Early Reflection Thresholds for Anechoic and Reverberant Stimuli within a 3-D Sound Display

Durand R. Begault[a]
Bryan U. McClain[b]
Mark R. Anderson[c]
[a]Human Factors Research & Technology Division
[b]San José State University Foundation, [c] QSS Group, Inc.
[a, b,c] Spatial Auditory Display Laboratory
MS 262-2, NASA Ames Research Center,
Moffett Field, California 94035-1000 USA
[a]Durand.R.Begault@ nasa.gov

Abstract

Data on auditory thresholds for virtual acoustic reflections were obtained from 9 participants as a function of spatial position, time delay and stimulus type in a simulated 5.1 surround sound listening room environment. First-order reflections (3, 15 and 30 ms) were determined from a ray-tracing model of the listening room; their level was manipulated to simulate the overall effects of absorptive treatment. The direction of the reflection varied from 0 – 164 degrees offset from the direction of the simulated direct sound (corresponding to either the center or the right surround channel). Absolute thresholds (perception of any type of change) were measured at the 70.7% level using a one up-two down staircase algorithm, for anechoic and reverberant speech stimuli, and for tone burst stimuli (125, 250, 500, 1k, 2k and 4k Hz). For anechoic speech and tone stimuli, the threshold was 12 – 31 dB below the level of the direct sound; the addition of a reverberant decay (mid-band T30 = .6 s) raised thresholds by 7 dB. The results were in good agreement with previous threshold studies using real sound sources. The information is useful for determining engineering parameters for the real-time simulation of virtual acoustic environments, such as head-mounted displays that include head tracking.

1. Introduction

A well-known method for characterizing the acoustical characteristics of a room is to measure the response at a particular microphone position to a brief source of energy, such as a pistol shot or a balloon burst. The use of a deterministic signal (e.g., maximum length sequence, sine sweep) is also possible via post-processing of the signal. From the perspective of room acoustic quality, the end result usually involves visual inspection of a graphic display of the "room impulse response", i.e., the squared pressure of the real part of the analyzed signal in decibels as a function of time. A similar sort of graphic can be obtained from a modeling program that uses ray tracing or other techniques for predicting, rather than measuring, the room impulse response. This information can be used for both analyzing the acoustics of a real room or for simulation of the acoustics of a virtual room.

In both applications, post-analysis of the reflection amplitudes relative to the level of the direct sound determines their significance in terms of audibility. Early reflections are well-known to be potentially detrimental to timbre reproduction, speech intelligibility, and the formation of spatial images in a loudspeaker sound field.

Auditory thresholds for early reflections have been reported by various workers using real sound sources [1-3]. The current study uses virtual simulation of real sources ('auralization' technique) for simulating direct and reflected sources corresponding to loudspeaker locations within a 5.1 listening room configuration. The correspondence between real and virtual sound source thresholds allows an estimate of the auralization technique's capacity to predict perceptual responses to more complex room models for both psychoacoustic investigations and sound quality evaluation. Establishment of thresholds for early reflections is pertinent to determining necessary absorptive treatment for building acoustic treatment. Another goal previously described in [4] is for management of computational resources for real-time auralization systems.

2. Methodology, subjects

Absolute thresholds were determined for time-delayed speech and tone burst signals, relative to a non-delayed version of the same signal corresponding to an acoustic "direct path". The delayed signals, corresponding to acoustic "reflections" within an enclosure, were manipulated in terms of both time delay and location between experimental blocks; the reflection level threshold was the dependent variable.

Nine subjects participated in the speech threshold experiments, and nine subjects participated in the tone burst threshold experiment. All were screened for normal hearing prior to participating in the experiment. Experimental blocks were conducted in double-walled soundproof booth having a background noise level of 15 dB (A-weighted).

Speech stimuli were formed from one of 36 randomly chosen anechoic speech segment .wav files 1.3 s in duration [5]. Tone burst stimuli were formed from one of 6 randomly chosen 80 ms duration, amplitude-ramped sinusoid .wav files that corresponded to octave-band center frequencies at 125, 250, 500, 1k, 2k and 4k Hz. The amplitudes of the sinusoidal stimuli were normalized to an equal loudness level of 65 phons. Stimuli were presented at a level of 65 dB (A-weighted) via stereo headphones (Sennheiser HD 430).

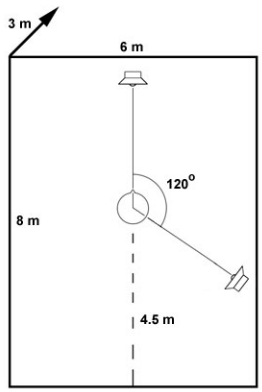


Figure 1. Layout of modeled room, listener position, and direct sound source configurations used in the experiment. Reflections correspond to boldface data in Table I.

Azimuth-elevation angles (referenced to 0° at a point directly in front of the listener) were simulated via real-time head-read transfer function (HRTF)-filtering. The SLAB real-time, software-based 3-D audio processor developed at NASA Ames Research Center was used [6]. An additional computer drove the experimental software that communicated to the SLAB server via a tcp/ip connection and gathered data from the subject via a two-button switchbox interfaced to the mouse serial port.

A room modeling software package (Odeon 4.0) was used to obtain image model reflection timings and azimuths for a surround sound loudspeaker array within a room conforming to listening test standards (ITU). The room dimensions were 8 x 6 x 3 m, with the listener centered between the loudspeaker array and the left and right walls, 4.5 m from the back wall (see Figure 1). Loudspeakers were modeled at 0° and 120° azimuth, corresponding to "center" and "surround" channels. For each direct path, 1^{st} and 2^{nd} order reflections were selected (ref. Table I). To establish reflection delay time as an independent variable, the derived azimuth and elevation for a given reflection was subsequently investigated at 3, 15, and 30 ms. Specifications in the "Az. Dif." column correspond to the inside angle subtended on the horizontal plane between the direct and reflected sound azimuths. The maximum lateral azimuth difference between the direct sound and the reflection is for the 72 and 164-degree azimuth difference angles (indicated in bold).

Table I. Experimental conditions. Time delays in bold type correspond to the room model results

Time delay ms.	Direct Az. (all at 0 elevation).	Reflection Az.	El.	Az. Dif.	Reflection surface
3, 15, 30	0	0	- 50	0	Floor
3, **15**, 30	0	0	72	**72**	Right wall
3, 15, **30**	0	0	151	151	Back wall
3, 15, 30	120	120	-50	0	Floor
3, **15**, 30	120	72	0	48	Right wall
3, 15, **30**	120	-76	0	**164**	Left wall

Using a two-alternative forced-choice paradigm, thresholds were obtained at the 70.7% level within a tolerance of 1 dB with a "one up-two down" adaptive staircase algorithm that adjusted the level of the reflection [7]. The reference stimulus consisted of one of the randomly chosen anechoic speech or tone bursts (a "direct path only" stimulus), while the probe stimulus consisted of the same stimulus with a direct path plus an amplitude-scaled reflection. Three sequential stimuli were presented; first the reference, and then either probe followed by reference, or reference followed by probe. The ordering of the last two stimuli was randomized between trials. For each trial, participants indicated their response via the push-button interface as to which of the final two stimuli were "different" from the first stimulus.

The reflection was initially presented at –4 dB relative to the direct sound. The staircase began with an 8 dB step size, and reduced in level by 50% until the 1 dB step size was reached. The staircase terminated after a total of eight "reversals" in direction. Thresholds were defined for each subject and for each block as the mean value of the five final staircase reversals at the minimum level of 1 dB.

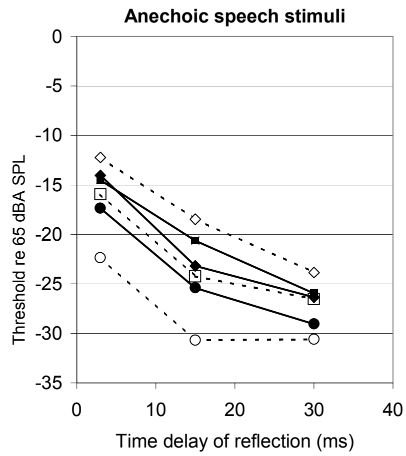
For speech stimuli, subjects were run under each of the time-location configurations indicated in Table I using both "anechoic" and "reverberant" stimuli conditions, for a total of 36 blocks. Block ordering was randomized across subjects. Anechoic stimuli included simulation of only the direct sound and a single reflection. Reverberant stimuli were generated via convolution of the direct sound with a synthetic reverberation decay, formed from exponentially-decaying white noise decorrelated between the left-right channels and at a level –20 dB below the direct sound. This corresponds to a non-acoustically damped version of the modeled room. The mid-band reverberation time corresponded to 0.63 s.

For tone burst stimuli, subjects were run under a subset of the direct and reflection azimuth-elevation locations, excluding the azimuth difference conditions at 151 and 48 degrees in Table I. All time delay conditions were used. The remaining conditions corresponded to the minimum (azimuth difference = 0 degrees) and maximum values (azimuth difference = 72, 164 degrees) for lateral azimuth difference.

3. Results

Figure 2 indicates mean values of the results across nine subjects for anechoic and reverberant speech stimuli. For both anechoic and reverberant stimuli, thresholds decrease monotonically with increasing time delay between the direct sound and the reflection. Compared to anechoic stimuli, thresholds are increased for reverberant stimuli by an average of 7 dB (range 3-11 dB). With increasing time delay, reverberant stimuli thresholds decrease less compared to anechoic stimuli.

For a given direction of the direct sound, increasing magnitude of the azimuth angle difference of the reflection (see Table I) generally corresponds to a decrease in thresholds. In the direct sound at 120° and reflection at -76° condition, i.e., the maximum lateral difference condition tested, thresholds for both anechoic and reverberant stimuli decrease by 7—15 dB, compared to when the reflection and direct sound are co-located. Comparatively, when the direct sound is at 0°, the effect of the azimuth angle difference is diminished. For example, the threshold for an anechoic reflection co-located with the direct sound at 0° azimuth and a time difference of 3 ms (corresponding to a floor reflection in the modeled room) is –14 dB, but decreases to only –17 dB when the reflection is located at 151°.



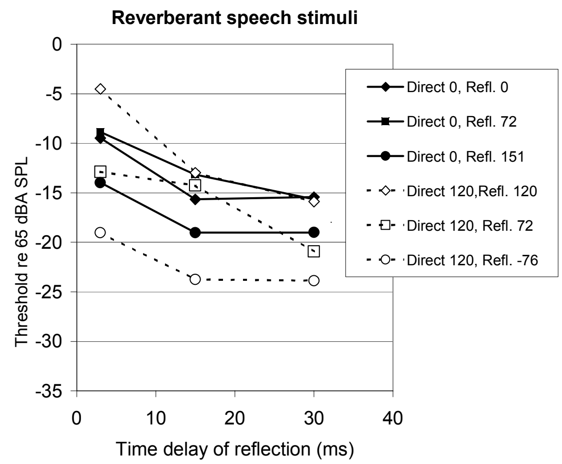


Figure 2. Mean threshold values for 9 subjects for anechoic and reverberant speech conditions. "Direct" refers to angle of direct sound in degrees, "Refl.." refers to the azimuth of the reflected sound in degrees (ref. Table I).

Figure 3 indicates results for tone burst stimuli. The thresholds are on average within 3 dB of equivalent speech stimuli. As with speech stimuli, the effect of spatially separating the direct and reflected sound is most apparent with the direct sound 120°, reflection -76° condition. However, there is only about a 5 dB decrease in threshold levels at 3 and 15 ms time delays. At 30 ms, the threshold levels are nearly the same across conditions.

For all stimuli, the lowest thresholds are for the direct sound at 120°, reflection at –76º. This represents a direct sound coming from the right rear surround loudspeaker and a reflection arriving from the left wall. For this case, the direct sound has a relatively high interaural time difference with a left ear lead-right ear lag, and for the reflection the same high interaural time difference but with a right ear lead-left ear lag. This situation represents the maximum lateral difference between the direct sound and the reflection, and has the lowest interaural cross-correlation for subjects. Under these conditions, it is likely that subjects attended to a binaural cue (image broadening) for that class of stimuli, which may be easier to detect compared to ascertaining the timbre cue present when the direct and reflected sound were azimuthally co-located at 0°, or separated by a smaller angle but impinging towards the same side of the head.

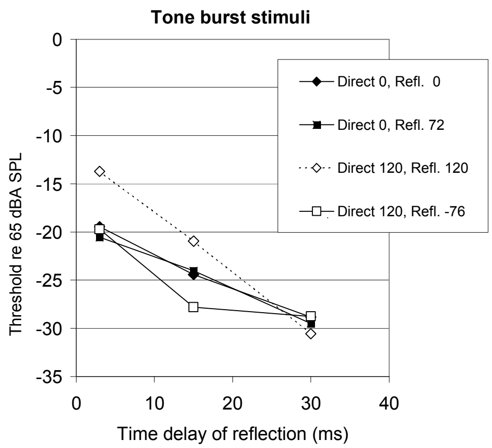


Figure 3. Mean threshold values for 9 subjects for tone burst stimuli conditions.

4. Discussion

There are many definitions for defining the concept of a "reflection" or "echo" threshold in the literature. This is in addition to the particular configuration of reflection angles, time delays, stimuli used, or methodology employed in a particular study. A comparison of thresholds must take all of these differences into account. Haas used the criteria of 'echo disturbance' in relationship to speech [8]. The 'echo threshold' as defined in [9] refers to the level at which a echo is perceived as a separate auditory event, whereas the 'image shift' threshold refers to a just-noticeable change in the spatial location of an auditory image.

For telecommunications applications, the threshold definition shifts to speech intelligibility and/or perception of inter-modal asynchrony, depending on the application. In applications related to audio reproduction, thresholds that influence the perception of audio quality become of interest, including spatial and timbral thresholds. Bech investigated reflection thresholds for changes in timbre, specifically for a pattern of reflections applicable to a listening room environment [1, 2]. Olive and Toole [3] and the present study focused on the "masked" or "absolute threshold", where the perception of any change in the stimulus is used as the definition of the threshold. One practical advantage to the absolute threshold is that subjects require no special training to discriminate between specific perceptual aspects of stimuli; *any* perceived change is a valid basis for indicating a "different" response in a two-alternative forced choice paradigm.

Overall, there is good agreement between the present results and studies by Bech [1,2], Olive and Toole [3] and Seraphim [10], for both speech and tone burst versus click stimuli. However, Olive and Toole's data indicates a 20 dB lower threshold for click stimuli at 30 ms (-50 dB). This may be due to the fact that their click stimuli extended across the full audio spectrum while the tone burst stimuli used here were band-limited. There may have also been a lower background noise level in their anechoic chamber compared to the background noise level in our soundproof booth (15 dBA).

The data presented here can be used, for many applications, to form "rules of thumb", such as: (1) early reflections will be inaudible when less than 22 dB below the direct sound at 3 ms, and less than 31 dB below the direct sound at 15-30 ms; (2) a modest amount of reverberation added to anechoic speech stimuli (reverberant-direct ratio of –20 dB) increases thresholds by up to 11 dB.

5. References

[1] Bech, S., "Timbral aspects of reproduced sound in small rooms. I" *J. Acoust. Soc. Am.*, 1995. 97: 1717-26.

[2] Bech, S., Timbral aspects of reproduced sound in small rooms. II" *J. Acoust. Soc. Am.,* 1996. 99: 3539-49.

[3] Olive, S.E. and F.E. Toole, "The detection of reflections in typical rooms" *J. Audio Eng. Soc.*, 1989. 37; 539-53.

[4] Begault, D.R., "Audible and inaudible early reflections: thresholds for auralization system design". *Audio Eng.Soc. 100th Convention (preprint 4244)*, 1996.

[5] *Music from Archimedes* (audio CD B&O 101) Bang and Olufsen, 1992.

[6] http://humanfactors.arc.nasa.gov/SLAB/

[7] Levitt, H., "Transformed up-down methods in psychoacoustics" *J. Acoust. Soc. Am.,* 1970. 49: 467-77.

[8] Haas, H., "The influence of a single echo on the audibility of speech". *J. Audio Eng. Soc.* 1972. 20: 146-159.

[9] Blauert, J., *Spatial hearing: The psychophysics of human sound localization.* Revised Edition ed. 1997, Cambridge: MIT Press.

[10] Seraphim, H., On the perceptibility of multiple reflections of speech sounds. *Acustica*, 1961. 11: 80-91.