



Audio Engineering Society

Convention Express Paper 400

Presented at the AES 159th Convention

2025 October 23-25, Long Beach, CA, USA

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Detection of individual reflections in a binaural presentation of a typical listening room

Matthew Azevedo¹

¹ University of Massachusetts, Lowell, MA, 01854, USA

Correspondence should be addressed to Matthew Azevedo (Matthew_Azevedo@uml.edu)

ABSTRACT

Comparing the acoustic conditions of real rooms is difficult due to the impracticality of moving from room to room quickly enough to maintain an accurate impression of each space. Presenting differing acoustic conditions binaurally allows for easy, rapid comparison but may not provide sufficient detail to allow for accurate comparison between small but perceptually meaningful aspects of room response. In this study, detection thresholds for individual reflections in binaural presentations of measured ambisonic room impulse responses are determined and then compared to published detection thresholds for direct in-room listening to sound reproduced by loudspeakers in a real room.

1 Introduction

The influence of room acoustics on listening is well established. Substantial effort has been made over many decades to analyse the effects of room features on listening conditions and listener preferences in concert halls[1], [2], [3], (and many more) but critical listening spaces for recorded music has not received the same level of attention. It would therefore be useful to establish a reliable, efficient, and inexpensive means of conducting the kinds of listener preference testing for critical listening spaces as has been done at great expense and effort for concert halls. Headphone-based presentations of measured impulse responses would be efficient and inexpensive, but the reliability of such an approach for comparing details of critical listening spaces near the threshold of perception has not been established. The goal of this project is to validate this approach by

determining perceptual thresholds for binaural presentation of measured impulse responses of listening rooms and comparing them to published thresholds for natural listening.

2 Background

The smallest unit of environmental acoustic influence on a perceived sound is an individual reflection. Toole & Olive[4] established perceptual thresholds for individual reflections in typical listening rooms with a loudspeaker source, specifically in a room that meets the IEC criteria for listening rooms[5] and a relatively reflection free (RRF) configuration of the same room with additional sound absorption. Some work has been done on reflection perception thresholds in a binaural presentation[6] but that research used a substantially different methodology using a measured binaural impulse response with and without an individual test reflection simulated in

MATLAB. The resulting threshold of audibility in Jensen & Welte is significantly different than that found by Toole & Olive. However, it is unclear if this discrepancy is due to the difference in methodology or due to shortcomings in a binaural approach in general.

To resolve the question, we are repeating Toole & Olive's experiment as closely as possible while replacing listening to speakers directly in the room under test with listening to a binaural rendering of audio convolved with measured ambisonic impulse responses of the room under test. If a similar perception threshold is found with the binaural presentation as Toole & Olive found for listening directly in the room, it will establish binaural presentation as similar enough to direct listening to be the basis for future work in evaluating subtle acoustic differences critical listening spaces.

3 Experiment

The experimental setup from Toole & Olive was replicated in a recording studio space with variable acoustics, UMass Lowell Moloney Performing Arts Center room 114 (114) with a receiver at the room's preferred listening position and loudspeaker sources on stands 2 m from the source position at 0° and 65° azimuth and 0° vertical. Second order ambisonic impulse responses were captured at the receiver position from both loudspeaker positions with the variable panels of the studio in both their absorptive and reflective positions to emulate the "IEC" (panels reflective) and "RRF" (panels absorptive) conditions in Toole & Olive as shown in Fig. 1. Impulse responses were captured in Cycling '74 Max using the HISSTools impulse response toolbox, which was also used for playback convolution.[7]

Anechoic speech was convolved with the second order impulse responses of both loudspeakers with the 65° azimuth source delayed between 5 and 80 ms to simulate a lateral reflection. The summed ambisonic signals were then decoded to a 12 channel t array[8] using the COMPASS decoder[9] in linear mode and then rendered to head-tracked binaural via convolution with a head related transfer function (HRTF) set derived from the B&K HATS binaural microphone system from the Meta Reality Labs "Sound Sphere 2" data set.[10]

Five subjects were recruited for testing after an IRB review. All five were assigned male at birth and

ranged in age from 20 to 48; none had any known hearing impairments. Each subject listened to the binaural render over headphones and completed 12 versions each of two tasks. In the first task, subjects were asked to adjust the level of the lateral reflection to the lowest level where it was still audible to determine an absolute threshold of perception. In the second task, subjects were asked to adjust the level of the lateral reflection to the lowest level where it produced a perceivable shift in either the apparent location or source width of the 0° source. For each task, evaluations were made with the lateral reflection delayed by 5, 10, 20, 30, 40, and 80 ms in the IEC and RRF room conditions. In both cases, level adjustments were made via on-screen up and down arrows with no feedback provided to the subject on the current absolute level.



Figure 1: Impulse response acquisition setup for the RRF condition. The three fabric areas on the sidewalls were covered for the IEC condition.

4 Results

The average binaural absolute detection thresholds and image shift thresholds for all subjects are presented in Fig. 2 and 3, along with the thresholds for loudspeaker presentation found in Toole & Olive. [4] There is generally strong agreement between the absolute thresholds for speaker and binaural IEC listening room conditions, which is a positive indicator for the use binaural presentation of ambisonic impulse responses for evaluation of such rooms. The binaural RRF condition largely matches the binaural IEC condition, but with thresholds elevated by 3 dB, which is counter intuitive. The binaural image shift thresholds are surprising; on average they are 4 dB lower than the binaural absolute

perception thresholds and 8 dB lower than the thresholds determined by Toole & Olive. Investigating the room impulse responses provides potential explanations for the unexpected results and points towards needed future work.

In Toole & Olive, the IEC condition is described as having a mid-frequency reverberation time of 400 ms, which is very similar to the reflective configuration of 114 with a mid-frequency reverberation time of 410 ms. No numerical data is provided in Olive for their RRF condition but the provided energy time functions (ETF) for IEC and RRF [Fig. 4] show a significant reduction in reflected sound level. In 114, the variable acoustic panels can be opened to eliminate the first-order sidewall reflections at the preferred listening position, but this does not dramatically change the total decay time, reducing it to just 340 ms at middle frequencies. This supports the strong correlation between the detection thresholds of the IEC conditions and the greater similarity of the 114 RRF condition to the Olive IEC than the RRF.

By comparing the energy-time functions (ETF) of the IEC and RRF conditions published in Toole & Olive [Fig. 4] to those measured in 114 [Fig. 5] we see an additional layer of differences. Early reflections in the Toole & Olive IEC room are much stronger than in either configuration of Moloney 114 which exhibit a strong floor reflection at 3 ms followed by a gap with very little energy until the first lateral reflections arrive around 18 ms. The 114 ETFs are much more similar than those in Toole & Olive, with the only significant difference being the amplitude of the first few sidewall reflections. The presence of distinct reflections from the reflection source in the 114 IEC condition in the initial gap of the direct source and the second order reflection from the left side wall visible at 45 ms could account for the unexpectedly lower detection thresholds in the IEC condition than the RRF one.

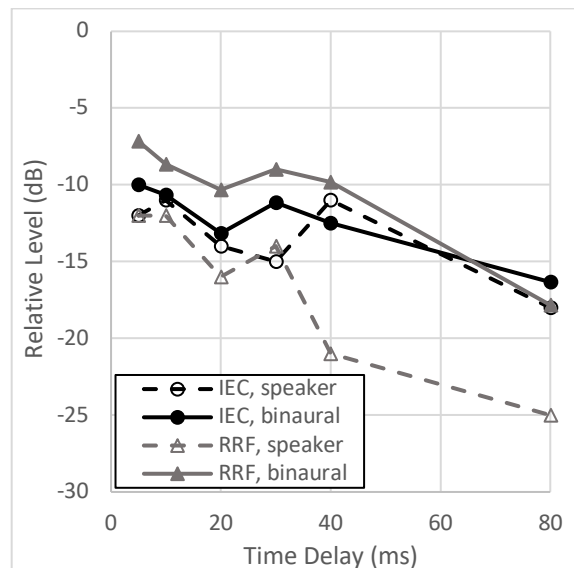


Figure 2: Absolute thresholds for a single lateral reflection (65° H, 0° V) relative to direct sound (0° H, 0° V) presented over speakers [4] and binaurally from a measured impulse response of a similar speaker setup in rooms with strong early reflections (IEC) and without (RRF).

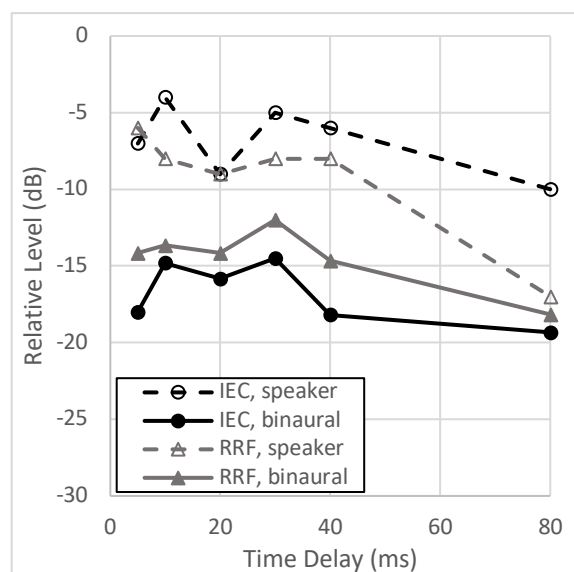


Figure 3: Image shift thresholds from the same experimental setup as in Fig. 2. The image shift threshold indicated the relative level of a lateral reflection to change the perceived location or source width of the direct sound.

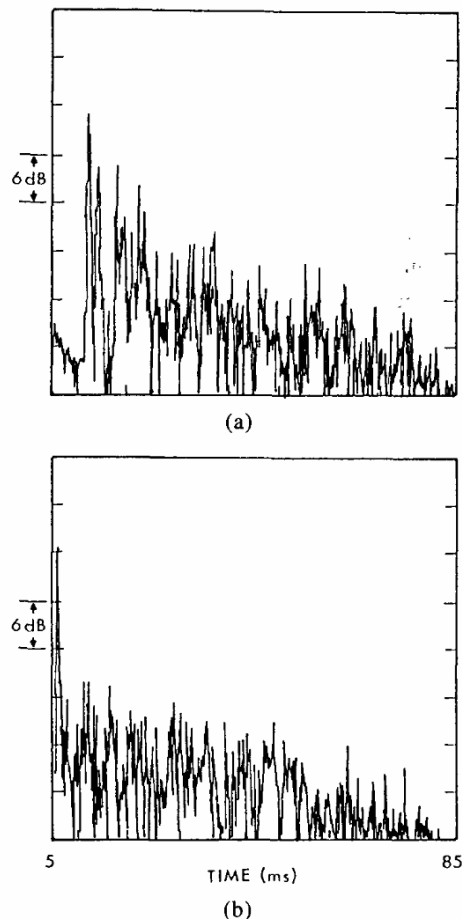


Figure 4: Energy-time functions published by Toole & Olive [4, p.545] of the direct-sound source (0° H, 0° V) in the IEC (a) and RRF (b) configurations of their test room.

Considering the very low image shift thresholds found in this study, a possible difference is the composition of the test subjects. All of this study's subjects were students and faculty from the UMass Lowell Sound Recording Technology program and as such all have training in making detailed observations of apparent source locations. In contrast, Toole & Olive describe this task as “exotic” [4, p.546] for their three subjects.

5 Conclusions

We have found some similarities in the perception thresholds for individual reflections in studio-like rooms between a binaural presentation of measured ambisonic impulse responses and direct listening in the room. However, we have also found that the results are highly sensitive to specific details of the acoustic environment and possibly the training of the

listeners that make a conclusive determination that a binaurally presented ambisonic impulse response is sufficient for comparisons between critical listening spaces.

To make this determination, we will extend this research through another round of testing with determination of single reflection perception thresholds for both speaker simulation and binaural presentation of measured impulse responses in the same room and testing sessions to eliminate all spurious variables.

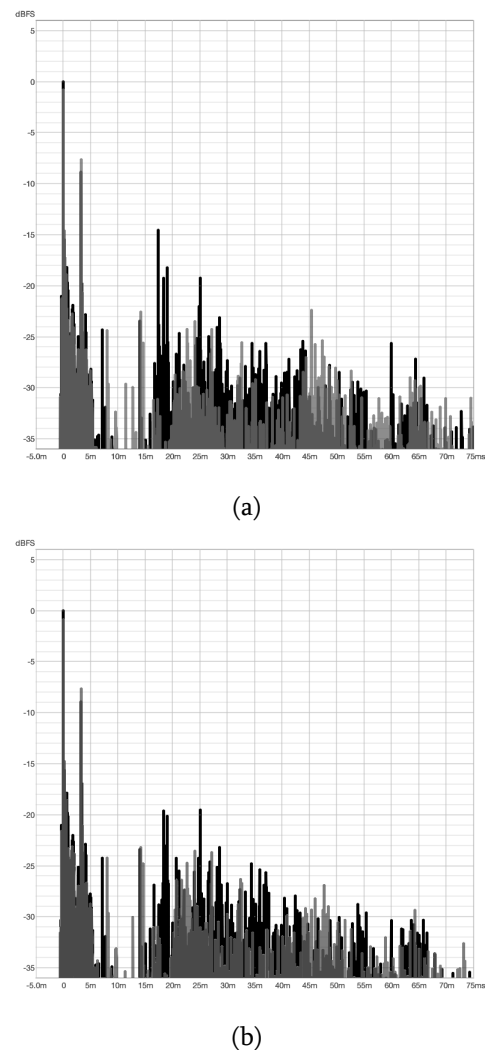


Figure 5: Energy-time functions for Moloney 114 in its IEC (a) and RRF (b) conditions. ETF for the direct sound source (0° H, 0° V) is shown in black, ETF for the reflected sound source (65° H, 0° V) is overlayed in gray.

References

- [1] L. L. Beranek, *Concert halls and opera houses: music, acoustics, and architecture*, vol. 2. Springer, 2004.
- [2] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, “Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters,” *The Journal of the Acoustical Society of America*, vol. 56, no. 4, pp. 1195–1201, Oct. 1974, doi: 10.1121/1.1903408.
- [3] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences,” *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 551–562, July 2016, doi: 10.1121/1.4958686.
- [4] F. E. Toole and S. E. Olive, “The Detection of Reflections in Typical Rooms,” *Journal of the Audio Engineering Society*, vol. 37, no. 7/8, pp. 539–553, Aug. 1989.
- [5] *Sound system equipment - Part 13: Listening tests on loudspeakers*, IEC Publication 60268-13, 1998.
- [6] T. Welti and R. E. Jensen, “The Importance of Reflections in a Binaural Room Impulse Response,” *Journal of the Audio Engineering Society*, vol. 5839, Mar. 2003.
- [7] A. Harker and P. A. Tremblay, “The HISSTools impulse response toolbox: Convolution for the masses,” in *Proceedings of the international computer music conference*, The International Computer Music Association, 2012, pp. 148–155.
- [8] R. H. Hardin and N. J. A. Sloane, “McLaren’s improved snub cube and other new spherical designs in three dimensions,” *Discrete & Computational Geometry*, vol. 15, no. 4, pp. 429–441, Apr. 1996, doi: 10.1007/BF02711518.
- [9] A. Politis, S. Tervo, and V. Pulkki, “COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB: IEEE, Apr. 2018, pp. 6802–6806. doi: 10.1109/ICASSP.2018.8462608.
- [10] M. Warnecke, S. Clapp, Z. Ben-Hur, D. L. Alon, S. V. A. Garí, and P. Calamia, “Sound sphere 2: A high-resolution hrtf database,” in *5th AES International Conference on Audio for Virtual and Augmented Reality, A VAR 2024*, 2024. Accessed: Sept. 23, 2025. [Online]. Available: <https://cris.bgu.ac.il/en/publications/sound-sphere-2-a-high-resolution-hrtf-database>