, Thousand Oaks, CA  
(sundar, kkwang, slchen, rbehringer, jhmcgee,  
cktam)@rWSC.com

Pavel Zahorik  
Waisman Center, University of  
Wisconsin, Madison, WI  
zahorik@waisman.wisc.edu

## Abstract

*Augmented Reality (AR) presentations may be visual or auditory. Auditory presentation has the potential to provide hands-free and visually non-obstructing cues. Recently, we have developed a 3D audio wearable system that can be used to provide alerts and informational cues to a mobile user in such a manner as to appear to emanate from specific locations in the user's environment. In order to study registration errors in 3D Audio AR presentations, we conducted a perceptual training experiment in which visual and auditory cues were presented to observers. The results of this experiment suggest that perceived registration errors may be reduced through head movement and through training presentations that include both visual and auditory cues.*

## 1. Introduction

In order to cue the user of an AR system about objects and events that are outside the narrow field-of-view (FOV) of a visual AR display, we developed a wearable 3D Audio Augmented Reality (3DAAR) system. We performed experiments to verify the ability of users to *localize* the audio signal. The results of one of these experiments suggest that there is a potential way to improve auditory localization through a perceptual training process that involves synchronous presentation of visual and auditory stimuli [1].

## 2. Wearable Implementation of 3D Audio

The wearable 3DAAR system is intended for mobile security applications, providing audible alerts to mobile personnel indicating the location of suspected/detected threats and waypoints for navigation to target locations. We built two of these systems, demonstrating situational awareness enhancement and collaboration through 3D Audio AR.

The 3DAAR system is built around the CharmIT wearable computer and features: Software-based 3D Audio presentation, wireless data communications (802.11b), Voice-over-IP (VoIP), small vocabulary speaker-independent Speech Recognition (utilizing IBM's ViaVoice®), GPS-based position measurement,

magnetometer-based head orientation tracking, and custom-built Push-To-Talk (PTT) buttons. The system is mounted on a vest and helmet configuration as shown in Figure 1.

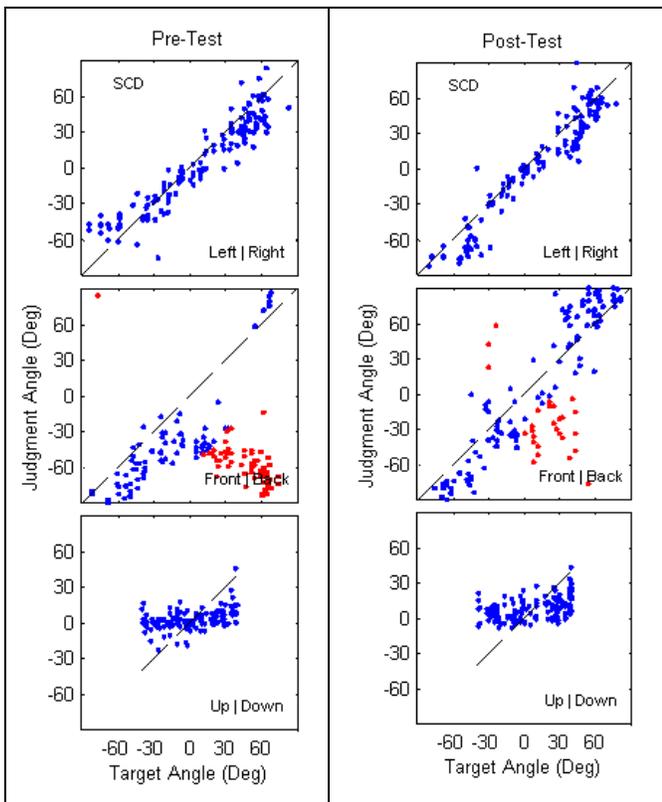


**Figure 1. 3DAAR with head tracker, GPS and 802.11b**

User head position and orientation are obtained from a Trimble GPS module and Leica DMC head tracker (respectively), while a 3D Audio Client forwards the data to the onboard 3D Audio Server application at 30 Hz. This server includes a custom-built HRTF engine, DirectSound3D, Aureal A3D HRTF engine, and a Sensaura HRTF engine. 3D audio icons are presented to the user over headphones using a *sweep* technique that plays the icons sorted sequentially by azimuth angle. Using 802.11b, each mobile user's GPS position is periodically reported to the Central Control Station, which in turn broadcasts the position of each tracked user & reported entities to all mobile users. In tests, mobile users were able to successfully navigate to the indicated positions of virtual entities. We observed informally that users tended to turn their head until the object seemed to be in the center ahead of them. Given this observation, we wondered whether this orientation response could be exploited via a training procedure to improve sound localization accuracy. By applying refined distance cueing and "aural magnification" that exaggerates the angular difference between the user's viewing direction and the acoustic source location, the navigation precision may be further improved.

### 3. Audio-Visual Experiments

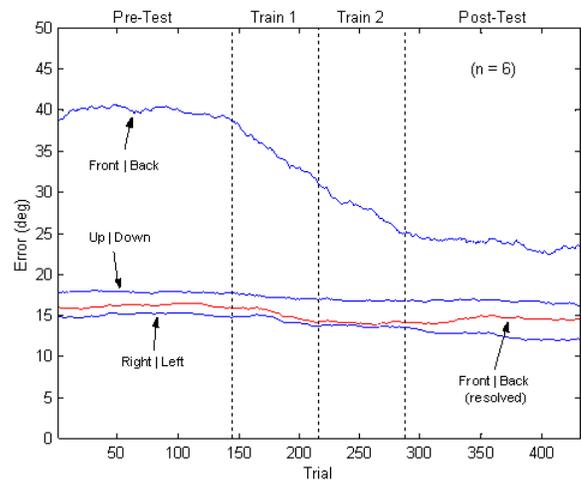
The goal of the audio-visual experiments was to determine the role of perceptual feedback training on localization accuracy under controlled conditions, using a large number of sound source positions presented with unambiguously paired visual targets shown on a Head-Mounted Display (HMD). The listener's head was tracked. Details on the experiments, the results, and their interpretation are described in [2]. A "pre-test" established a baseline for evaluating the user's ability to judge 3D orientation from sound. Subsequently, in a training phase, visual stimuli were displayed to show the correct source location. This visual feedback allowed the user to learn the association of the played sound to a 3D location. Finally, the post-test phase was identical to the pre-test phase and was conducted at least 4 days after completing the training phase and was used to assess lasting effects of the training. Figure 2 shows the results of one listener. Perfectly accurate responses in Figure 2 would lie on a positive diagonal shown by the dashed line.



**Figure 2. Results for one listener as a typical example (front | back confusions are shown in red).**

The error analysis shown in Figure 3 included a measure of front | back error in which confusions were

resolved by identifying confusions and multiplying them by  $-1$ . This allows a separate consideration of the angular location error. The most dramatic reduction of error occurred for the front | back dimension, falling from approximately 40 degrees to less than 25 degrees of error. This result provides strong evidence that the perceptual training alone was effective in reducing localization error in the front | back dimension. Other dimensions showed only minimal error reduction, on the order of a few degrees. A possible reason is that a perceptual remapping to degraded HRTFs may cause 2 spatial maps to form in the brain. Particularly useful is the lasting effects of the perceptual training. We observed that at least 4 days between training and the post-test did not affect accuracy.



**Figure 3. Localization errors, averaged over the subject population. Dramatic reduction of error occurred in the front | back case.**

Parts of this work were funded by AFRL Contract No. F33615-02-C6002 and ARL Contract No. DAAL01-96-2-0003

### 4. References

- [1] K. Wang, V. Sundareswaran, C. Tam, P. Bangayan, and P. Zahorik, "Efficient and effective use of low-cost 3D audio systems," in Proc. Intl. Conf. on Auditory Display 2002, Kyoto, Japan, July 2-5, 2002
- [2] P. Zahorik, C. Tam, K. Wang, P. Bangayan, & V. Sundareswaran, Localization accuracy in 3D sound displays: The role of visual-feedback training. In P. N. Rose (Ed.), Proceedings of the Advanced Displays Consortium: ARL's 5th Federated Laboratory Annual Symposium, College Park, MD, pp. 17-22, 2001
- [3] P. M. Hofman, J. G. Van Riswick, & A. Van Opstal. Relearning sound localization with new ears. Nature Neuroscience, 1(5), 417-21, 1998.