Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems*

DURAND R. BEGAULT, *AES Member*

Aerospace Human Factors Research Division, NASA-Ames Research Center, Moffett Field, CA 94035-1000, USA

A psychoacoustic investigation was conducted in which five subjects gave localization judgments for headphone-delivered speech stimuli processed by nonindividualized head-related transfer functions, with and without synthetic "spatial" reverberation added to the stimuli. Spatial reverberation minimized intracranially heard stimuli, but increased the magnitude of azimuth and elevation localization errors. The results are applicable to three-dimensional sound systems and spatial sound field processors designed to increase the sensation of auditory "spaciousness".

0 INTRODUCTION

There has recently been considerable interest in implementing three-dimensional spatial auditory displays in a variety of fields in audio. The focus of research and development for three-dimensional audio has centered either on commercial music recording, playback, and "playback enhancement" techniques [1], or on utilizing the technology in advanced human–machine interfaces such as computer workstations [2], aeronautics [3], and virtual reality systems [4]. The technology is based on the implementation of digital filters that reproduce the filtering characteristics of the pinnae and head, the head-related transfer function (HRTF), which, along with interaural level and time differences, has been found to be an important cue for auditory localization (see [5] for an overview).

For these systems to be successful, the problem of perceptual mismatch between the intended location of virtual sound sources and their actually perceived positions must be addressed [6]. These perceptual errors can be roughly categorized into the following three areas:

1) Localization errors in azimuth and elevation estimation of virtual sound source targets

2) Reversal errors (hearing a virtual sound source at its mirror position in the rear hemisphere instead of the front or, less frequently, in the front hemisphere instead of the rear)

3) Distance errors (inability to hear sound outside of the head, and inability to predict distance of the virtual sound source).

Localization studies have cited a number of reasons for these perceptual mismatch problems. There are hard limits to localization acuity, with both simulated and actual localization cues, and these limits appear to vary from person to person [7], [8]. The influence of visual and cognitive cues can also affect auditory localization, as exemplified by cinema sound techniques. Overall spectral content and amplitude envelope rise time are also factors. For example, we can localize the chirping of a bird more easily than low-frequency tones with relatively slow amplitude envelope onset times. Perhaps more relevant is the fact that the representation of normal hearing localization cues in three-dimensional sound systems is typically impoverished. Our real-world ability to use head movement to focus on the location of a sound is considered an important factor in eliminating reversals, but very few systems track listener head position (e.g., [9]). Most significant is the fact that most systems use and will continue to use "nonindividualized" HRTFs, since measuring a unique set of HRTFs for each individual using a particular three-dimensional audio system is cumbersome. Finally, these nonindividualized HRTFs are usually measured in an anechoic chamber, where reverberation cues normally present in stimuli localized in the real world are absent.

As digital signal processing becomes cheaper, both the number of proposals and the complexity for implementing synthetic reverberation based on ray-tracing models of acoustic spaces have increased, as described in [10]–[12]. Some of these proposals have been worked

* Presented at the 91st Convention of the Audio Engineering Society, New York, 1991 October 4–8; revised 1992 February 10.

into three-dimensional sound systems, where HRTF processing is applied not only to the direct sound, but to the indirect sound field as well, resulting in a synthetic "spatial reverberation" [13]–[16]. The purpose for doing so is to allow modification of the original recording so that it seems to have been recorded within an arbitrary environmental context, such as a concert hall. However, few comparative studies have been done on the effects of simulated spatial reverberation on localization performance. The vast literature on the precedence effect (such as [17]–[19]) applies to the perceptual effects of actual early reflections only to the extent that a small number of time-delayed, amplitude-scaled copies of the direct sound resembles an actual reverberant field. In one headphone study of reverberation, very small thresholds were found for the detection of changes in impulse response patterns, but not for late reverberant decay [20]. Regarding the effect of real reverberation on the location of impulsive stimuli, studies have shown that lateral early reflections can result in a breakdown of the precedence effect, resulting in degraded azimuthal location [21], [22].

In the present study, localization performance was compared for five subjects listening with headphones under two conditions: with stimuli processed with anechoic, nonindividualized HRTFs, and with a synthetic "spatial reverberation" field added to the same stimuli. The results are presented here in terms of the perceptual errors described previously. Speech stimuli were used because of their relevance to many potential applications of three-dimensional sound, such as teleconferencing. The two stimulus condition types are referred to here as "dry" (stimuli processed with anechoic HRTFs) and "reverberant" (the same stimuli but with added synthetic spatial reverberation).

1 METHOD

1.1 Stimulus Generation

The dry stimuli were generated by digitally filtering a set of 45 one- or two-syllable words, each representing a particular phoneme from an international phonetic alphabet list, with a binaurally measured HRTF pair. Digital filtering was used to generate a version of each speech segment at the following 10 azimuth target positions: 0, 180, and left and right 30, 60, 120, and 150°. The target elevation was 0° (eye level) for all targets. The anechoic HRTFs were derived from a representative subject (SDO) who was measured extensively in [23] and whose measurements were used in [24] and [25]. These particular HRTF measurements were chosen because the overall localization performance of SDO in both free-field and headphone localization studies was better than the average performance of other subjects measured in the original study [8]. The transfer function of the headphones and measurement apparatus was divided out of the HRTF measurements as described in [23].

The reverberant stimuli were generated by additional processing of the dry stimuli, and then adjusting the resulting rms level to equal that of the dry stimuli. The reverberation simulation consisted of two parts—an early reflection pattern and a late reverberation pattern (Fig. 1), based on a model of a listener 1 m distant from an omnidirectional sound source. The early reflection pattern itself consisted of two parts—an initial reflection pair to represent two floor reflections (HRTF filtered at −36° elevation, ±30° of the target azimuth, delayed 5 and 5.5 ms, 6 dB below the level of the direct sound); and a later group of 64 early reflections calculated from a simple two-dimensional ray-tracing program for each target position. This program was based on a technique described in [26] and implemented as outlined in [3], [14].

The later group of 64 early reflections began 17–21 ms after the direct sound, depending on the target direction, at about 19 dB below the level of the direct sound. The enclosure specifications used here were asymmetrical, in the hope that differences in arrival-time patterns between front and back source target positions could alleviate localization reversal errors. Hence a unique pattern of early reflections was calculated for each target position within this modeled enclosure. Fig. 2 illustrates the modeled enclosure and the early reflection pattern for the target at left 30° azimuth.

The late reverberation portion of the stimuli was modeled as two separate distributions of exponentially decaying noise, one for each output channel. Slight differences were implemented to decrease correlation between the channels for greater realism in the simulation. The noise was generated digitally with a pseudorandom number generator as a mixture of white and $1/f$ noise (1:1 ratio for the left channel, 3:2 ratio in the right channel), decaying exponentially at slightly different rates to 60 dB below its initial level at 750 ms. The late reverberation was delayed by 50 ms so that the final early reflection overlapped with its beginning. This resulted in a final R_t time of 800 ms for all of the reverberant stimuli.

The signal processing involved sequential steps of finite-duration impulse response (FIR) filtering of the direct sound according to the model described, using floating-point impulse responses at a 50-kHz sampling rate. The left and right channels of the dry stimulus

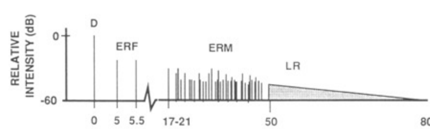

Fig. 1. Reverberation simulation used for current study. D—direct sound, HRTF-filtered at target direction, identical to dry stimuli; ERF—two floor reflections, 6 dB down, HRTF filtered at ±30° azimuth of target azimuth, at an elevation of −36°; ERM—is pattern of 64 early reflections calculated from ray-tracing model, each reflection is HRTF filtered and scaled according to model; LR—late reverberation modeled with exponentially decaying noise.

were first filtered by the two floor reflections. Then, for each of the 64 modeled early refections, the angle of incidence to the reflecting surface, the absorptivity of the reflecting surface, and the path length given by the ray-tracing program determined an initial attenuation and time delay of a copy of the direct sound. Subsequently the angle of incidence to the listener was calculated, and the nearest measured HRTF (every 30° azimuth) was used to create a two-channel binaurally processed early reflection. The impulse responses for all 64 reflections were then summed for filtering. Finally the left and right outputs were convolved with the late reverberation noise signals and then adjusted by an overall gain to match the level of the dry stimuli. The playback level of 70 dBc SPL (±2 dB) corresponded to the rms level of a person speaking at a distance of 1 m from the listener (5 dB above the average and 5 dB below the maximum long-term rms values for normal speech [27]).

A detailed explanation of the signal processing scheme for the 64 reflections is shown in Fig. 3. Fig. 3(a) is a simplified illustration of a single reflection and a direct source, comprised of a sound source S, a wall W, and a listener L. Fig. 3(b) is a diagram of the digital signal processing algorithm; the numbers show each stage of calculation based on the ray-tracing program, corresponding to the numbers in the following description.

1) The digital sound source input $x(n)$ is split into three paths. The center path involves the processing for the 64 early reflections, based on the ray-tracing method. This corresponds to the reflected sound S–W–L in Fig. 3(a). The upper and lower paths correspond to the direct sound shown as the path S–L in Fig. 3(a) and the two floor reflections (not shown).

2) The direct sound is spatialized by a single HRTF

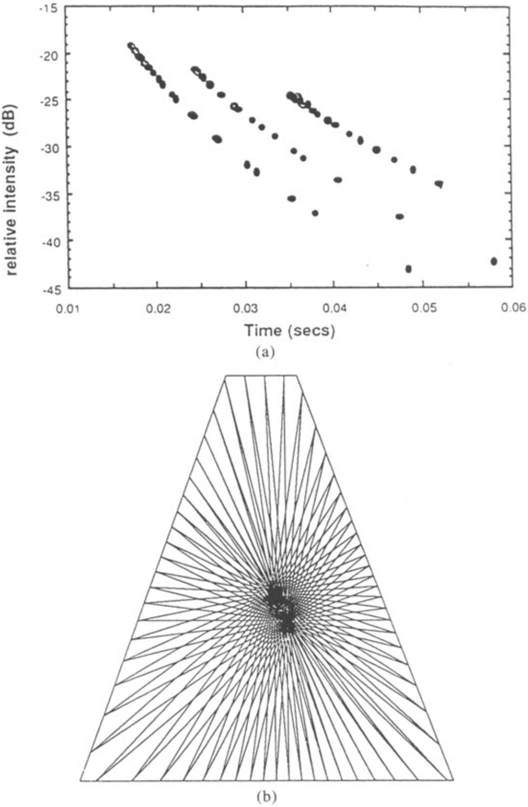

Fig. 2. (a) Reflection intensities over time for modeling of 64 early reflections from source to listener. (b) Modeled environmental enclosure. Dimensions: length of top wall 2.8 m; bottom wall 10.8 m; distance of listener from top wall 5 m; listener centered with sound source moving at 30° increments along an azimuth at a distance of 1 m.

filter pair, corresponding to the angle of incidence of the direct sound to the listener. Furthermore, two additional HRTF filter pairs delayed by 5 and 5.5 ms are used to simulate the floor reflections.

3) An attenuation based on the inverse square law and a delay based on sound velocity (334 m/s) are calculated for the distance of the sound path S–W.

4) A magnitude transfer function corresponding to the approximate frequency-dependent characteristics of a specified surface (here, plaster) is convolved with the signal at this point. This filtering is only approximate since the transfer function in reality changes as a function of the incident angle of the sound.

5) The angle of incidence of the reflected sound to the surface is analyzed and then attenuated according to a model of specular reflection, such that for frequencies above 500 Hz the amplitude is decreased as the angle of incidence of the reflection to the surface W becomes smaller.

6) An attenuation and a delay are applied for the path W–L, in the same manner as in step 3.

7) The angle of incidence of the reflected sound from the surface to the listener is calculated and then used to determine which HRTF filter pair is used to spatialize the reflection.

8) Each channel of the resulting spatialized reflection is summed with the direct sound and floor reflections, and is ready for late reverberation processing.

1.2 Experiment

Five adults served as paid volunteers in the study (ages 23, 24, 31, 33, and 40). All reported no known hearing loss or history of hearing problems, and none had previously participated in a headphone localization experiment. Subjects were screened with an initial training block of 20 trails and a headphone-centering task using a 440-Hz tone.

For each trial, the particular combination of speech segment, condition (dry or reverberant), and target location was chosen randomly. All stimuli were presented to blindfolded subjects via headphones (Sennheiser HD-430s) in total darkness within a soundproof chamber. Playback was from a Masscomp computer with 16-bit digital-to-analog converters. During a trial, subjects heard a given speech segment repeated five times and then called out estimates of the azimuth using a modified spherical coordinate system. Azimuth was defined as 0 to 180° left or right (where 0° is directly in front) and elevation was defined as 0 to 90° up or down (where 0° is at ear level). For distance, subjects were instructed to use "0 inches" if the sound was directly at the center of their head, between 0 and 4 in for positions inside the head, exactly 4 in for a verged-cranial impression (at the edge of the head), and greater than 4 in for externalized sounds. For example, a sound that seemed outside the head, slightly elevated, and to the right of the median plane might be reported as "right 30°, up 15°, and 30 in." Over the course of two to three days, each subject listened to 10 blocks of 40 stimuli containing a randomized ordering of the target azimuth positions. In sum, each subject gave a total of 40 estimates for each of the 10 targets: 20 judgments for the dry stimuli, and 20 judgments for the reverberant stimuli.

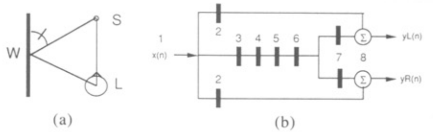

Fig. 3. Signal processing scheme for creating synthesized, early reflection portion of stimulus. (a) Simplified illustration of sound source S, wall W, and listener L, with direct sound S–L and one early reflection path S–W–L traced. (b) Steps involved in signal processing a one-channel digital input $x(n)$ into a two-channel binaural digital output $yL(n)$ and $yR(n)$. Refer to text for explanation of numbers in illustration.

2 RESULTS

2.1 Reversals

Front–back "reversals" are responses which indicate that a source in the front hemisphere, usually near the median plane, was perceived to be in the rear hemisphere. Occasionally the reverse situation is also heard. In the literature, reversals have generally been resolved when computing descriptive statistics (that is, the responses are coded as if they had indicated the correct hemisphere), and then the number of reversals is reported as a separate statistic. The argument for treating reversals as a separate issue is based on the premise that localization blur would be unfairly inflated if reversals are left "uncorrected" in reporting the results of an experiment [8], [28].

The procedure for resolving reversals used here is to test whether the angle between the target and the judged location is made smaller by reflecting the judgment about the vertical plane passing through the subject's ears. If the test proves true, the judgment is coded in reflected form and the percentage of reversals is increased.

In Table 1, the percentage of reversals under dry and reverberant conditions for each subject is shown, along with chi-square scores and probability levels for significant differences. The overall percentage of reversed judgments for both dry and reverberant conditions was the same, about 33%, a ratio close to the 27.5% rate found in a previous study that used the same dry speech stimuli [25]. However, individual differences were quite apparent. For two of the subjects, r_5 and r_7, reverberation made no significant difference in the reversal rate. Subject r_5 made significantly more front–back reversals, while subject r_7 made almost the same number of reversals but predominantly back–front (both at a ratio of about 30:1, $p < 0.001$). On the other hand, the reversal rates for subjects r_8, r_9, and r_{11} were affected by presence of reverberation, although in different ways. With reverberation present, subjects r_8 and r_{11} made significantly fewer front–back reversals, but they also made significantly more back–front reversals. It is possible that the presence of reverberation caused an overall bias toward frontal judgments for these subjects, although their overall ratio of front–back reversals is about 2.5:1. Finally, subject r_9 made significantly more front–back reversals with reverberation present than without, but had a very significant bias toward front–back reversals, independent of stimulus condition (a 79:1 ratio of front–back to back–front, $p < 0.001$).

2.2 Localization Error

Two measures of azimuth and elevation judgments, analogous to means and variances, were computed using spherical statistics [29], 1) the judgment centroid, the "mean direction" of a set of judgment vectors for a given target, and 2) inverse kappa κ^{-1}, a measure of dispersion around the centroid on the surface of a sphere. (See [8] and [24] for additional information on the application of spherical statistics to localization studies.) Table 2 shows the reversal-corrected judgment centroids and mean κ^{-1} values for each target position. The mean value of κ^{-1} from Table 2 is 0.116 for the dry condition and 0.195 for the reverberant condition, indicating a greater dispersion of judgments under the latter. This could possibly be due to the fact that the extent of the auditory image, or its "auditory spaciousness," was increased by the presence of decorrelated reflected energy. The κ^{-1} values are larger than the corresponding values computed in [8] for subject SDO listening with their own HRTFs; κ^{-1} was 0.06 and the mean error angle was 20.5° for a range of different azimuths at middle elevations (0 and up 18°).

Fig. 4 shows a plot, based on Table 2 of azimuth centroids for each condition averaged across subjects. The distance of the points from the diagonal line in the center of the plot represents the deviation of judgments from perfect agreement with the target location. The results for reverberant stimuli show a greater degree of "pulling" toward the left for 0 and left 30° targets than those for dry stimuli, and a greater degree of pulling toward the right at 180° and right 150°. Plots for individual subjects were near the diagonal, except for subject r_8, who showed a tendency to collapse judgments toward left and right 90° (Fig. 5). This pulling toward the vertical–lateral plane was also observed in a free-field study [28] and in a headphone localization study

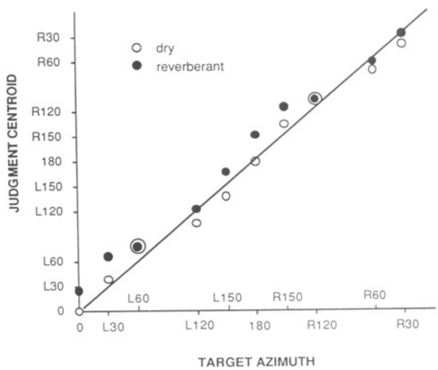

Fig. 4. Scatter plot of centroids based on mean values for all subjects.

Table 1. Percentages of front–back and back–front reversals by subject.

Subject	Percent Front–Back Reversals			Percent Back–Front Reversals		
	Dry	Reverberant	Chisquare*	Dry	Reverberant	Chisquare*
r_5	58	70	—	4	0	—
r_7	0	2	—	60	59	—
r_9	66	92	20 ($p < 0.001$)	0	2	—
r_8	64	32	14 ($p < 0.001$)	5	24	14 ($p < 0.001$)
r_{11}	69	27	35 ($p < 0.001$)	8	30	15 ($p < 0.001$)

* Chi-square values and probability levels given where significant differences between conditions exist.

Table 2. Mean azimuth and elevation centroids for each target for reversal-corrected judgments, and inverse kappa κ^{-1} means, based on five subjects.

Target azimuth	Azimuth centroids		Elevation centroids		κ^{-1} Means	
	Dry	Reverberant	Dry	Reverberant	Dry	Reverberant
L0	R0	L25	U0	U51	0.128	0.179
L30	L38	L66	U19	U33	0.158	0.123
L60	L78	L78	U20	U25	0.077	0.102
L120	L106	L123	U8	U17	0.065	0.164
L150	L138	L167	D11	U20	0.117	0.268
L180	L179	R149	D7	U32	0.076	0.289
R150	R136	R115	U12	U20	0.153	0.218
R120	R106	R105	U14	U13	0.112	0.129
R60	R71	R60	U24	U23	0.104	0.215
R30	R45	R27	U33	U44	0.176	0.268

for three out of nine subjects [25].

Fig. 6 shows a plot, based on Table 2, of elevation centroids averaged across subjects for each target azimuth. These averaged centroids are almost all elevated from the target elevation of 0° (eye level), except for the dry stimuli at 0, 180, and left 150°. The mean value across azimuths for the dry stimuli is up 11°, lower than the mean value of up 17° found in a previous study using the same stimuli [25]. The elevation centroids for the reverberant stimuli are higher; the mean across all target azimuths is up 28°.

In order to compare localization judgments between dry and reverberant conditions, values for absolute azimuth and elevation error were calculated for each target. Absolute azimuth (elevation) error is defined here as the absolute value of the difference in degrees between each subject's mean azimuth (elevation) centroid and the target position. Table 3 shows means and standard deviations across subjects for absolute azimuth errors, and Table 4 for absolute elevation errors. Analyses of variance were done on the conditions used (target azimuth versus reverberant or dry stimuli). The mean values of absolute azimuth error for the two conditions [dry = 11.9 (standard deviation 9.4), reverberant = 22.9 (standard deviation 16.9)] were significantly different across subjects ($F_{(1,4)} = 19.8$, $p = 0.011$), as was the interaction between target azimuth and condition ($F_{(9,36)} = 7.89$, $p < 0.001$). The mean values for absolute elevation error shown in Table 4 were also found to be significantly different in the interaction between target azimuth and condition ($F_{(9,36)} = 2.74$, $p = 0.015$).

Fig. 7 summarizes the difference in absolute azimuth and elevation error between the two conditions. It can be seen that for most targets the addition of reverberation increased the absolute error for both azimuth and elevation; however, the absolute azimuth error was comparatively less for reverberant targets at right 30° and left 120°. Across all the target azimuths, the greatest differences in error between conditions were for azimuth judgments on the median plane (0 and 180°): the mean azimuth error was lowest under the dry condition (1.6 and 2°) and highest under the reverberant condition (33.2 and 36.8°).

2.3 Distance

All subjects made relative increases in their distance judgments when reverberation was added to the stimuli.

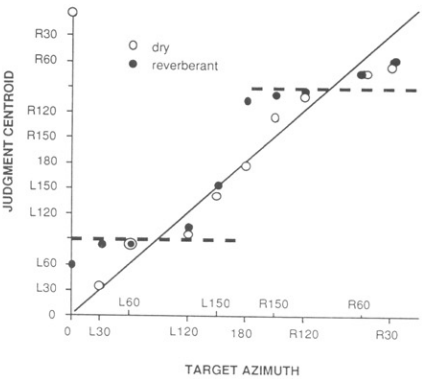

Fig. 5. Scatter plot of centroids based on mean values for one subject whose judgments "pulled" toward left and right 90°.

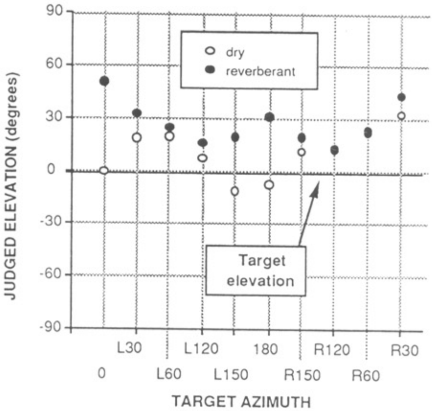

Fig. 6. Elevation centroids (averaged across subjects) for each target azimuth.

Table 3. Mean azimuth errors and standard deviations based on mean centroids of five subjects.

Target azimuth	Mean azimuth error		Standard deviation	
	Dry	Reverberant	Dry	Reverberant
0	1.6	33.2	1.9	19.8
L30	8.0	36.4	6.9	12.7
L60	17.6	17.8	3.1	8.5
L120	14	9.2	11.7	4.0
L150	12.6	15	7.2	12.1
180	2.0	36.8	2.0	22.4
R150	14.0	35.6	9.0	20.0
R120	18.2	22.6	8.0	5.3
R60	11.6	13	6.7	12.2
R30	19.4	9.4	13.7	11.7

Fig. 8 shows mean distance estimates under both conditions for each target azimuth. Note that with the dry stimuli, the judgments for target positions on the median plane (at 0 and 180°) are closer to the center of the head than judgments for other targets. The mean distance for the reverberant stimuli (42 in) is close to the distance of 1 m used in the early reflection simulation model. However, absolute distance judgments varied greatly between subjects. Table 5 gives mean values for each subject's judgments averaged across all azimuths. But in spite of these large differences in absolute distance judgments between subjects, the relative increases between dry and reverberant stimuli are fairly close (ranging from 2.3:1 to 3.8:1).

Another way of viewing the distance judgments is in terms of whether or not the sound was externalized outside of the head (judgments > 4 in) or heard inside or on the edge of the head (judgments ⩽ 4 in). In what follows, "unexternalized" refers to sounds ⩽ 4 in and "externalized" to judgments > 4 in. The right three columns of Table 5 summarize the percentage of unex-

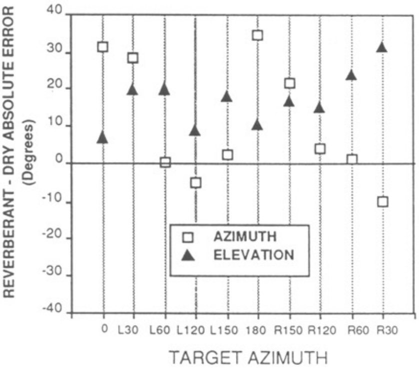

Fig. 7. Difference in absolute errors for azimuth and elevation between reverberant and dry stimuli.

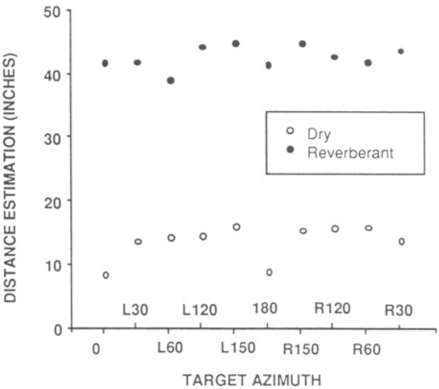

Fig. 8. Mean distance estimates for dry and reverberant stimuli.

Table 4. Mean elevation errors and standard deviations based on mean centroids of five subjects.

Target azimuth	Mean elevation error		Standard deviation	
	Dry	Reverberant	Dry	Reverberant
0	7.0	48.4	6.1	28.7
L30	19.6	32.4	13.7	13.8
L60	20.0	24.6	11.0	12.9
L120	8.8	16.6	9.4	15.8
L150	18.2	26.2	15.2	24.2
180	10.8	30.8	7.3	31.1
R150	16.8	21.6	13.9	23.7
R120	15.2	12.6	15.1	13.5
R60	24	23.4	12.6	20.6
R30	31.6	43.6	21.2	30.9

Table 5. Distance judgment data for five subjects.

Subject	Mean distance estimate (inches)			Unexternalized stimuli† (percent)		
	Dry	Reverberant	Relative increase*	Dry	Reverberant	Chi square
r_5	5	13.1	2.6:1	46	10.50	62.1 ($p < 0.001$)
r_7	5	14.5	2.9:1	59.50	2	151 ($p < 0.001$)
r_8	14.9	34.9	2.3:1	3.50	0.50	—
r_9	26.8	90.4	3.3:1	0	0.50	—
r_{11}	15.3	59.4	3.8:1	15.50	0	33.6 ($p < 0.001$)

* Ratio of reverberant to dry estimates.
† Percentages of all judgments by each subject that were unexternalized (heard less or equal to 4 in) along with chi-square scores and probability levels where differences were significant.

ternalized judgments for each subject, along with chi-square scores and probability levels. Two subjects, r_8 and r_9, externalized all of their judgments, regardless of whether or not there was reverberation present. But for the other three subjects the presence of reverberation was significant for allowing externalized stimuli; particularly for subjects r_5 and r_7 whose mean distance estimates of 5 in for dry stimuli were barely externalized.

3 DISCUSSION

To determine the effects of synthetic reverberation on three-dimensional audio systems, the present study gathered comparative data on perceived azimuth, elevation, and distance of speech stimuli filtered by nonindividualized HRTFs. The results are inherently limited due to the specialized character of the stimuli. In other words, one particular set of nonindividualized HRTFs was used, as well as one particular synthetic reverberant environment. It is possible that parametric variation of either of these variables could yield somewhat different results. Nevertheless, some of the results fall in line with results gathered in other headphone and free-field or real reverberation studies.

The fact that inexperienced subjects were used is an important factor in considering the outcome of the data. No extended period of training or feedback was given, and no improvement in localization error over time was noticed. However, experienced subjects have been observed both informally and formally to perform better than inexperienced subjects [1], [30]. This means that it is probably possible to "learn" to adapt to another set of HRTFs over time. The use of inexperienced subjects could also account for the degree of individual differences in judgments.

The general conclusion regarding azimuth and elevation is that the error between the target and the perceived virtual sound source increases with the presence of reverberation. Nevertheless, the azimuth centroids were only an average of about 12 and 23° off from their intended target positions under dry and reverberant conditions. The increase in azimuth error is possibly due to the early reflection portion of the stimulus. In an actual room study [22] it was found that lateral early reflections influence localization performance. Hence, physiological advantages associated with binaural hearing and the precedence effect allow us to only partially suppress acoustic information from both real and simulated environmental contexts. The increase in the inverse kappa value κ^{-1} for reverberant stimuli (see Table 2) perhaps represents an increase in the "image broadening" (auditory spaciousness) of the stimulus. This could in turn affect the subject's ability to consistently place a "center of gravity" in their azimuth (and elevation) judgments.

Regarding elevation judgments, the bias toward an upward direction above the target with dry and especially with reverberant stimuli is surprising, although this same bias was found previously with dry stimuli [25]. Also, elevation judgments were worse than azimuth judgments under both dry and reverberant conditions. This could possibly be a result of the use of speech stimuli. The most important spectral region for elevation is thought to be above 7 kHz [31], [32]. The spectral energy of speech is relatively less in this region compared to frequencies below 7 kHz. However, one study using HRTFs derived from a KEMAR mannequin head found veridical elevation judgments for speech low-pass-filtered at 5 kHz [33]. There is no suitable explanation for the elevation bias found in the current study.

One rationale for the asymmetrical design of the modeled enclosure used for generating the early reflections was to determine whether reversals could be mitigated by the presence of a unique pattern of early reflections from specific directions. The use of the directionalized "floor" reflections was similarly motivated. Two subjects seemed to have an overall bias toward reversing judgments to the front with reverberation present. Two other subjects were biased toward reverals in directions opposite of each other, independent of stimulus condition. Hence there was no evidence that the directionality of the early reflection pattern used here had any consistent influence across subjects on the percentage of reversed judgments. Since relatively untrained subjects were used, perhaps environmental information could not be used to advantage. It has been shown, for example, that with training both blind and blindfolded subjects can improve their ability for distance perception of objects to a high degree of accuracy through "echolocation" techniques [34].

The fact that reverberation increased perceived distance was not surprising in light of the well-known reverberant-to-direct sound ratio as a cue to distance [35]. It was found here that the synthetic reverberation used provided a fairly consistent relative increase in distance across subjects, but that absolute distance judgments varied widely. There is a possible relationship between externalization and localization accuracy. Specifically, Fig. 8 shows that under the dry condition the 0 and 180° targets were the least externalized, but in Table 3 these targets are perceived with the least amount of absolute azimuth error. Both azimuth errors and externalization increased with the presence of spatial reverberation.

More important is the effect of synthetic reverberation on externalizing HRTF stimuli. In a previous study using the same dry stimuli, 33% of all stimuli were not externalized [25]. In this study, 25% of the dry stimuli were not externalized, compared to only about 3% of the reverberant stimuli. This in itself may be important for the eventual successful implementation of three-dimensional sound systems, since externalization is important for the realism of a virtual sound-source simulation, and is sometimes taken for granted as an inherent perceptual result of HRTF filtering. In previous headphone studies that were not directly concerned with three-dimensional sound techniques, the proportion of externalized localization judgments was shown to increase when either actual or simulated reverberation was added to the stimulus [36], [37].

In summary, a welcome improvement in externalization can be had with the addition of synthetic reverberation, but there is a decrease in localization accuracy, at least with the particular model used in this study. Future work should use parametric variation of synthetic reverberation to explore its effect on the localization errors described here. Another important factor for future research is to evaluate the role of familiarity and training with simulated environmental contexts. Hopefully it will be possible to describe the perceptually salient features of rooms that would allow three-dimensional sound systems to create perceptual experiences found in everyday localization.

4 REFERENCES

[1] D. Griesinger, "Equalization and Spatial Equalization of Dummy-Head Recordings for Loudspeaker Reproduction," *J. Audio Eng. Soc.*, vol. 37, pp. 20–29 (1989 Jan./Feb.).

[2] L. F. Ludwig, N. Pincever, and M. Cohen, "Extending the Notion of a Window System to Audio," *Computer*, vol. 23, no. 8, pp. 66–72 (1990).

[3] D. R. Begault and E. M. Wenzel, "Techniques and Applications for Binaural Sound Manipulation in Human–Machine Interfaces," *Intl. J. Aviation Psychology*, 2(1), 1–22 (1992).

[4] E. M. Wenzel, S. Fisher, P. K. Stone, and S. H. Foster, "A System for Three-Dimensional Acoustic Visualization in a Virtual Environment Workstation," in *Visualization '90* (IEEE Computer Society Press, San Francisco, CA, 1990), pp. 329–337.

[5] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1983).

[6] D. R. Begault, "Challenges to the Successful Implementation of 3-D Sound," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 39, pp. 864–870 (1991 Nov.).

[7] R. A. Butler and K. Belendiuk, "Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane," *J. Acoust. Soc. Am.*, vol. 61, pp. 1264–1269 (1977).

[8] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation," *J. Acoust. Soc. Am.*, vol. 85, pp. 868–878 (1989).

[9] E. M. Wenzel, F. L. Wightman, and S. H. Foster, "A Virtual Display System for Conveying Three-Dimensional Acoustic Information," in *Proc. 32nd Annual Human Factors Society Meeting*. (Human Factors Society, Santa Monica, CA, 1988), pp. 86–90.

[10] J. B. Allen and D. A. Berkley, "Image Model for Efficiently Modeling Small-Room Acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950 (1979).

[11] J. Borish, "Electronic Simulation of Auditorium Acoustics," Ph.D. dissertation, Stanford University, Stanford, CA (1984).

[12] M. Kleiner, "Auralization: Experiments in Acoustical CAD," presented at the 89th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 38, pp. 874–895 (1990 Nov.), preprint 2990.

[13] G. S. Kendall and W. L. Martens, "Simulating the Cues of Spatial Hearing in Natural Environments," in *Proc. 1984 International Computer Music Conf.* (Computer Music Assoc., San Francisco, CA, 1984).

[14] D. R. Begault, "Control of Auditory Distance," Ph.D. dissertation, University of California San Diego, La Jolla, CA (1987).

[15] H. W. Gierlich and K. Genuit, "Processing Artificial-Head Recordings," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 37, pp. 34–39 (1989 Jan./Feb.).

[16] F. Richter and A. Persterer, "Design and Applications of a Creative Audio Processor," presented at the 86th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 35, p. 398 (1989 May), preprint 2782.

[17] H. Haas, "The Influence of a Single Echo on the Audibility of Speech," Reprinted in *J. Audio Eng. Soc.*, vol. 20, pp. 146–159 (1972 Mar.).

[18] M. B. Gardner, "Some Single- and Multiple-Source Localization Effects," *J. Audio Eng. Soc.*, vol. 21, pp. 430–437 (1973 July/Aug.).

[19] P. M. Zurek, "Measurements of Binaural Echo Suppression," *J. Acoust. Soc. Am.*, vol. 6, pp. 1750–1757 (1979).

[20] G. Plenge, "Über die Hörbarkeit kleiner Änderungen der Impulsantwort eines Raumes" (On the Audibility of Small Changes in the Impulse Response of a Room), *Acustica*, vol. 25, pp. 315–325 (1971).

[21] W. M. Hartmann, "Localization of Sound in Rooms," *J. Acoust. Soc. Am.*, vol. 74, pp. 1380–1391 (1983).

[22] W. M. Hartmann and B. Rackerd, "Localization of Sound in Rooms II: The Effects of a Single Reflecting Surface," *J. Acoust. Soc. Am.*, vol. 78, pp. 524–533 (1985).

[23] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis," *J. Acoust. Soc. Am.*, vol. 85, pp. 858–867 (1989).

[24] E. M. Wenzel, F. L. Wightman, and D. J. Kistler, "Localization of Non-Individualized Virtual Acoustic Display Cues," in *Proc. CHI'91 ACM Conf. on Computer–Human Interaction* (Assoc. for Computing Machinery, New York, 1990), pp. 351–359.

[25] D. R. Begault and E. M. Wenzel, "Headphone Localization of Speech Stimuli," in *Proc. 35th Annual Human Factors Society Meeting* (Human Factors Society, Santa Monica, CA, 1991), pp. 82–86.

[26] M. R. Schroeder, "Ditigal Simulation of Sound Transmission in Reverberant Spaces," *J. Acoust. Soc. Am.*, vol. 47, pp. 424–431 (1970).

[27] K. D. Kryter, "Speech Communication," in H. P. V. Cott and R. G. Kinkade (Eds.), *Human Engineering Guide to Equipment Design* (McGraw-Hill, Washington, DC, 1972), pp. 162–226.

[28] S. R. Oldfield and S. P. A. Parker, "Acuity of