Evaluation of Digital Sound Spatialization Accuracy Over Commodity Audio Channels in a Personal Computer

Omar Grafals, Navarun Gupta, Gualberto Cremades
Armando Barreto, and Malek Adjouadi
Electrical and Computer Engineering Department
Florida International University

Abstract

This paper presents an analysis of the accuracy achievable in the spatialization of digital sound using generic head related transfer functions (HRTFs) played back over commodity audio channels readily available in most personal computers (PCs) now in market, using off-the-shelf headphones. We believe that this study is of interest because these are the conditions in which the vast majority of computer users actually experience 3D sound in consumer applications, such as games. Our analysis suggests that under these conditions localization is much less accurate to the sides of the head, and that there is an average localization error of approximately 15^o in the azimuth range [-75^o, 75^o].

I. Introduction

Research on Human-Computer Interaction (HCI) continues to gain prominence among computer research topics. One particular area of HCI research is that of 3-dimensional auditory displays. Researchers have already explored potential applications of these techniques in several areas. For example, 3-D audio implementations in fighter aircraft have been used to enhance situational awareness (threat location warning, wingman location indication, spatially separated multi-channel communications, etc.)[4][5], and the use of multimedia systems based on spatial audio has been proposed to provide access to GUIs for blind users [2]. In general, the systems that implement the 3D sound utilize a binaural headphone system with HRTF (Head Related Transfer Functions) processing technology. The HRTFs represent the modifications of phase and amplitude experienced by the different frequency components of a sound originating at a give spatial location with respect to the listener, before they reach each of the listener's eardrums. These amplitude and phase alterations are caused by the way the travelling sound interacts with the listener's torso, head, pinnae (outer ears), and ear canals. [3] The complexity of this interaction makes the HRTF at each ear heavily dependent on the location from where the sound originates Thus, there will be a pair of HRTFs associated with each sound source localization in the space around the listener. In fact, because of the individual characteristics of the listener's body HRTFs should, in principle, be measured for each person individually. In practice, however, this is not practical, and "generic" HRTFs, measured with a mannequin, are commonly used for the synthesis of 3D-sound.

The importance of 3-D audio displays lies within their potential to improve certain characteristics of human-computer interfaces. Whether it is to facilitate the use of a computer for a blind person, enhance game play for the younger generation, or increase situational awareness for individuals in critical environments. It is known that 3-D sound systems perform best when individual HRTFs, measured for the intended listener, are used for the synthesis and high-end audio components are used for the delivery of 3-D audio. The concern addressed in this paper is how well HRTF systems will perform under the following constraints:

(1) Use of generic as opposed to individual HRTFs.
(2) Use of commodity audio channels to deliver the specialized sound.

The purpose of the study is to evaluate the performance of 3D-sound emulation under these constraints and the impact on potential 3D sound applications.

II. Methods

Subjects:

Ten college-aged volunteers of normal hearing participated as subjects in our experiment (6 males, 4 females) their average age was 27.7 years. Each volunteer spent about ten minutes to complete the experiment. The subjects tested used a platform we built that allowed them to easily record their observations on a sheet of 8.5" 'by 11" paper. The platform is essentially a flat surface with guides for the paper and a pointing device at the bottom edge, which the subjects could use optionally to pinpoint the observed location on the sheet of paper. Figure 1 is the actual platform used by the subjects to record their data.

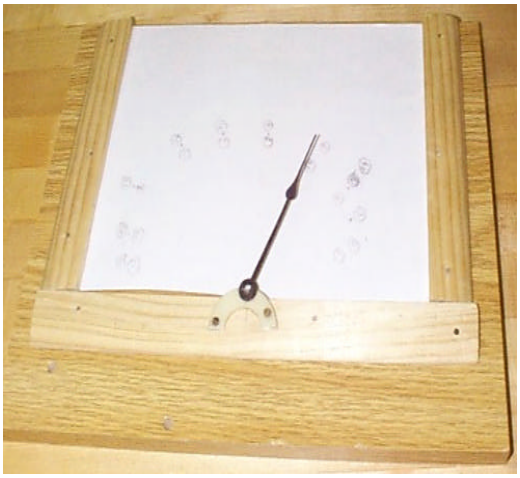

Figure 1. Data recording platform

Apparatus:

Using the compact (128-point) HRTFs measured by Bill Gardner and Keith Martin in an anechoic chamber at MIT, we imparted a direction to a sound of relatively short duration. Care was taken to choose a sound, which was of such a duration, that gave the listener enough time to decide about its direction. The sound used was also rich in many frequencies (broad band), particularly higher frequencies since these provide better localization cues than lower frequencies. The sound was played on a Pentium PC, and we used AKG K-270s as headphones for the test. Using MATLAB, we processed this sound with the HRTFs by using finite impulse response (FIR) filters for the convolution process. The resulting left and right signals were played to the test subjects, using a sampling rate of 22.05 kHz for playback. The HRTFs we used spatialize a sound by affecting the left and right channels in the same way they would be affected if the sound traveled from a given azimuth and elevation to the ears of the subject. We chose to restrict our investigation to sound locations at "ear level", i.e., with 0^o elevation. To further restrict the scope, we only used the HRTFs in the range from -90° (left of the head) to +90° (right of the head) azimuth. We did not implement directions that occurred behind the head for they are known to present front-to-back ambiguities [1] that are beyond the scope of this study.

Procedure:

To perform the test, each volunteer was first trained with a set of sounds from six different directions. This helped to acclimate them to the 3D sound environment and understand how to record their observations, reporting the perceived direction of arrival of each sound. Using the platform described above, the volunteers listened to sounds played to them from nine different emulated directions (0°, 3 -30°, -45°, -75° and -90°). As they would go from trial to trial, the volunteers recorded the directions from which they perceived the sound on a data chart. After the test was done, the random order in which the sounds were played would be recorded by the tester for the analysis and comparison between the actual direction emulated and the direction perceived and noted by the subject. In total, the subjects were put through twenty trials that were randomized while repeating each direction once. Figure 2 shows a volunteer in the process of taking the test.

Figure 2. Volunteer during testing process

Statistical Analysis:

A reliability analysis was conducted to validate this method of assessing the level of error in sound localization using generic (i.e, non-individual) HRTFs.

A repeated-measures Analysis of Variance (ANOVA) was carried out to investigate significant differences in level of error (true angle - perceived score) among emulated sound directions (angles measured in degrees).

III. Results

The Alpha-Cronbach coefficient for the reliability of the test was .95, which indicates that the test is statistically reliable.

Means and standard deviations for the levels of error found at different emulated sound locations are listed in Table 1. A graphic representation of the mean error detected at each one of the emulated sound locations is shown in Figure 3 (solid line). This figure also includes a representation of the standard deviation for the localization error at each emulated sound location.

Table 1. Descriptive Statistics for the error in perceived sound location for each emulated direction

Angles	Mean	Std.Dev.	N
0	7.65	5.8312	10
-30	22.15	12.8540	10
-45	12.85	7.4314	10
-75	15.55	11.1242	10
-90	31.4	11.7587	10
30	10.25	11.3315	10
45	12.35	9.8631	10
75	29.55	13.6554	10
90	46.45	14.2915	10

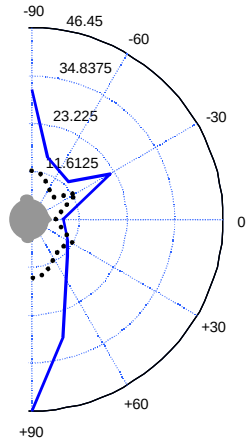

Figure 3. Mean (solid) and standard deviation (dotted) for the error in sound localization at each of th emulated sound directions

The ANOVA calculations revealed that there were significant differences in the level of error among sound locations ($p<.05$). Post-hoc Bonferroni tests revealed that the level of error was significantly greater ($p<.05$) at the $+90^o$ angle when compared to all other angles (except for the -90^o and $+75^o$ angle). Furthermore, Bonferroni tests revealed that the level of error was significantly greater ($p<.05$) at the -90^o and $+75^o$ angle when compared to the 0^o angle.

IV. Discussion

Both the values in Table 1 and the plots in Figure 3 indicate that the error in the perception of an emulated sound location is greater towards the sides of the listener's head, under the conditions of our test. The grand mean of the errors in localization can be calculated from the values in Table 1 to be 15.76^o in the azimuth range [-75^o , 75^o] (that is, excluding the emulated locations at and near +/-90^o, where the perception is particularly erroneous).

These findings may have important implications towards the application of 3-D sound systems, whenever generic HRTFs and commodity audio equipment will be used for its implementation. For example, in attempting to use 3-D sound in spatially separated multichannel inter-communication systems, such as the one needed in an aircraft or a command center, the level of error reported from our measurements may limit the number of spatial channels that are effectively distinguished by the listeners. If the emulated location of several listeners is to be accommodated within the interval [-75^o,75^o], each emulated spatial location should be separated from the next one by about 30^o, to account for the margin of error in perceiving each one of two adjacent emulated locations. Thus, the total available span of 150^o would only allow for the establishment of 5 effective spatial channels in the system (This is, of course, keeping our limitation of not using emulated locations to the back of the listener).

V. Conclusions

The purpose of this study was to determine the level of performance in localization achieved through generic HRTFs and off-the-shelf commercial audio components. To determine this level of performance an experimental method of assessment was proposed. Test-retest alpha reliability analysis revealed that the method of assessment utilized in this study was a valid measure to quantify the level of error in sound localization using generic HRTFs.

Significant differences were found in the level of error involved in the perception of emulated locations at +90^o azimuth (sound emulated as originated to the right of the listener). Similarly significant differences were found for sound locations emulated at +75^o and –90^o, with respect to the smaller level of error at 0^o. This suggests that a more efficient sound location emulation should be restricted to the azimuth interval [-75^o, 75^o], when generic HRTFs and commodity audio components are to be used.

These results may be important in the assignment of emulated locations for 3-D sound applications such as spatially separated multichannel inter-communications systems.

Acknowledgement

This work was supported by NSF grants EIA-9812636 and EIA-9906600, with Florida International University.

References

[1] Begault D., "3-D Sound for Virtual reality and Multimedia", Academic Press, 1994.

[2] Crispien K., and Petrie H., "Providing Access to GUIs for Blind People Using a Multimedia System Based on Spatial Audio Presentation". 5th Audio Engineering Society Convention (Preprint No. 3738), 1993.

[3] Kendall G.,"A 3-D Sound Primer: Directional Hearing and Stereo Reproduction", Computer Music Journal, 19:4, pp. 23-46, 1995.

[4] Mckinley, R., Erickson M., and D'angelo -Dimensional Auditory Displays: Development, Applications, and Performance". Aviation, Space, and Environmental Medicine, v. 65, no. 5, pp. A31-A38, May, 1994.

[5] Wenzel E., Wightman F., and Foster S., "Development of a Three-Dimensional Auditory Display System". SIGCHI Bulletin V. 20, no. 2, pp.52-57, October 1998.