AUDIBLE AND INAUDIBLE EARLY REFLECTIONS: THRESHOLDS FOR AURALIZATION SYSTEM DESIGN

Durand R. Begault, Ph.D.
San José State University
Flight Management and Human Factors Research Division
NASA Ames Research Center
Mail Stop 262-2
Moffett Field, California 94035-1000
Tel: (415) 604 3920 Fax: (415) 604 3323
Email: db@eos.arc.nasa.gov

ABSTRACT

This paper gives thresholds for direct sound and early reflections simulated by a virtual reality system. An auralization system would ideally utilize a large number of accurately-modeled early reflections parameters, but there are practical DSP limitations for real-time rendering. A perceptual solution for data reduction involves determining early reflection thresholds as a function of time and spatial incidence.

1. OVERVIEW

In a previous paper by the author [1], auralization system calculations made with reference to a fairly typical set of room/listener/sound source configurations were shown to yield early reflections below auditory threshold, based on data from [2-4]. In rendering an auralizable version of a room model, a complex matrix of digital filter impulse responses must be calculated that include the transfer functions of the pinnae and wall surfaces. The complexity of digital signal processing (DSP) parameters for auralization rendering is proportional to the number of early reflections modeled; see [5-7] for different approaches. The DSP computational limit can become quickly exhausted, particularly in real-time systems where digital signal processing (DSP) parameters must be updated in response to a head-tracking device such as described in [8].

The overall goal of this and our future early reflection studies is to determine absolute thresholds as a function of angle of incidence and temporal distribution.

Prior to auralizing a room modeling program, one can save computational resources by determining if individual reflections are indeed audible.

There is ample evidence that the absolute threshold for an early reflection changes as a function of angle of incidence and sound source type [2, 3, 9-11], and as a function of forward or backward masking (see [4]). For example, Bech [10, 11] examined early reflection thresholds in the context of a loudspeaker-based simulation system, using 17 early reflections and a simulated diffuse field; a criteria of including only those reflections with intensity greater than –20 dB was used. The current study is similar to that described in [11] in the determination of the "threshold of detection": the goal being to determine at what level of intensity is an early reflection indistinguishable from being completely absent, in the presence of other reflected energy. However, a complete reverberant field is absent in the current study; the conclusion of [11] was that sensitivity may be increased anywhere from 2 to 5 dB when comparing early reflections without the presence of a dense reflection field ("late" reverberation).

Below, the results from two preliminary investigations are given. In the first investigation, the absolute threshold was examined for individual listeners, using non-individualized HRTFs and anechoic music stimuli, for both a single reflection and a pattern of three reflections. In the second investigation, individualized HRTFs were measured for each of four subjects, and then applied to two patterns of early reflections representing a "large room" and a "small room." Anechoic speech stimuli were used in the second investigation. Variables included the angle of reflections ("wide" or "narrow") and which particular reflection was evaluated.

2. METHOD

While an infinite number of temporal and spatial distributions exist for early reflections, the current investigations used a simplified model, in that only HRTF processing and amplitude scaling is applied; no filtering for simulating wall surfaces or angle of incidence is included. Additional filtering would lower the intensity and therefore most likely increase the threshold for any of the reflections examined in the current context. The current studies are further restricted in that only one direct sound and 3 early reflections were HRTF filtered; and that a limited number of azimuths were evaluated. The direct sound was always synthesized to a virtual position directly ahead of the listener (0° azimuth, 0° elevation); details on the reflection angles are given below.

Absolute thresholds were determined at a 70.7% level within a tolerance of 1 dB, using a staircase algorithm described in [12]. The threshold is defined for each subject as the mean of 5 staircase direction reversals at the 1 dB level. No special training was given subjects to adopt a particular criteria; their task was to listen in terms for <u>any</u> detectable difference. They were informed in a pre-experiment briefing that they might hear spatial, timbral or loudness differences.

A three-alternative forced-choice paradigm was used, where the subject was asked to identify which of three stimuli heard in succession is different from the other two. Two of these are reference stimuli and one is the probe stimulus, with presentation randomized. The probe is the direct sound and reflections at an initial level and azimuth, such as illustrated in Figures 1 and 2; the level of one of the reflections is adjusted by the staircase algorithm. Any one of the reflections may be the target reflection for staircasing within a particular experimental block. The reference stimulus is the same, except that the reflection being investigated was absent.

Each stimuli consisted of 3-4 sec of spatially-processed anechoic music (first investigation) or speech stimuli (the second investigation) from a single, randomly chosen sound file [13, 14]. The hardware system (see Figure 3) included a Crystal River Engineering Acoustetron capable of spatializing 4 sound sources. Three copies of the sound source were passed through MIDI-controlled Sony DPS-D7 digital delay units (20-bit quantization, 44100 kHz sampling rate) before arriving at the Acoustetron, while another version passed directly to it to simulate the direct sound.

In the first investigation, we used a minimum-phase, non-individualized Head-Related Transfer Functions (HRTFs) of a “good localizer.” In the second investigation, we used individualized HRTFs, obtained using a Crystal River Engineering “Snapshot” HRTF measurement system (see Figure 4). This system uses a blocked meatus technique with a Golay-code–pseudo random signal, along with post-processing to remove effects of the listening environment, loudspeaker and microphones. This allows measurement within a non-anechoic environment, since the post-processing windows the direct sound portion of the signal. It is generally assumed that the use of one's own pinnae transform functions allows for optimal localization (see summary in [15]), although this assumption was not tested directly here.

Table I summarizes the experimental conditions used in the first investigation. First, the threshold for a single reflection delayed 18 msecs was obtained, at 90° and 180° azimuth. Second, the threshold for each one of a pattern of reflections at 10, 14, & 18 msec was evaluated, with the angle of incidence as the independent variable (45°, 90° and 180°)—see Figure 1. The relative level of the reflections in relationship to the probe was –18 dB. Direct sound and reflections were filtered by non-individualized HRTFs.

Figure 2 shows the configuration of the distribution of reflections for the probe stimulus in the second investigation. This configuration of timings and levels were obtained from a ray tracing technique of first-order reflections in a simple room model. The first reflection was used to simulate a floor reflection at 0 degrees azimuth, –36° elevation. The second or third early reflections were set to either "narrow" or "wide" incidence, meaning that the lateral incidence was either set to left and right 30 degrees, or left and right 90 degrees, all at 0° (ear-level) elevation. This was to determine if an increase in interaural time delay would correspond to a lowered threshold. In addition, the threshold for the first reflection was obtained in the presence of both narrow and wide distributions of lateral reflections. All reflections were filtered by the HRTFs of the individual subject.

3. RESULTS AND CONCLUSIONS

In Figure 5, the difference in thresholds for a single spatialized early reflection is evident between the three subjects. Individual differences are apparent; note that subject 3's highest threshold is about the same level as the lowest threshold for subject 1 and 2. In Figure 6, preliminary data for two subjects listening to a pattern of reflections at 10, 14, & 18 msec is shown, with the angle of incidence as the independent variable. In both Figure 5 and 6, it appears that the interaural time delay present in the reflections at 45° or 90° assists in lowering the threshold, compared to the almost 0 interaural time delay present in the 180° reflection. This observation is supported in the second investigation as well.

In the second investigation, a 3-way analysis of variance (ANOVA) using each subject's 1 dB reversal values as the dependent variable was conducted. This analysis (*Room size–reflection number–incidence angle)* was conducted to determine if a significant interaction existed between the size of room used, the sequential order of the reflection, and the angle of incidence. Only one significant difference was found, between narrow (left and right 30°) and wide (left and right 90°) lateral angles of incidence, F (1, 3) = 34.3, p = 0.01; see Figure 7. However, the mean values are so close—narrow angle: –18.5 dB (*SD*, 3.1), wide angle: –20.4 dB (*SD*, 2.9)—as to be negligible in practical applications. The grand mean for the threshold is –19.45 dB (*SD*, 3.11), which is very close to the –20 dB cut-off level used in [11]. For comparison, we also obtained the threshold of the floor reflection in the presence of both small and wide reflection patterns; the threshold is higher, –14.5 dB (*SD*, 4.2), indicating, as with the first investigation, the release from masking of a reflection occurs with the introduction of a significant interaural time difference.

Based on the individual differences seen in the first investigation, it is recommended that the threshold for inclusion of an early reflection be based on a dB value lower than the mean; e.g., the lower boundary of the first standard deviation. Based on the grand mean and standard deviation found in the second investigation, it would be prudent to assume that individual, isolated early reflections below –23 dB need not be calculated. This threshold will certainly rise (1) in the presence of a dense pattern of reflections, and or (2) for angles of incidence with minimal interaural time difference (0° azimuth, –36 elevation, or 180° azimuth, 0° elevation). As more data is obtained, the beginning of a set of guidelines for this particular approach to computational efficiency in auralization systems should be possible.

ACKNOWLEDGMENTS

This study was made possible in part by the contributions of Elizabeth Wenzel (NASA-ARC); Kevin Jordan (SJSUF); Joel Miller (programmer, Sterling Software); and Rick Shrum (experiment manager, Sterling Software).

REFERENCES

[1] Begault, D.R., "Binaural Auralization and Perceptual Veridicality". In *Audio Engineering Society 93rd Convention* Preprint No. 3421 (1992). New York: Audio Engineering Society.

[2] Zurek, P.M., "Measurements of binaural echo suppression". *Journal of the Acoustical Society of America*, vol. 66, pp. 1750-1757 (1979).

[3] Olive, S.E. and F.E. Toole, "The detection of reflections in typical rooms". *Journal of the Audio Engineering Society*, vol. 37, pp. 539-553 (1989).

[4] Blauert, J., *Spatial hearing: The psychophysics of human sound localization*. (MIT Press, Cambridge, 1983).

[5] Kleiner, M., B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview". *Journal of the Audio Engineering Society*, vol. 41, pp. 861-875 (1993).

[6] Reilly, A., D. McGrath, and B.-I. Dälenback, "Using auralisation for creating animated 3-D sound fields across multiple speakers". In *AES 99th Convention* preprint 4127 (1995). New York: Audio Engineering Society.

[7] Reilly, A. and D. McGrath, "Convolution processing for realistic reverberation". In *AES 98th Convention* preprint 3977 (1995). New York: Audio Engineering Society.

[8] Foster, S.H., E.M. Wenzel, and R.M. Taylor, "Real-time synthesis of complex acoustic environments (Summary)". In *Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics* (1991).

[9] Ebata, M., T. Sone, and T. Nimura, "On the perception of direction of echo". *Journal of the Acoustical Society of America*, vol. 44, pp. 542-547 (1968).

[10] Bech, S., "Perception of reproduced sound: Audibility of individual reflections in a complete sound field, II". In *Audio Engineering Society 99th Convention* Preprint 4093 (1995). New York: Audio Engineering Society.

[11] Bech, S., "Timbral aspects of repoduced small rooms. I". *Journal of the Acoustical Society of America*, vol. 97, pp. 1717-1726 (1995).

[12] Levitt, H., "Transformed up-down methods in psychoacoustics". *Journal of the Acoustical Society of America*, vol. 49, pp. 467-477 (1970).

[13] European Broadcast Union, *Sound quality assessment material recordings for subjective tests (EBU SQAM)*. Polygram: Hanover, Germany (1988).

[14] Denon, *Anechoic Orchestral Music Recording*, , M.E.O.P. Orchestra, Editor. DENON/Nippon Columbia Co.: Japan (1987).

[15] Begault, D.R., *3-D Sound for Virtual Reality and Multimedia*. Academic Press Professional, Cambridge, MA. (1994).

Figure reference	Direct	Refl. at 10 ms	Refl at 14 ms	Refl at 18 ms
5	0°	n/a	n/a	S
6A	0°	S	R90°	L90°
6B	0°	L90°	S	R90°
6C	0°	L90°	R90°	S

TABLE 1. Experimental conditions: reference and probe configurations for the first investigation. Key: n/a= not applicable (the early reflection was absent in both probe and reference); S = the early reflection was absent in the reference; but staircased in the probe, at the azimuths indicated in Figure 1.

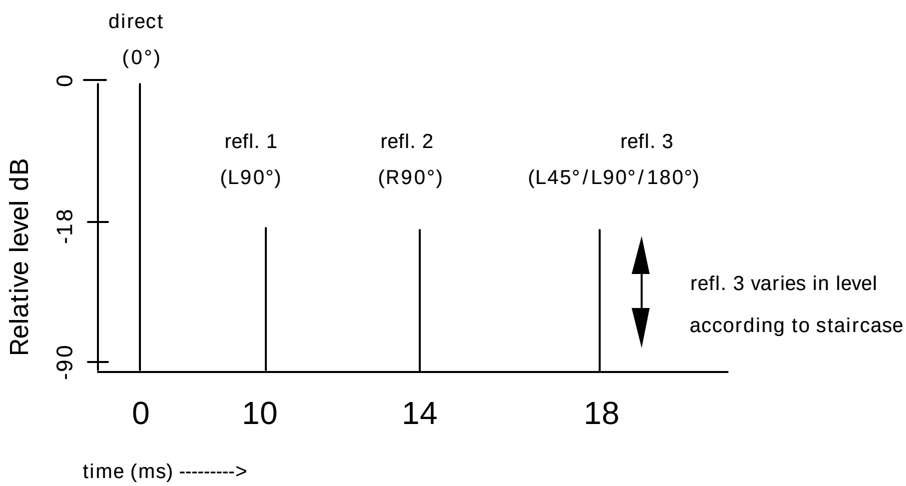


FIGURE 1. Configuration of the distribution of reflections for the probe stimulus in the first investigation. The level of any of the early reflections might be staircased within an experimental block. The reference stimulus was the same except that the staircased reflection was absent.

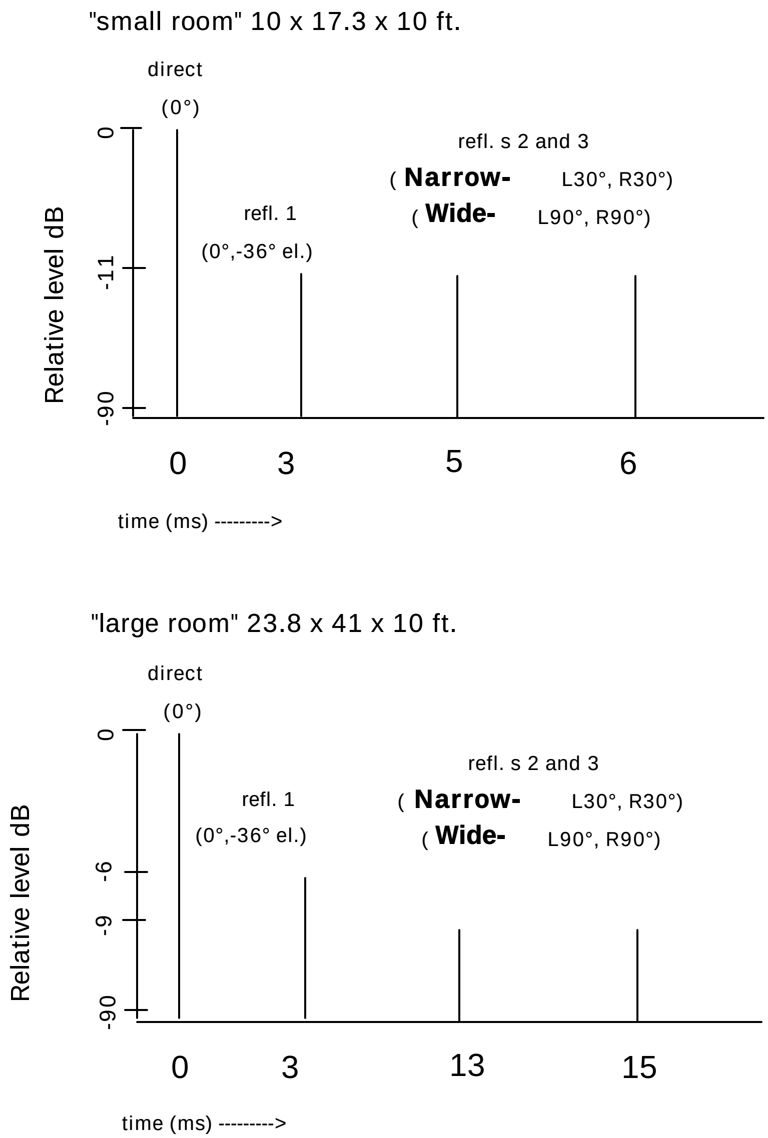


FIGURE 2. Configuration of the distribution of reflections for the probe stimulus in the second investigation. The second or third early reflections were staircased; either a “narrow” or “wide” distribution of reflections were used. The first reflection was used to simulate a floor reflection. Timings and levels were obtained from a ray tracing technique of first-order reflections in a simple room model. The second and third reflections were at 0° (ear-level) elevation.

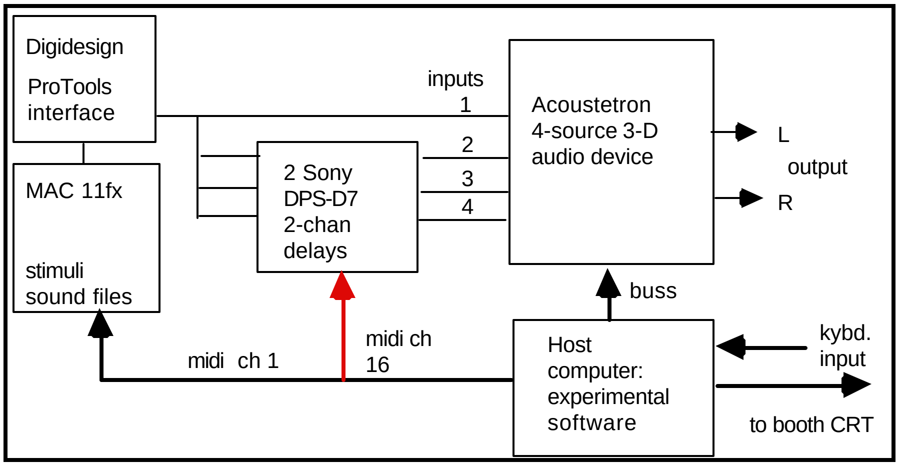
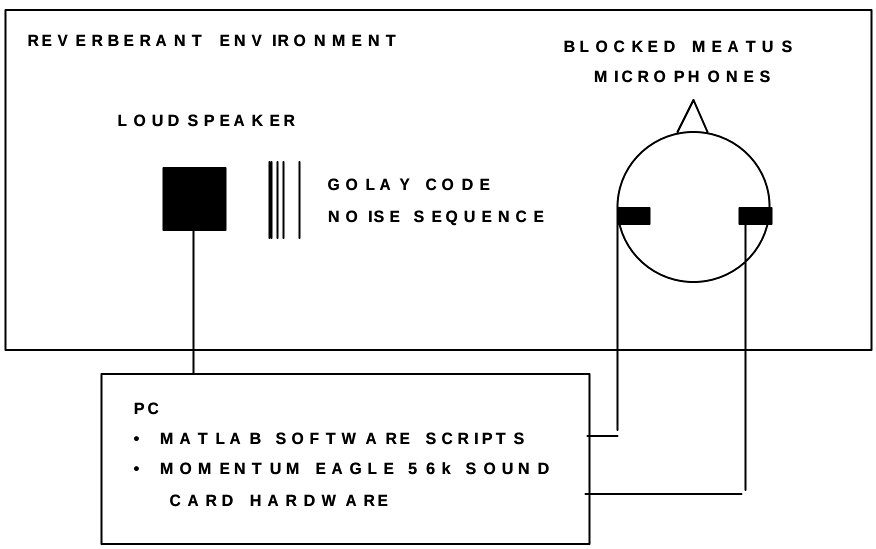


FIGURE 3. Experimental hardware setup.



SNAPSHOT HRTF MEASUREMENT SYSTEM

FIGURE 4. The Crystal River Engineering “Snapshot” system allows measurement of diffuse-field HRTFs in a reflective environment. Early reflections from the measurement environment are removed, allowing replication of an anechoic measurement.

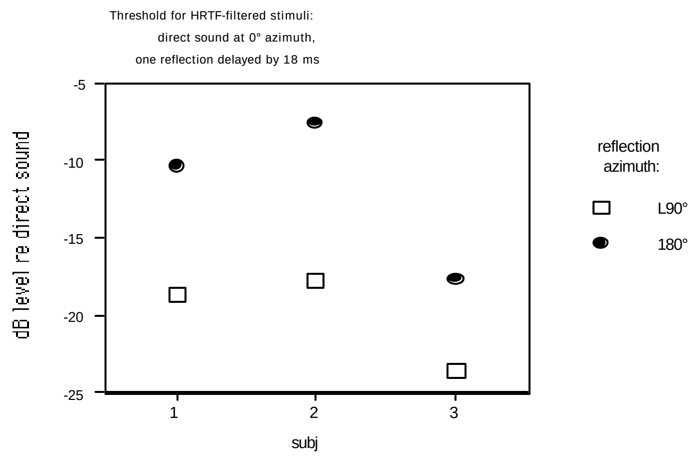


FIGURE 5. Illustration of threshold for a single reflection at two different azimuths, three different subjects. Note that subject 3's highest threshold is about the same level as the lowest threshold for subject 1 and 2.

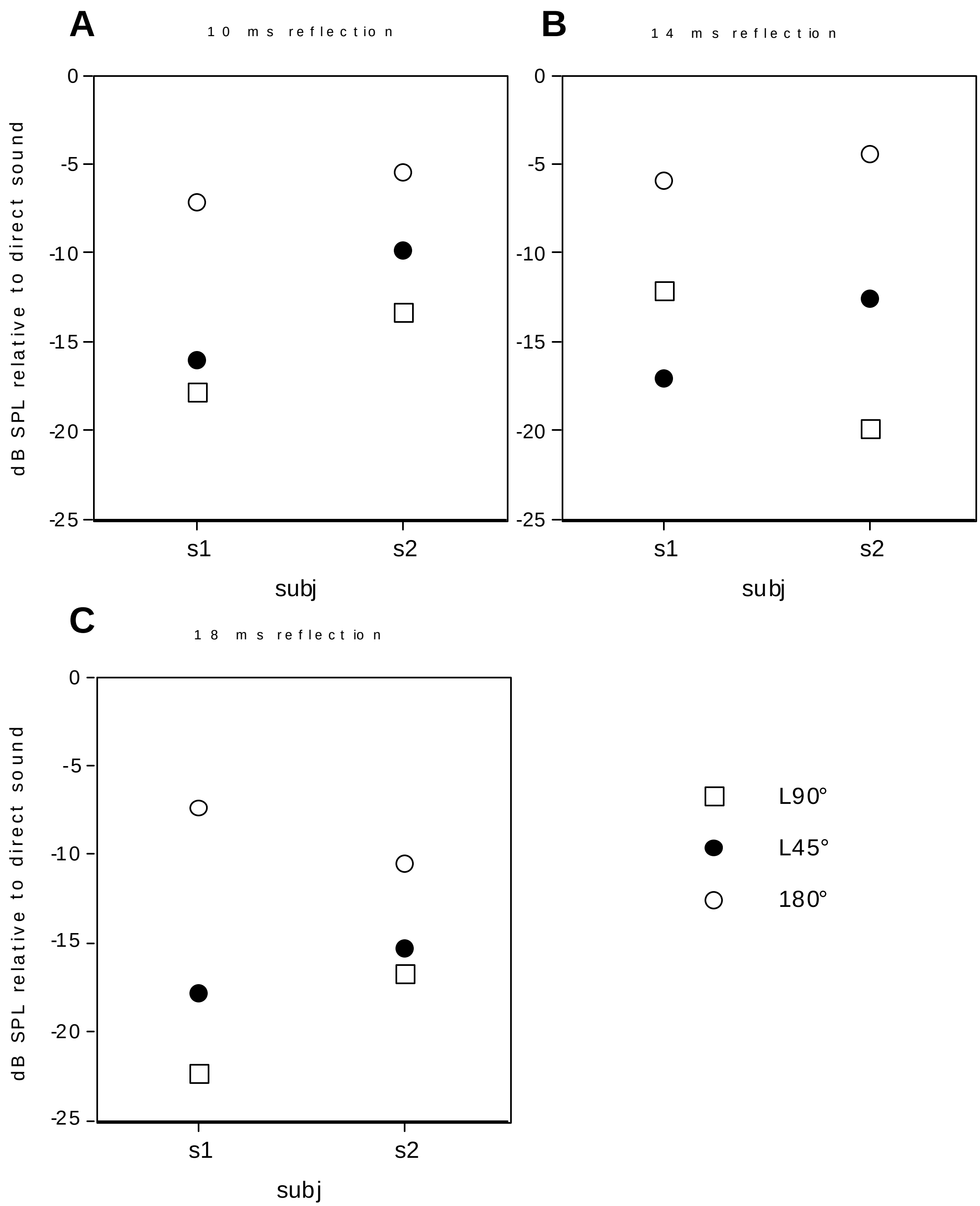


FIGURE 6. Data for two subjects from the first investigation, showing results for the first, second and third early reflections as a function of angle of incidence (see Table I).

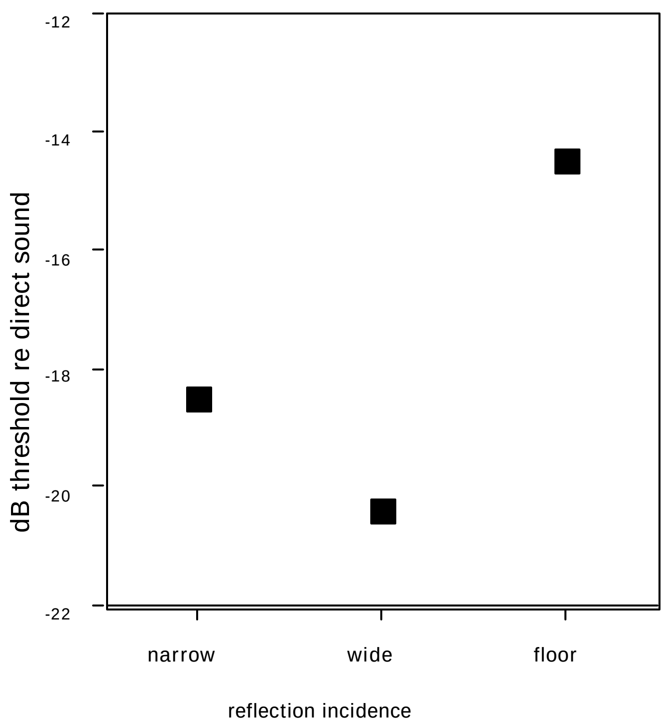


FIGURE 7. Data for four subjects from the second investigation, showing results for "narrow" (lateral reflections at 30°) and "wide" (lateral reflections at 90°) early reflections. For comparison the threshold is given for the first "floor" reflection.