Octave-band Thresholds for Modeled Reverberant Fields

4662 (P4-5)

Durand R. Begault
NASA Ames Research Center (Flight Mgt. & Human Factors Div.) Moffett Field USA

Elizabeth M. Wenzel
NASA Ames Research Center (Flight Mgt. & Human Factors Div.) Moffett Field USA

Laura L. Tran
NASA Ames Research Center (Flight Mgt. & Human Factors Div.) Moffett Field USA

Mark R. Anderson
NASA Ames Research Center (Flight Mgt. & Human Factors Div.) Moffett Field USA

Presented at the 104th Convention
1998 May 16-19
Amsterdam

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

OCTAVE-BAND THRESHOLDS FOR MODELED REVERBERANT FIELDS

Durand R. Begault[1]
Elizabeth M. Wenzel
Laura L. Tran
Mark R. Anderson
Flight Management and Human Factors Division
MS 262-2 NASA Ames Research Center
Moffett Field, California 94035-1000

(Preprint for Audio Engineering Society 104th Convention, 16-19 May 1998)

Abstract

Auditory thresholds for 10 subjects were obtained for speech stimuli reverberation. The reverberation was produced and manipulated by 3-D audio modeling based on an actual room. The independent variables were octave-band-filtering (bypassed, 0.25 - 2.0 kHz Fc) and reverberation time (0.2- 1.1 sec). An ANOVA revealed significant effects (threshold range: -19 to -35 dB re 60 dB SPL).

0. INTRODUCTION

In rendering an auralized 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 computation for auralization rendering is proportional to the number of early reflections modeled; see, e.g., [1-3] for different approaches. The computational limit can become quickly exhausted [4, 5]; particularly in real-time systems where digital signal processing parameters must be updated in response to a head-tracking device [6, 7]. These engineering constraints motivated the current study which evaluates thresholds for a three-dimensional reverberant field. Auditory threshold data for reverberation can be utilized in conjunction with previously

---

[1] Author for contact regarding additional information (db@eos.arc.nasa.gov).

obtained early reflection threshold data [8, 9] to develop computationally-simplified auralizations of acoustic spaces.

Specifically, we investigated thresholds for speech stimuli in four octave bands that correspond to maximal energy for speech stimuli (0.25, 0.5, 1 and 2 kHz center frequencies). We also investigated the effect of reverberation time, by varying the scalar dimensions of a room modeled with an auralization program, to roughly 75% and 200% of a reference set of dimensions. Hence, the temporal and spatial distribution of reflected energy over time, as well as the particular absorptive characteristics of the modeled room, determine the idiosyncratic aspect of the stimuli. On the other hand, such stimuli are more representative of many real-world modeling situations. Investigation of such stimuli may be particularly relevant when the approach of the modeling software uses frequency domain transformations within octave bands. Octave band data is also useful in developing analysis-synthesis algorithms, which use measured impulse responses to generate virtual synthetic reverberation.

1. EXPERIMENTAL METHOD

1.1 Subjects

Ten paid, volunteer subjects, ages 19-30 (mean = 23) were recruited for the study. They generally finished the study in two separate 2-hr. sessions with multiple rest breaks. Before the study began, a hearing questionnaire and an audiogram were administered. All subjects had hearing thresholds ≤ 30 dB HL and reported no incidence of hearing-related pathology.

1.2 Method

Absolute thresholds were determined at the 50% level within a tolerance of 1 dB using a staircase algorithm [10]. Absolute threshold experiments are advantageous in that no special training is required; the subject can base their responses on any perceived aspect of the stimuli. The threshold was defined for each subject as the mean of 3 staircase direction reversals at the 1 dB level. A three-alternative forced-choice paradigm was used, where the subject identified which of three stimuli heard in succession was different from the other two. Two of these were reference stimuli and one was the probe stimulus, with presentation order randomized. The reference stimulus was the same as the probe, except that the reverberant field was absent. Over the course of a given staircase, the probe stimulus contained a progressively decreasing reverberant field level (with the direct sound level remaining constant), until the threshold for each individual was reached for a given condition (3 room models with associated RTs and 4 octave-band center frequencies). In addition, a full-bandwidth (filter bypass) condition was run under each room condition.

Figure 1 shows all of the staircases for each condition, for an example subject. The figure highlights two staircases: one where the subject remained near the same stimulus level at the minimum dB adjustment step size, and one where the threshold moves upwards before the last reversal. The upward drift is probably due to fatigue. In preliminary testing, three reversals were determined to be the best compromise between fatigue and minimization of the standard deviation between threshold levels. The mean standard deviation was 1.54 dB across all subjects.

1.3 Stimuli

Stimuli consisted of 3-4 sec of spatially-processed speech, one of eight randomly chosen anechoic speech sound files [11]. A commercially available software package (CATT Acoustic) was used to generate binaural impulse responses based on a model of a conference room with an overall medium reverberation time (RT) of 0.5 sec [12]. Two other rooms ("small room" and "large room" with RTs of 0.2 and 1.1 sec respectively) were obtained by scaling the overall absorption or room dimensions. Figure 2 indicates the variations on the room models.

The spatialization of the direct sound used minimum-phase, non-individualized Head-Related Transfer Functions (HRTFs) of a "good localizer." To create binaural impulse responses for the reflected sound, analytical HRTFs based on a solid sphere (i.e., interaural level and time differences but no pinnae cues) were used. It is doubtful that the inclusion of pinnae cues would have affected the results, since the 2 kHz octave band was the highest frequency range tested, and because the pinnae exerts its greatest spectral modification in the > 5 kHz frequency range.

The resulting binaural impulse responses were convolved with test material and then stored in a high-quality stereo sampler (Roland S-760). The direct sound was spatialized with HRTFs described in [4], and overall levels and MIDI triggering were controlled by an *Acoustetron* (Crystal River Engineering) and a PC platform. For octave-band filtering at 0.25, 0.5, 1.0 and 2.0 kHz, the signals were passed through an audio DSP card (Lake CP4), using the "Lake EQ" software package. These filters were linear phase with a roll-off approximating standard settings for octave-band filters [13].

The medium room was based on a model of an empty, windowless conference room at NASA. The RT of the real conference room was measured via a reverse integration of its impulse response [14]; the model agreed within 0.1 sec in the 0.25-2 kHz octave bands. In all room sizes, the source-receiver configuration was held constant; the receiver is facing the source. The reverberation time variable in the experiment was manipulated for the small and large conditions by scaling the surface dimensions and for the small room, doubling the absorption of the surface materials (except for the ceiling, which was held constant). Most surface material absorption coefficients were based on published data, including gypsum board on 90 mm studs [15]; acoustical ceiling tiles [16]; and thin carpet cemented to concrete [17]. Figure 3 indicates the resultant reverberation times and mean percentage absorption for each room. The interaural cross-correlation (IACC) coefficients were .23, .26 and .15 for the small, medium and large rooms, respectively.

The direct sound and filtered reverberant fields were all calibrated to a 60 dB SPL level at the headset, using pink noise; data were adjusted to compensate for the difference between pink noise and the average octave-band spectral energy of the stimuli. Calibration of the stimuli was accomplished by first convolving each of the room models with pink noise, and then adjusting the resulting level for each octave band to be with 0.5 dB of 60 dB SPL (unweighted) at the headphones (Sennheiser HD 545), using an artificial ear and sound level meter (Brüel and Kjær 4153, 2230). To compensate for the difference between the speech stimuli used and pink noise, a frequency analyzer was used (Brüel and Kjær 2123) to obtain the unweighted relative dB level within each octave band (see Table 1). The mean value across the eight source stimuli files were used to adjust the data relative to the 60 dB SPL level, prior to analysis. This was necessary since, as Table 1 shows, the mean spectral content of the speech rolls off gradually with increasing frequency, relative to pink noise.

3. RESULTS AND DISCUSSION

A repeated-measures analysis of variance (ANOVA) revealed a significant effect for RT, $F(2, 18) = 86.0$, $p < .0001$, octave-band center frequency, $F(3, 27) = 36.6$, $p < .0001$, and their interaction, $F(6, 54) = 3.1$, $p < .05$. Overall, the mean thresholds ranged from -19 to -35 dB re 60 dB SPL. Figures 4 and 5 show the thresholds as a function of reverberation time (based on the mean RT for each room) and octave band center frequency.

Figure 6 shows the interaction between room size and octave-band center frequency, along with standard error bars. The distribution of the standard error, indicated by the error bars, suggests that the primary effect was caused by the RT of the large room. The curves for the small and medium room overlap more, as do the standard error distributions. The exception is the data for the 250 Hz octave band, where the difference between each room size appears significant.

Figures 7 and 8 show the thresholds for unfiltered, full-bandwidth (fbw) stimuli, compared to the octave-band results. These data were not included in the main ANOVA since they constituted a different condition. As might be expected, the results are generally somewhere inbetween the results for individual frequency bands. Figure 7 shows the results as a function of the mean value across RTs for individual octave bands. The results for different RTs shown in Figure 8 generally follow the overall effect of the filter bandwidth shown in Figure 7. These results may be helpful in developing non-frequency specific thresholds for reverberant speech.

These data constitute an additional step towards a set of guidelines for data reduction in auralization systems and for analysis-synthesis of virtual reverberation [4, 8, 18]. In general, these data suggest that reverberation will most audible in a comparative sense when modeling larger rooms, and that energy in the higher frequency bands contributes most to the audibility or unmasking of reverberant fields. Thus, greater fidelity of modeled or synthetic reverberation may be a more critical factor for larger rooms and for stimuli with higher frequency content.

Several factors have not been 'teased out' of the stimuli used, including the different temporal and spatial distributions in each of the rooms, and their relative contribution to unmasking of the reverberation. For instance, it may have been that the large room produced a greater release from masking because the reflections arrive relatively later, and less coherently at the listener (IACC = 0.15). Greater sensitivity may exist for music or other broadband stimuli, and perhaps in specific reverberant fields. Such issues will be the focus of future work.

4. ACKNOWLEDGMENTS

Many thanks to Joel Miller, whose software algorithms and discussion were immeasurably helpful; and to Charles M. Salter Associates for the use of their frequency analyzer. This work was supported by NASA-San Jose State University Foundation co-operative grant NC2-237.

5. REFERENCES

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

[2] 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).

[3] Reilly, A. and D. McGrath, "Convolution processing for realistic reverberation". In *AES 98th Convention* preprint 3977 (1995). .

[4] Begault, D.R., "Virtual acoustics: Evaluation of psychoacoustic paramaters". In *Proceedings of theThird International Workshop on Human Interface Technology IWHIT '97, University of Aizu (Japan)* (1997).

[5] Lehnert, H. and J. Blauert, "Principals of Binaural Room Simulation". *Applied Acoustics*, vol. 36, pp. 259-91 (1992).

[6] 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).

[7] Wenzel, E.M., "What peception implies about implementation of interactive virtual acosutic environments". In *Audio Engineering Society 101st Convention* Preprint 4353 (1996).

[8] Begault, D.R., "Audible and inaudible early reflections: thresholds for auralization system design". In *Audio Engineering Society 100th Convention* Preprint 4244 (1996).

[9] 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). .
[10] Levitt, H., "Transformed up-down methods in psychoacoustics". *Journal of the Acoustical Society of America*, vol. 49, pp. 467-477 (1970).
[11] E.B.U. (European Broadcast Union), *Sound quality assessment material recordings for subjective tests (EBU SQAM)*. Hanover, Germany: Polygram (1988).
[12] Dalenbäck, B.-I., *A new model for room acoustic predicition and absorption*. Dissertation, Chalmers University of Technology (Göteborg, Sweden) (1995).
[13] ANSI, *Specification for octave-band and fractional octave-band analog and digital filters*, ANSI S1.11-1986. American National Standards Institute/Acoustical Society of Amercia (1986).
[14] Schroeder, M.R., "New method of measuring reverberation time". *Journal of the Acoustical Society of America*, vol. 37, pp. 409-412 (1965).
[15] Bradley, J.S., "Sound absorption of gypsum board cavity walls". *Journal of the Audio Engineering Society*, vol. 45, pp. 253-259 (1997).
[16] USGC (United States Gypsum Corporation), Acoustone Frost: One-third-octave band absorption coefficient data sheet (measured at Riverbank Laboratories).
[17] Beranek, L., *Concert and opera halls. How they sound.* Woodbury, NY: Acoustical Society of America (1996).
[18] Begault, D.R., "Binaural Auralization and Perceptual Veridicality". In *Audio Engineering Society 93rd Convention* Preprint No. 342 (1992).

Fc of filter used (Hz) ⇒	none	250	500	1000	2000
dB SPL ⇒	60	54.2	55.5	50.7	45.7

TABLE 1. Mean levels of speech stimuli relative to pink noise, used in adjusting the final data.

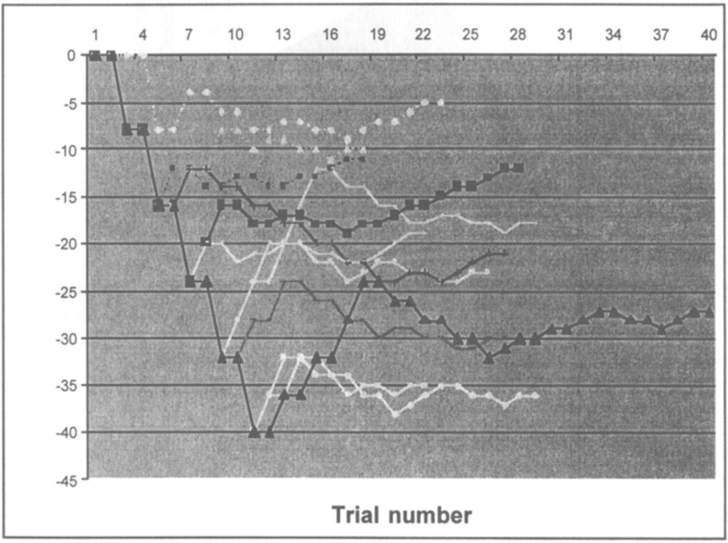

FIGURE 1. Example of stimulus level adjustments made by the Levitt staircase, across all conditions for a sample subject. Staircase begins with 6 dB adjustments in level, decreasing by 50% to 1 dB with each reversal in direction. The threshold is defined as the mean value for three directional reversals at the 1 dB level. Usually, the reversals have a small (> 1.0 dB) standard deviation (e.g., the condition indicated via filled triangles); but, probably due to subject fatigue, the threshold can "drift" upwards, as in the staircase indicated by filled squares.

Room I.D.	L	W	H	vol. (m^3)	α
small	75%	75%	75%	21	75%
medium	(5.2 m)	(3.58 m)	(2.74 m)	50	100%
large	300%	300%	200%	909	100%

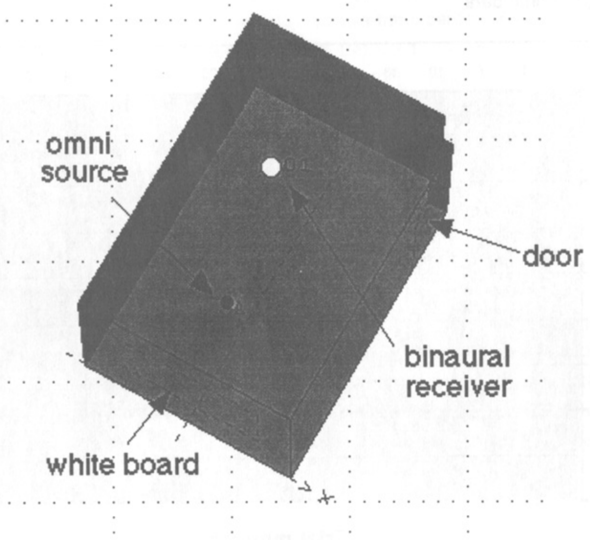

FIGURE 2. Top: variants on room dimension and absorption specifications used (identified in the study as "medium", "large" and "small" rooms) in the room modeling program. Medium room dimensions given in meters; percentages are relative to the medium room. The column labeled α indicates relative absorption. Bottom: view of the medium room, modeled after a conference room at NASA.

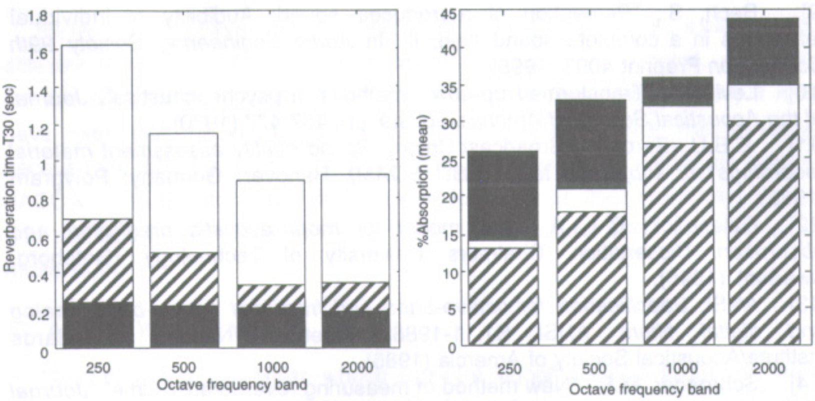

FIGURE 3. Room modeling results for reverberation times (left) and mean percentage of absorbed reflected energy (right), for each octave band used in the experiment. Solid bars: small room; hatched bar: medium room; open bar: large room.

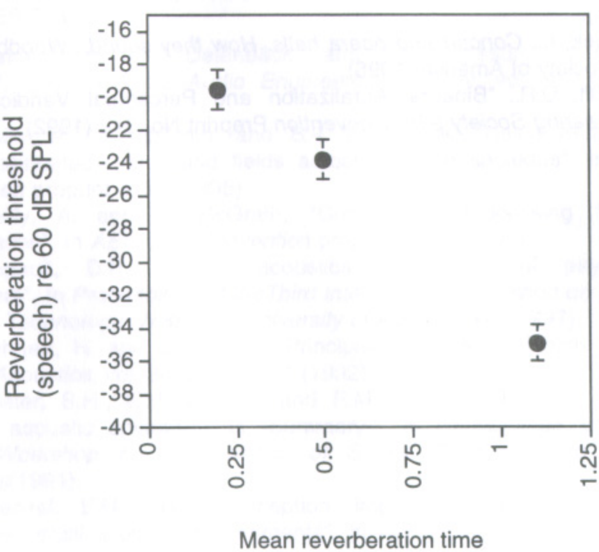

FIGURE 4. Mean and standard error for 10 subjects: thresholds for speech stimuli as a function of mean reverberation time for the small, medium and large room conditions (0.2, 0.5 and 1.1 sec).

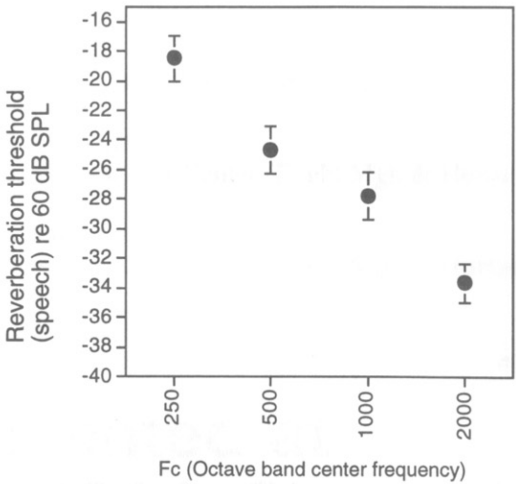

FIGURE 5. Mean and standard error for 10 subjects: thresholds for speech stimuli as a function of octave band center frequency.

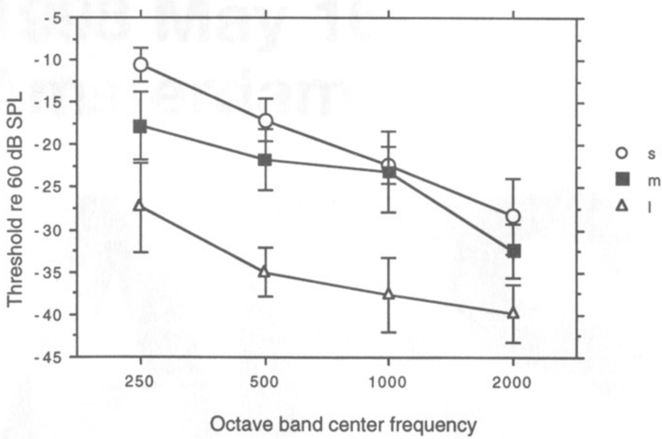

FIGURE 6. Interaction between room size (s = small; m = medium; l = large) RT and octave band center frequency.

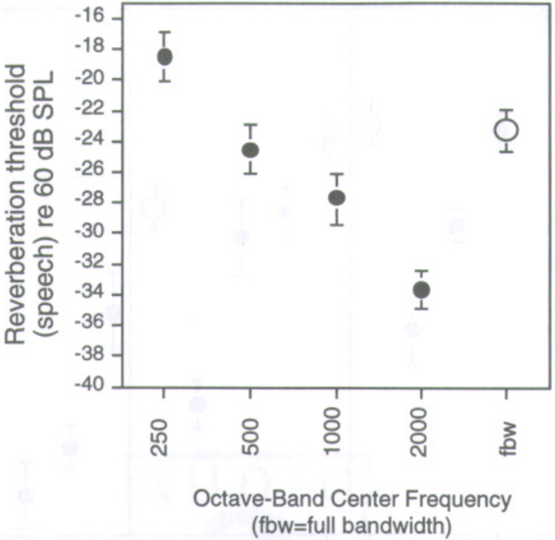

FIGURE 7. Thresholds for full bandwidth stimuli, compared to the mean values for each frequency band tested.

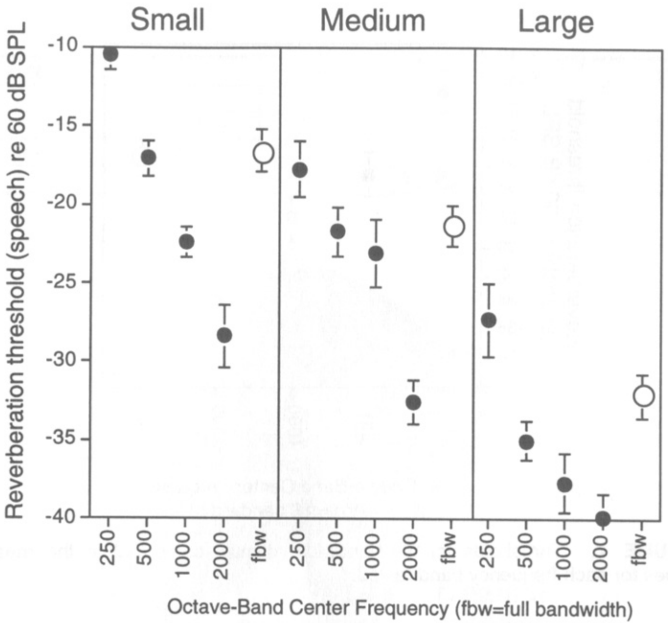

FIGURE 8. Thresholds for full bandwidth stimuli, compared to the mean values for each frequency band tested and within each room size.