



Audio Engineering Society Convention Paper 5714

Presented at the 113th Convention
2002 October 5–8 Los Angeles, California, USA

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

On the design of canonical sound localization environments

Eric J. Angel^{1,2}, V. Ralph Algazi¹, and Richard O. Duda¹

¹ Interface laboratory, CIPIC, University of California, Davis, Ca. 95616-8553, USA

² Correspondence should be addressed to ejangel@cipic.ucdavis.edu

ABSTRACT

This paper addresses the design of virtual auditory spaces that optimize the localization of sound sources under engineering constraints. Such a design incorporates some critical cues commonly provided by rooms and by head motion. Different designs are evaluated by psychoacoustics tests with several subjects. Localization accuracy is measured by the azimuth and elevation errors and the front/back confusion rate. We present a methodology and results for some simple canonical environments that optimize the localization of sounds.

1. INTRODUCTION

The approximation of the perceptual characteristics of physical environments is commonly the objective in the design of a virtual auditory space. By contrast, in this paper, our objective is the design of an environment that optimizes the localization of sounds.

The accurate localization of sound sources depends on the conjunction of a number of cues that are due to the nature of the sound, to the anthropometry and hearing characteristics of the listener, to the voluntary or involuntary motion of the source or the listener, and to the physical environment in which the listener is immersed. While some of these cues are of

value for accurate localization, others, such as room echoes and reverberation, are often detrimental. In our work, we start with a reference environment that consists of a personalized head-related transfer function (HRTF) of a listener and of noise bursts as sound sources with no echoes or reverberation. The approach to the design is to incrementally add to such an environment cues that will improve the localization accuracy. A particular application will specify or constrain the options or parameters of the environment. The minimal such environment that provides the best localization accuracy under a set of constraints will be denoted a canonical localization environment. The paper reports on the determination

of such environments and on some of their characteristics.

Four specific design assumptions differentiate our work from previous related studies [1, 2, 3]:

1. Use of personalized HRTFs. We limit ourselves to that case because: (a) a non-personalized HRTF will generally result in a poorer localization of sounds and (b) a generic HRTF may differ from a specific personalized HRTF in more ways than can readily be investigated and resolved.
2. Room acoustics. Since accurate localization rather than realistic reproduction of sound is the criterion, the determination and simulation of the acoustics of actual rooms is not considered.
3. Azimuth and elevation localization. Accurate localization in both azimuth and elevation is needed to exploit the auditory space fully. Since localization in elevation relies heavily on high-frequency pinna cues, broadband test stimuli are used.
4. Head motion. This produces an important localization cue that is investigated. However, since unrestricted head motion may compromise the accurate assessment and reporting of location, we analyze experimental results in order to clarify the contribution of head motion to localization.

1.1 Organization of Paper

The paper is organized as follows: Section 2 is an overview of the major sound localization cues that a virtual spatial sound system has to provide. Section 3 reviews relevant previous work. The rationale for the localization cues used in the proposed environments is presented in section 4. Section 5 provides details on the equipment and methods used in the localization tests. The results of these tests are organized and summarized in section 6. The canonical environments that are suggested by this work are presented in section 7, the concluding section, with a discussion of application-specific localization environments.

2. LOCALIZATION CUES

We are restricting our work to the use of customized HTRFs and therefore we only consider the contribution of the environment to the perception of the relative direction of the sound source. The environment will provide cues that may be filtered by the HRTF and will combine with the direct sound to enhance or weaken the perception of direction. The main cues for relative direction and distance are discussed briefly to set the stage for the parameters available for the design of a localization environment.

Interaural Cues: The interaural time difference (ITD), and the interaural level difference (ILD) are primary cues for localization in azimuth. Interaural cues generally do not provide enough information to determine the elevation of a sound source. Sources at many different locations produce essentially the same ITD and ILD if they are on the so-called "cone of confusion" shown in Figure 1¹.

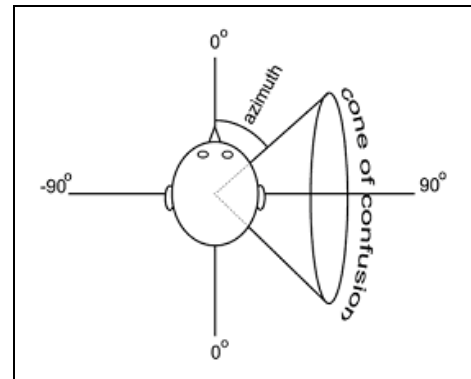


Figure 1: A cone of confusion.

Spectral Cues: The external ear, or pinna, can be viewed as a direction dependent filter that produces localization cues. The reflections and resonances of the pinna alter the spectrum of the sound received at the eardrum in a manner dependent on source direction. The flare of the pinna provides some directivity at high frequencies, resulting in a cue for front/back discrimination. The pinna, or more specifically the cavum concha, causes a deep spectral notch on sound in the 5 kHz - 10 kHz range. The center frequency of the notch depends on source elevation [4] and provides a primary cue for elevation. Because of its dimensions, the pinna mainly affects frequencies above 3 kHz. Below 3 kHz, head diffraction and torso reflections produce elevation dependent spectral changes. Algazi, Avendano, and Duda have shown that these low-frequency elevation cues are perceived and can be utilized [5]. However, it is generally accepted that the high frequency content of the sound is most important to elevation discrimination based on spectral cues.

Dynamic Cues: In addition to the static elevation cues provided by the pinnae, head, and torso, strong

¹ In this paper we report angles using an interaural-polar coordinate system. Readers who are more familiar with vertical-polar coordinates should be warned that interaural-polar azimuth is limited to the range from -90° to $+90^\circ$. Points that are in back of the subject are found at 180° elevation.

dynamic elevation cues result from head movements. A head rotation changes interaural differences, which produces a critical cue for determining whether a sound source is in front or behind. In addition to simple front/back discrimination, a head rotation also produces absolute elevation cues. A head rotation causes a maximum rate of change in ITD and ILD for sources in the horizontal plane, and that rate of change will decrease with increasing absolute elevation. Elevation cues derived from head rotations have been shown to be more important than spectral cues [6].

Externalization and Distance Cues: There are several known cues for distance that rely on prior knowledge of the acoustic environment, on the power of the source, or on a comparison of two sounds. In the research reported in this paper, distance judgments were not collected. Thus the interest in distance cues was only related to the general sense of distance or of externalization. One distance cue is based on the intensity of the pressure wave that, in free space, is inversely proportional to distance squared. Thus relative distances can be judged by comparing the intensity of two sounds, but for a single source absolute distance cannot be judged without prior knowledge of the source. Externalization and distance cues are also provided by room reflections that in a few tens of milliseconds merge into a non-directional reverberation. In most reverberant environments, the intensity of reverberation is roughly the same everywhere. Since the intensity of the sound received directly from the source varies with distance, the ratio of direct to reverberant energy is a cue for distance.

Precedence Effect: The precedence effect describes how localization cues are suppressed in a reverberant environment. Wallach [7] reported this phenomenon, and demonstrated that the locations of delayed copies of a sound, coming from different directions, are "fused" together with the sound that arrives first. As long as the copies are delayed by no more than approximately 40 msec, only the location of the first sound is perceived. The precedence effect is effective to varying degrees depending on the direction of the reflections, as is explored further in the review of previous work.

3. PREVIOUS WORK

Several key research papers served as motivation for the design of the canonical localization environments. These papers focus primarily on the perceptual effects of reverberation and head tracking.

Hartmann [8] and Rakerd and Hartmann [2] studied the effect of early reflections on azimuth localization accuracy using acoustically adjustable rooms. They found that reflection from the floor and ceiling gave more accurate azimuth judgments for sounds in the horizontal plane, while lateral early reflections blurred the location of the source.

Begault [9] considered the effect of artificial reverberation on localization accuracy for speech stimuli by simulating the acoustics of a specific room. The artificial reverberation consisted of early reflections, determined through geometric ray-tracing, and a decaying-noise late-reverb model. The first two early reflections were floor reflections, and the rest were lateral reflections from the vertical walls in a trapezoidal shaped room model. Using non-individualized HRTFs, virtual sound sources were placed at locations in the horizontal plane. The average front/back reversal rate for the group of subjects was about the same with or without reverberation with substantial individual differences. Elevation judgments were noticeably higher when reverberation was used. Additionally, azimuth judgments were less accurate and more spread out with reverb. In light of Hartmann's work, this might be attributed to the many lateral early reflections. Externalization was consistently improved by reverberation.

In a recent study, Begault et al [1] reexamined the effects of artificial reverb, and considered the effects of head tracking, reverberation, and individualized HRTFs on the localization of virtual sources in the horizontal plane. A rectangular shaped room was simulated with the early reflections obtained by ray tracing and the generic late-reverberation model for that room added as an option. Both reverberation conditions reduced azimuth error, but raised elevation judgments. Additionally, sounds were externalized twice as often when either reverberation was used. Localization was not affected by the late-reverb as compared to using only early reflections. Head tracking did not significantly affect absolute localization error, nor did it affect externalization rates, but overall reversal rates were reduced by about a factor of two.

Perret and Noble [3] studied the perceptual importance of changes in ITD and ILD caused by head rotations. Subjects localized low-pass, high-pass, and broadband noise from concealed speakers in the median and lateral vertical planes. Improvements in localization accuracy from head rotations were only seen when the stimuli contained

energy below 2 kHz, indicating that changes in ITD are the most important perceptually. Head rotations decreased the average reversal rate from about 30% to less than 1%. Head rotations also improved elevation judgments for low-pass and broadband noise, but not high-pass noise.

Wightman and Kistler [10] also considered the effects of head movements on localization of real and virtual sources. Subjects either remained motionless, made unrestricted head movements, or oriented toward the sound. Azimuth accuracy was roughly the same in all cases, with correlation coefficients ranging from 0.87-0.98. Front/back reversals were almost completely eliminated when head movements were allowed. Contrary to the findings in [3] for broadband stimuli, Wightman and Kistler found that head movements did not improve elevation accuracy at all. Analysis of subjects' head movements showed that subjects generally oriented toward the location of the sound when head movements were unrestricted.

Sandvad [11] investigated the effects of degrading key dynamic parameters of a spatialization system by considering the effect of update rate and the HRTF resolution. He found that system latency was the one parameter that, when degraded, significantly increased localization error.

4. LOCALIZATION ENVIRONMENTS

Based on these previous studies, we now discuss the rationale for adding specific features to a reference virtual spatial sound simulation environment that uses only "dry" sounds and a personalized HRTF. Note that although our goal is to design canonical environments that improve localization accuracy, we shall find that features that improve localization will simultaneously increase perceived realism.

4.1 Virtual Floor

With substantial evidence that lateral reflections may be detrimental to azimuth localization, and following the work of Hartmann and Rakerd who found that a "room" comprised solely of a physical floor degraded localization accuracy by only a modest amount, we concluded that for a virtual environment the externalization provided by a floor reflection would be of value to localization. Thus a room with only a virtual floor was chosen as the basis for a new canonical localization environment. Such a floor would reinforce the direct sound azimuth cue and may not be detrimental to elevation localization. Adding the virtual floor required only a modest increase in the computational requirements of the localization environment. Furthermore, compared to

the anechoic conditions of the reference environment, any room reflections invariably increased the perceived sense of "realism" of the virtual sound.

4.2 Late Reverberation

We also looked at how localization is affected by the inclusion of the spatially-diffuse late portion of a room response. We considered a simple late-reverb model, and varied the parameters to test for sensitivity. Late reverb increases the sense of externalization of virtual sound, which could lead to improved localization accuracy.

4.3 Head Tracking

Dynamic cues from head tracking were the next addition to the reference environment. The literature unanimously shows that head tracking with virtual sources reduces the frequency of front/back reversals and this is a critical component of sound localization. The effect of head tracking on angular localization accuracy for broadband sound sources was not definitely resolved in previous work and is addressed in this paper.

5. METHODS

We now describe the laboratory environment, the test signal used, the localization reporting methods and the statistical tools used in this work.

5.1 Localization Environment Equipment

The equipment used in the experimental environment consisted of a PC running MATLAB² from MathWorks, a PD1 Power SDAC Convolver and HB6 stereo headphone buffer from Tucker-Davis-Technologies, a Polhemus FASTRAK system, and Beyerdynamic DT-770 circumaural headphones. All localization tests were conducted in a sound treated room. Subjects used a self-guided MATLAB program to listen to sounds, and to give their localization judgments. MATLAB generated the stimulus, sent it to the PD1 along with HRTF coefficients, and determined new HRTF coefficients based on the head position and orientation information reported by the PD1. The PD1 convolved the stimulus with the HRTFs in real-time, and relayed head position and orientation information from the FASTRAK to MATLAB.

5.2 Stimulus

The stimulus was a pair of one-second long gaussian noise bursts, with a half-second of silence separating the two. Each noise burst was amplitude modulated

² MATLAB is a registered trademark of The MathWorks, Inc.

100% at a rate of 40 Hz, with the envelope beginning and ending at zero-crossings. The modulation increased the number of onsets, which provided additional potential localization cues. Broadband gaussian noise was used to maximize the number of localization cues that could be utilized. The PD1 convolved the stimulus with the appropriate HRTFs in real-time. Two 16-bit digital-to-analog converters on the PD1 were connected to the HB6 headphone buffer, which controlled the volume of the sound delivered to the headphones.

5.3 HRTF Set

An individualized HRTF set for each subject was used. The HRTFs were measured according to the methods described in [12]. The measured HRTFs had a spatial resolution of 5° in azimuth from -45° to 45° , and the $\pm 55^\circ$, 65° , and 80° azimuths were also measured; the spatial resolution was 5.625° in elevation from -45° to 230.625° . To ensure smoothness when head tracking was used, HRTFs were interpolated to have a spatial resolution of 1° in azimuth.

5.4 Headphone Compensation

To compensate the response of the headphones, as measured at the blocked entrances to the ear canals, repeated headphone response measurements were made and averaged in the frequency domain to remove variation caused by the inconsistent coupling between headphones and pinnae. This average response was flattened outside the range from 100 Hz to 10 kHz, and a $1/6^{\text{th}}$ octave smoothed response was used to design an inverse filter [13].

5.5 Head Tracking

Head movements were tracked by the Polhemus FASTRAK electromagnetic tracking system. The receiver was mounted on top of the headphones, the transmitter was fixed on the workstation desktop, and the distance between the two was always less than 1 m. Even accounting for the delays due to the MATLAB programs, the relative position of the source with respect to the head was updated at an average rate of more than 40 Hz, which is comparable to the reported update rates in other virtual source localization studies [1, 10, 11]. An initial calibration procedure was used to correct for inexact placement of the receiver.

5.6 Reverberation

The virtual floor was defined to be 1.3 m below the subject's head, and the source was at a range of 3 m. The virtual floor was modeled as a hard reflective surface with a frequency-independent reflection

coefficient of 0.5. The azimuth angle of the floor reflection was the same as that of the direct sound. The elevation angle θ_r was computed using the equation given in Figure 2, where fl is 1.3 m, R is 3 m, and θ_e is the elevation of the source relative to the horizontal plane. In tests where head movements were allowed, ϕ was the elevation of the subject's head relative to the horizontal plane. The HRTF corresponding to the direction of the floor reflection³ was scaled by the reflection coefficient and inversely with the range of the total distance traveled by the sound. Additionally, the reflected HRTF was delayed by the difference in time of travel.

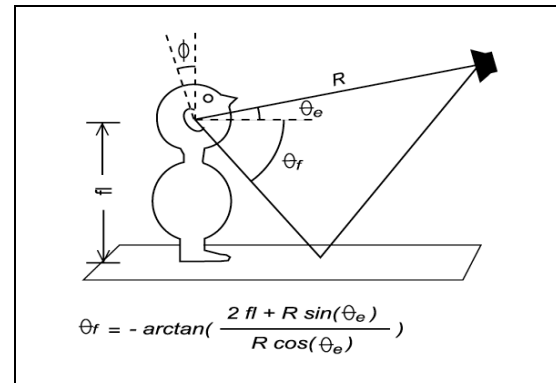


Figure 2 : Virtual floor diagram for a source in the median plane.

Spatially-diffuse late reverberation was used in a limited number of tests. The virtual-room model

³ The azimuth angle α_v for the floor reflection is defined in a vertical-polar coordinate system, while the azimuth angle α_i used for the HRTF measurements employs an interaural-polar coordinate system [5, 12]. To avoid having to perform a coordinate system translation, we neglected the difference between the two, after correcting for the fact that the range of α_v is 360° while the range of α_i is 180° . This did have the effect of deflecting the direction of the echo somewhat from the exact geometrical image location, or, equivalently, of introducing a distortion into the floor surface. In some cases, the floor reflection came from such a low elevation angle that there was no available HRTF data. When that happened, we used the HRTF data for the nearest available direction. Because the precedence effect causes the initial wave front to have dominant importance for localization, these kinds of small variations in the exact characteristics of the floor reflection are not likely to significantly affect the results.

then consisted of the following: 1) the direct sound, 2) the floor reflection, and 3) late reverb. The late reverb that was used was simply exponentially decaying gaussian noise. Separate left and right late reverb tails were generated, and the interaural cross-correlation coefficient was less than 0.05. The first two components of the virtual-room model were fixed by the room geometry, i.e. the floor position with respect to the ears, but the late reverb could be adjusted to achieve desired objective measures. The room impulse response $h[n]$ is shown below. $direct[n]$ and $floor[n]$ represent the HRTFs for the direct sound and floor reflection, respectively, $LR[n]$ is the late reverb, and $g[n]$ represents a sample from the normal distribution.

$$h[n] = direct[n] + floor[n] + LR[n]$$

$$LR[n] = a * \exp((n - t_d) / b) * g[n], \text{ for } n \geq t_d$$

$$0, \text{ for } n < t_d$$

The three parameters of the late reverb (a , b , t_d) control the total energy in the late reverb, the decay-rate of the late reverb, and the time between the direct sound and start of the late reverb, respectively. The ratio of direct-to-reverb energy is considered a measure of the loudness of the reverb. Because the floor reflection is considered part of the reverb response, there was an upper bound on the direct-to-reverb ratio that could be obtained by scaling the late reverb. This upper bound varied slightly from subject to subject, because it was dependent on the personalized HRTFs. Two “rooms” were considered with an approximate direct-to-reverb ratio⁴ of 5 dB, and a late reverb decay rate of 60 dB per 0.5 s, denoted LR1 and 60 dB per 1.0 s denoted LR2.

The late reverb parameters were customized for each subject prior to the experiment. The delay time t_d was the same for all subjects. The start of the late-reverb portion of a room response is usually considered to be about 80 msec after the arrival of the direct sound. However, because of the missing early reflections in our virtual room, a much smaller t_d had to be used. Qualitative listening tests were performed where a subject listened to different types of sounds placed in virtual rooms with different values of t_d . Not surprisingly, we found that sounds placed in virtual rooms with the longest t_d 's produced the sense of the greatest perceived distance. However, the sound began to be perceived as

unnatural when t_d exceeded 30 msec. For the present localization environment, a t_d of 20 msec was chosen. A 5-msec cosine-squared taper window was applied to the start of the late reverb. A representative room impulse response is shown in Figure 3.

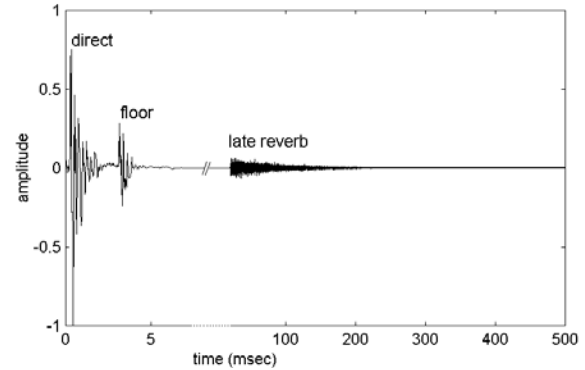


Figure 3: Room impulse response with the HRTFs for the direct sound and floor reflection, and late reverb tail.

5.7 Reporting Sound Location

The way in which subjects report perceived sound location is a fundamental problem in localization studies. We chose to have subjects report sound source location graphically. During the tests, subjects looked at one or more representation of themselves and their surrounds, and pointed where they perceived the location of the source in that representation. A MATLAB graphical user interface (GUI) program guided subjects through the experiments, allowing them to move at their own pace. In each trial, subjects heard the stimulus placed at a certain location, and then gave their localization judgment by placing cross-hairs on one or two images, one for azimuth and one for elevation.

5.8 Experiments

Experiments were conducted to compare localization accuracy in different environments. Each environment consisted of different combinations of two of the conditions under study: a room comprised of a virtual floor with or without reverberation and with or without the use of head tracking. In the first experiment, subjects localized sounds placed in the horizontal plane. The two main goals of this experiment were a) to reveal changes in front/back discrimination performance, and b) to reveal changes in localization accuracy in terms of azimuth error. A secondary goal was to analyze head-movement patterns to discover strategies used by subjects to localize sound sources. In the second experiment, subjects localized sounds on the 45° cone of confusion. The goal of this experiment was to determine the effect of head movement and of a virtual floor on elevation localization accuracy.

⁴ Reported direct-to-reverb ratios are with respect to the mean energy of the left and right ear HRTFs for a source at azimuth = 0°, elevation = 0°.

5.9 Test Subjects

Nine test subjects (2 female, 7 male, ages 21-52, median age 24) completed the horizontal-plane experiment over the course of one or two days. Five test subjects (1 female, 4 male, ages 21-52, median age 23) completed the cone of confusion experiment in one day. Four of these subjects participated in both experiments and three of them completed additional tests with the reverberation conditions. None of the subjects was involved in the research project, and all were ignorant of the test conditions. Each subject's hearing was checked with the Ear Q Reference Hearing Analyzer, which tested the threshold of hearing in 16 frequency bands from 63 Hz to 20 kHz. No significant hearing loss was discovered with any of the subjects. Before the start of the experiments, subjects received oral instructions on how to report their location judgments

5.10 Source Locations

In the experiment with virtual sound locations on the horizontal plane, a total of 14 locations in the horizontal plane were tested: 7 in front and 7 behind the subjects. The azimuth angles of the locations were -65° , -45° , -20° , 0° , 20° , 45° , and 65° . Each location was presented 10 times, giving a total of 140 trials for each test. Each test was divided into two blocks of 70 trials. Subjects needed about 10 minutes to complete each block.

In another experiment a total of 21 locations on the 45° cone of confusion were tested: 10 in front, 10 behind, and 1 over head. The elevations ranged from -22.5° to 202.5° in steps of 11.25° ; these were locations where measured HRTF data existed. Each location was presented 5 times, giving a total of 105 trials for each test. Subjects completed each test in separate sessions of about 10 minutes.

5.11 Measures of Error

In the localization experiments, errors were reported as angular differences (azimuth, elevation) between the target and reported locations, and as front/back reversals.

Before analyzing angular error, front/back reversals were always corrected for. If the reported and actual source locations were on opposite sides of the lateral vertical plane (LVP), then the reported location was reflected across the LVP (corrected elevation = 180° - reported elevation). When computing the reversal rate, trials in which the source was near the LVP required special care. In these cases, a small angular error might put the target and reported locations on

opposite sides of the LVP. This type of error should not be considered a reversal, and a "windowing" procedure was used to ensure that it was not. In the cone of confusion experiment, only trials where the target elevation was less than 70° or greater than 110° could possibly contribute to the reversal rate. However, elevation judgments were corrected for in all trials where reversals occurred.

5.12 Angular Errors

For angular errors, both conventional and spherical statistics have been used in previous localization studies. There are problems with the use of conventional statistics such as mean and variance for spherical data, as noted by Wightman and Kistler [14]. This is because angles of the same magnitude translate to different absolute distances, depending on where they are on the unit-sphere. However, since we considered source locations restricted to 2-D slices of a sphere, conventional statistics were acceptable for this study. To capture both the consistency and the uncertainty in the perception of angular errors both bias and standard deviation were computed. Bias in perception of the location of sound source location occurs commonly and is of special interest in a virtual environment.

5.13 Compiling Statistics

In the horizontal-plane experiment, bias and standard deviation of azimuth judgments were calculated for each of the 7 azimuths tested, so that each value was based on 20 judgments. Additionally, the values were averaged across all azimuths, to represent overall localization accuracy. The absolute value of bias was used in the average to prevent leftward and rightward biases from canceling each other. The mean unsigned bias is the metric that is reported in the results section.

In the cone of confusion experiment, bias and standard deviation of elevation judgments were calculated in basically the same way they were for azimuth judgments in the first experiment. The only difference is that in the second experiment all locations had a unique elevation, so results from different locations were never combined. Therefore, 5 judgments contributed to the bias and standard deviation values for each location.

5.14 Statistical Analysis

A one degree-of-freedom chi-square test for significant differences was performed on the reversal rates from two tests with different environments. The null hypothesis was that front/back discrimination was the same in both environments, so

the two reversal rates should have been the same. The null hypothesis was rejected when the calculated chi-square indicated a significant difference beyond either the 0.01 or 0.05 level.

6. RESULTS

This section presents the experimental localization accuracy results. The goal of the experiments was to see if localization would be improved by the modifications used in the different environments. An overall improvement was achieved when at least one of the measures of localization accuracy - reversal rate, azimuth error, or elevation error was improved, and none was degraded. A number of tests were conducted over several months and not all subjects were available for all tests. Therefore all the results could not be aggregated. Instead, they are presented for each individual subject and aggregated into subgroups consistent with the tests.

6.1 Effects of the Virtual Floor Without Head Tracking

6.1.1 Reversal Rates.

Tests on the horizontal plane: These tests were performed with nine subjects. Reversal rates from the reference and virtual floor environments are compared in Figure 4. The differences in reversal rates were significant for six subjects; one beyond the 0.05 level (subject 9), and five beyond the 0.01 level (subjects 1, 2, 5, 7, 8).

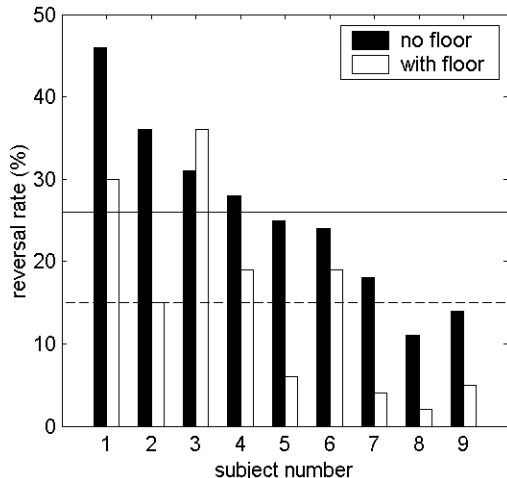


Figure 4: Reversal rates in the horizontal-plane experiment with and without the floor. Solid line is the mean without the floor, and the dashed line is the mean with the floor.

Tests on the 45° cone of confusion: These tests were performed with five subjects. The results are shown in Figure 5. The reversal rate with a virtual floor

alone was smaller for most subjects but the reduction was not significant except for subject 9⁵.

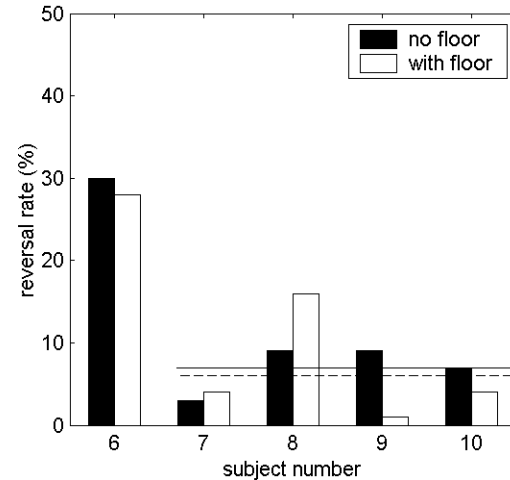


Figure 5 : Reversal rates in the cone of confusion test with and without the floor. Solid line is the mean without the floor, and the dashed line is the mean with the floor. Means are based on subjects 7-10.

Note that for the three subjects (7, 8 and 9) that completed all experiments, their average reversal rate on the horizontal plane dropped from 14% for the reference environment to 4% with the virtual floor. On the 45° cone of confusion, the reversal rate without the virtual floor was quite low at an average of 7% and trends were more difficult to infer for the small individual variations.

6.1.2 Azimuth Errors

The azimuth localization errors in the horizontal plane experiment are shown in Table 1. We note a significant decrease in the mean unsigned errors for a majority of the subjects as well as on the average.

⁵ In the cone of confusion experiment the overall means are evaluated without the results of subject 6, who was considered to be an outlier.

subject number	floor?	
	no	yes
1	18 (17)	12 (18)
2	17 (14)	18 (17)
3	28 (22)	19 (17)
4	11 (18)	13 (16)
5	26 (14)	8 (15)
6	15 (13)	17 (12)
7	18 (14)	9 (12)
8	12 (11)	5 (10)
9	18 (11)	9 (10)
mean	18 (15)	12 (14)

Table 1: Azimuth error in the horizontal-plane experiment with and without the floor. Mean unsigned bias, and standard deviations (in parentheses) are reported.

6.1.3 Elevation Errors

Elevation error results comparing the reference and virtual floor environments are given in Table 2. There is no significant gain or loss in elevation errors caused by the virtual floor.

subject number	floor?	
	no	yes
6	21 (31)	21 (38)
7	9 (13)	11 (14)
8	24 (15)	24 (13)
9	21 (12)	19 (10)
10	9 (17)	11 (15)
mean	16 (14)	16 (13)

Table 2 : Elevation error in the cone of confusion experiment with and without the floor. Mean unsigned bias, and standard deviations (in parentheses) are reported. Means are based on subjects 7-10.

6.2 Effects of Head Tracking

6.2.1 Reversal Rates

Tests on the horizontal plane: These tests were performed with nine subjects. Reversal rates from environments with and without head tracking and a virtual floor are compared in Figure 6. The differences in reversal rates without the virtual floor (black and gray bars) were significant for seven subjects; one beyond the 0.05 level (subject 2), and six beyond the 0.01 level (subjects 1, 4, 5, 7, 8, 9). Differences in reversal rates between the reference environment and the environment with both head tracking and the virtual floor (black and white bars) were significant at the 0.01 level for seven subjects (1, 2, 4, 5, 7, 8, 9).

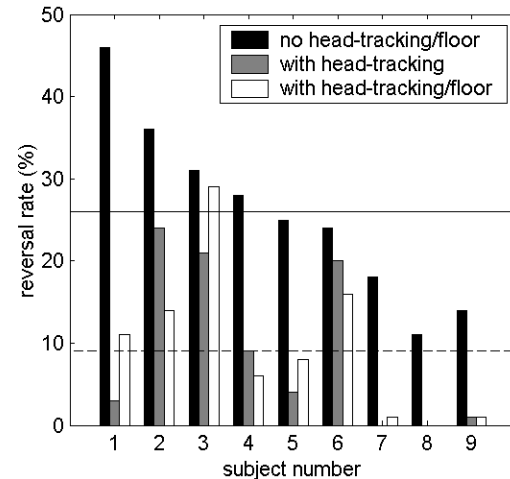


Figure 6 : Reversal rates in the horizontal-plane experiment with and without head tracking and the floor. Solid line is the mean without head tracking, and the dashed-line is the mean with head tracking, and also the mean with head tracking and the floor.

The decrease in reversal rate with head tracking was highly variable and quite large for most subjects, but may be dependent on the head motion patterns used by each subject. This issue will be discussed in section 6.4.

Tests on the 45° cone of confusion: These tests were performed with five subjects. The results are shown in Figure 7. The differences in reversal rates between the reference environment and either head-tracking environment were significant beyond the 0.05 level for only one subject (9), and the reversal rates were smaller in the head-tracking environments. The overall mean reversal rates in the head-tracking environments were almost half the reversal rate in the reference environment.

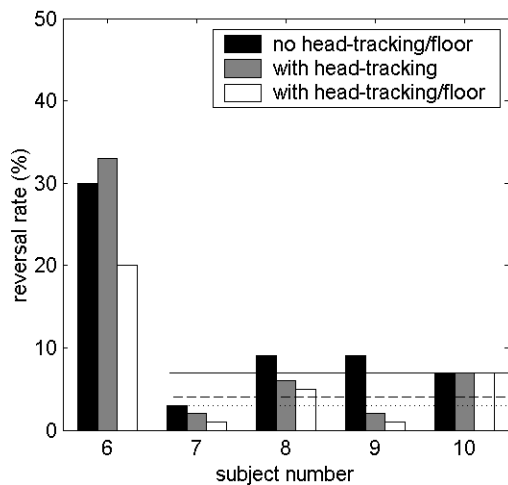


Figure 7 : Reversal rates in the cone of confusion experiment with and without head tracking and the floor. Solid line is the mean without head tracking, the dashed line is the mean with head tracking, and the light dashed line is the mean with head tracking and the floor. Means are based on subjects 7-10.

It is interesting to note that for all subjects the reversal rates on the cone of confusion were lowest when both head tracking and the virtual floor were used. Subject 6 did not make any head movements and therefore the improved performance is solely due to the virtual floor.

6.2.2 Azimuth Errors

These tests were conducted on the horizontal plane and the effect of head tracking on azimuth error is shown in Table 3.

subject number	head tracking/floor?		
	no/no	yes/no	yes/yes
1	18 (17)	4 (18)	7 (23)
2	17 (14)	9 (26)	11 (19)
3	28 (22)	20 (22)	23 (13)
4	11 (18)	6 (20)	10 (18)
5	26 (14)	11 (15)	8 (18)
6	15 (13)	15 (12)	19 (12)
7	18 (14)	5 (13)	3 (12)
8	12 (11)	6 (10)	6 (11)
9	18 (11)	5 (9)	6 (10)
mean	18 (15)	9 (16)	10 (15)

Table 3 : Azimuth error in the horizontal-plane experiment with and without head tracking and the floor. Mean unsigned bias, and standard deviations (in parentheses) are reported.

Head tracking has now a major effect, while adding the floor reflection does not result in a further improvement.

6.2.3 Elevation Errors

These test were conducted for sound locations on the 45° cone of confusion. Elevation error results from the head-tracking environments are compared to those from the reference environment in Table 4.

subject number	head tracking/floor?		
	no/no	yes/no	yes/yes
6	21 (31)	24 (29)	16 (31)
7	9 (13)	9 (11)	7 (9)
8	24 (15)	22 (14)	20 (11)
9	21 (12)	13 (8)	18 (9)
10	9 (17)	9 (15)	9 (14)
mean	16 (14)	13 (12)	13 (11)

Table 4 : Elevation error in the cone of confusion experiment with and without head tracking and the floor. Mean unsigned bias, and standard deviations (in parentheses) are reported. Means are based on subjects 7-10.

Adding head tracking to the reference environment reduced elevation errors overall. However, adding the virtual floor to the head-tracking environment did not reduce elevation errors further.

6.3 Effects of Late Reverberation

A limited experiment was performed to examine the effect of late reverb on localization. The same methods were used as in the horizontal-plane experiments with head tracking and a virtual floor. Three subjects completed tests where the environment consisted of a virtual floor and with or without a late reverb simulator. The goal was to see if adding late reverb to a virtual floor would affect localization. To test for sensitivity to the parameters of the late reverb, two different late reverb simulations were tested. The late reverb decay rate was 60 dB per 0.5 s for the first (LR1), and 60 dB per 1 s for the second (LR2). The direct-to-reverb ratio was 5 dB for both late reverbs.

6.3.1 Reversal Rates

As shown in Figure 8 the reversal rates were virtually identical in the different late-reverb environments. The late reverb had no effect on the subjects' front/back discrimination performance.

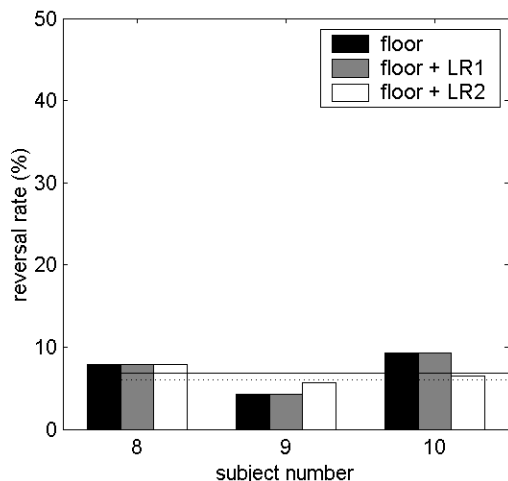


Figure 8 : Reversal rates in the horizontal-plane experiment with the floor and different late reverbs. Solid line is the mean with the floor, the dashed line is the mean with the floor and LR1, and the light dashed line is the mean with the floor and LR2.

6.3.2 Azimuth Errors

The effect of late reverb on azimuth error is reported in Table 5. This effect was not significant for either late reverb condition.

subject number	floor		floor + LR1		floor + LR2	
8	6	(10)	4	(12)	5	(10)
9	8	(8)	10	(12)	12	(8)
10	4	(13)	6	(14)	5	(12)
mean	6	(10)	7	(13)	7	(10)

Table 5: Azimuth error in the horizontal-plane experiment with the floor and different late reverbs. Mean unsigned bias, and standard deviations (in parentheses) are reported.

6.4 Analysis of Head Movements

This research brought to light the need for a more precise mechanism for relating localization accuracy to head motion. Because subjects were not given specific instructions on how to move their heads, the data reveals different head-movement patterns. With the exception of subject 6, all of the subjects made substantial head movements in most trials where head tracking was used. Subject 6 was reminded that head movements were allowed, and the subject acknowledged the fact, but the subject still rarely made any head movements. The most common pattern of head motion was to rotate the head so as to place the sound source at 0° azimuth in front or behind, thus “nulling” the azimuth. Such a pattern was only possible because of the use of a single virtual sound source and of long enough duration to permit “nulling”. As we discuss in the next section,

some applications may not allow the time for “nulling” the azimuth. In preliminary tests we observed that elevation errors increased when subjects made vertical head movements that were not tracked. However, when vertical head movements were tracked, elevation errors were slightly reduced. This finding agrees with the improvement in elevation localization from head movements reported by Perret and Noble [3].

7. DISCUSSION AND CONCLUSIONS

The results reported lead to the following conclusions with respect to canonical localization environments for a broadband noise stimulus:

1. When head tracking is not possible, an environment with a single floor reflection is a canonical environment in that it reduces reversal rates by more than 40%, decreases the bias of azimuth errors on the average by 30% and does not increase elevation errors. Adding reverberation to that environment does not significantly change the localization accuracy.
2. When head tracking is used, then an environment that includes a floor reflection as well as head tracking is canonical. For most subjects, the reversal error rate was the lowest in such an environment, and on the average was 65% lower than for dry sound and 40% lower than for the canonical environment without head tracking. The bias in azimuth errors was also reduced further by 15% from the canonical environment without head tracking and the elevation errors were slightly reduced.

We expect these results to hold for any broadband sound source and not only for noise. The results are likely to be different for speech or any other sound source with no high frequency content, in that the key contributions of the pinna to the determination of elevation will be absent or reduced. In that case we expect that head tracking may lead to a greater improvement in elevation localization.

Although the methods used to determine localization accuracy followed the procedures widely accepted in psychoacoustics evaluation, such methods do not necessarily provide a good basis for the use of the results in applications. A major thrust of our applied research is in the localization and discrimination of multiple sound sources that are either simultaneous or closed spaced in time. For such applications the results for a single source do not provide conclusive

answers. In particular, the effect of head tracking on the localization of multiple sources requires further work. This is because head motion cannot result in simultaneously “nulling” for all the source locations. Further, we expect that such tests will be much more robust with head motion.

8. ACKNOWLEDGMENTS

We are indebted to Dennis Thompson for his assistance in the experimental phases of this research. This work was supported by the National Science Foundation under grants IIS-00-97256 and ITR-00-86075. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the National Science Foundation.

9. REFERENCES

- [1] Begault, D.R., Wenzel, E.M., and Anderson, M.R., “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source,” *J. Audio Eng. Soc.*, Vol. 49, No. 10, pp 904-916, (October, 2001)
- [2] Rakerd, B., and Hartmann, W.H., “Localization of sound in rooms, II: The effects of a single reflecting surface,” *J. Acoust. Soc. Am.*, Vol. 78, pp 524-533 (1985)
- [3] Perret, S., and Noble, W. “The effect of head rotations on vertical plane sound localization,” *J. Acoust. Soc. Am.*, Vol. 102, pp 2325-2332 (1997)
- [4] Butler, R.A., and Belendiuk, K., “Spectral cues utilized in the localization of sound in the median sagittal plane,” *J. Acoust. Soc. Am.*, Vol. 61, No. 5, pp 1264-1269 (1977)
- [5] Algazi, V.R., Avendano, C., and Duda, R.O., “Elevation localization and head-related transfer function analysis at low frequencies,” *J. Acoust. Soc. Am.*, Vol. 109, No. 3, pp 1110-1122 (2001)
- [6] Wallach, H., “On Sound Localization,” *J. Acoust. Soc. Am.*, Vol. 10, pp 270-274 (1939)
- [7] Wallach, H., Newman, E.B., and Rosenzweig, M.R., “The precedence effect in sound localization,” *American Journal of Psychology*, Vol. 62, (3), pp 315-336, (1949)
- [8] Hartmann, W.H., “Localization of sound in rooms,” *J. Acoust. Soc. Am.*, Vol. 74, pp 1380-1391 (1983)
- [9] Begault, D.R., “Perceptual effects of synthetic reverberation on three-dimensional audio systems,” *J. Audio Eng. Soc.*, Vol. 40, No. 11, pp 895-904, (November, 1992)
- [10] Wightman, F.L., and Kistler, D.J., “Resolution of front-back ambiguity in spatial hearing by listener and source movement,” *J. Acoust. Soc. Am.*, Vol. 105, pp 2841-2853 (1999)
- [11] Sandvad J., “Dynamic aspects of auditory virtual environments,” presented at the 100th Convention of the Audio Engineering Society, Copenhagen, Denmark, May 1996 (preprint 4226)
- [12] Algazi, V.R., Duda, R.O., and Thompson, D.M., “The CIPIC HRTF database,” presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, October 2001
- [13] Angel, E. J., “Design And Validation Of Canonical Localization Environments” M.S. thesis. Report CIL-2002-3, CIPIC Interface Laboratory, University of California, Davis CA 95616 (June, 2002)
- [14] Wightman, F.L., and Kistler, D.J., “Headphone Simulation of Free-Field Listening. II: Psychophysical Validation,” *J. Acoust. Soc. Am.*, Vol. 85, pp 868-878 (1989)