# Towards determining thresholds for room divergence: A pilot study on perceived externalization

Sebastia V. Amengual Gari
*Facebook Reality Labs Research*
*Facebook Inc.*
Redmond, USA
samengual@fb.com

Henrik G. Hassager*
*GN Audio*
*GN Group*
Copenhagen, Denmark
hghassager@jabra.com

Florian Klein*
*Institute for Media Technology*
*TU Ilmenau*
Ilmenau, Germany
florian.klein@tu-ilmenau.de

Johannes M. Arend*
*Institute of Communications Engineering*
*TH Köln - University of Applied Sciences*
Cologne, Germany
johannes.arend@th-koeln.de

Philip W. Robinson
*Facebook Reality Labs Research*
*Facebook Inc.*
Redmond, USA
philrob22@fb.com

***Abstract*—In binaural rendering, the room divergence effect refers to the decrease on perceived externalization due to the mismatch between the room acoustics of the virtual sounds and those of the listening space. In this work we report on the results of a 2-AFC pilot experiment where 5 expert subjects evaluated the impact of the room divergence effect by comparing real sources and head-tracked virtual sounds generated using the Binaural Spatial Decomposition Method (BSDM) presented over headphones. By applying an exponential weighting function on the measured room impulse responses (RIR) we render binaural RIRs with arbitrary reverberation time (RT), ranging from 50% to 150% of the original RT, while maintaining the temporal and spatial patterns of the original RIR. Preliminary results for a test conducted in a small room (RT $\simeq$ 0.55 s at 1 kHz) suggest that the perceived externalization degree depends on the played stimulus, and progressively degrades with an increasing mismatch. Castanets sounds present externalization ratings comparable to those of a real loudspeaker for RTs ranging from $\sim$90% to $\sim$125%, while for male speech the externalization ratings degrade significantly for sounds with RTs greater than $\sim$110%. Furthermore, we discuss the potential effects of listener adaptation to virtual sounds and its impact on the externalization ratings.**

***Index Terms*—room divergence, binaural, reverberation**

## I. Introduction

Externalization refers to the perception of a sound originating outside of the listener's head. While this is generally the case when listening to real sources, binaurally generated sounds often suffer from a lack of externalization, resulting in what is known as in-head localization. In mixed reality (XR) applications, sources that are externalized and seamlessly integrate with the listener environment aid in enhancing the sense of presence and immersion. However, virtual sounds are often generated using computational models or artificial reverberators that are not able to fully replicate the acoustic properties of the room. Determining acceptable perceptual deviations from the actual listening space is important in this applications in order to ensure that virtual sources are reliably externalized.

Early literature emphasized the importance of rendering "ear-adequate" signals as a requirement to produce externalized sounds [1]. This implies providing all the aural cues that an external source would present, i.e. source, space, and listener properties. From a perceptual perspective, externalization related cues can be grouped in three main categories: acoustic cues of the direct sound, reverberation-related cues, and multimodal factors [2].

With regards to direct sound cues, it is well documented that sources at the median plane tend to be more likely internalized than lateral sources [3], [4], where larger interaural differences are present. The impact of individualized Head-Related Transfer Functions (HRTF) on externalization is not fully understood, and contradicting results are present in the available literature [5]–[8].

Multiple studies have confirmed that reverberant sounds are more likely to be externalized than anechoic sounds [5], [7], [9]. However, it is unclear what specific properties of reverberation lead to higher externalization. In [7] it is reported that the externalization ratings do not improve further when comparing renderings produced using full Binaural Room Impulse Responses (BRIR) as compared to BRIRs truncated after 80 ms. Furthermore, for lateral sources, reverberation at the contralateral ear is critical, likely due to it being the main contributor to the total energy arriving at the ear [10]. It is also known that binaural information in reverberation is necessary to induce externalization [5], [11], although its spectral detail

* Henrik G. Hassager, Florian Klein, and Johannes M. Arend performed the work while at Facebook Reality Labs Research.

is less important than that of the direct sound [12], [13]. However, it is worth noting that externalization gains from reverberation can be limited if the acoustical properties of the listening space and those of the binaural sounds are different, due to a mismatch between expectations and reproduced sounds [1]. This is known as *room divergence effect*, and past studies reported a decrease in externalization in situations of acoustical divergence [8], although training can lead to increased externalization in those scenarios [14]. Nonetheless, the tolerable differences between the acoustics of the listening space and those of the virtual sources to avoid situations of room divergence are currently unknown. The main focus of the current work is to explore perceptual tolerances with regards to reverberation time mismatch of the listening space and the virtual sources presented to listeners.

Among multi-modal cues that significantly impact externalization we find head movements and vision. Head movements contribute to a reduction of front-back confusions [15], [16], more accurate sound localization [15], [16], and higher externalization that persists beyond exposure [4]. It is thus desirable to render binaural virtual sources that allow head movements, presenting world-locked sources. Regarding vision, externalization is reportedly higher in scenarios where the rooms are visible, as opposed to darkness [8] or incongruent visual spaces [8], [17]. However, it is noteworthy that in the presence of visual and auditory feedback, auditory awareness dominates over visual feedback [18].

In this work we present a pilot test that evaluates externalization of 2DOF+[1] binaural sounds generated using the Binaural Spatial Decomposition Method (BSDM) [19], [20] by comparing renderings of measurements taken in the listening room directly with a real loudspeaker at the same position. The renderings are manipulated to modify their reverberation time without affecting their temporal structure. These are compared in a pairwise comparison test to determine what is the acceptable extent of reverberation mismatch before externalization is impaired due to the room divergence effect.

## II. Experiment

### A. Binaural Rendering

Multi-channel Room Impulse Responses (RIR) from a single loudspeaker were obtained at the listening position in a shoe-box room using an open microphone array composed of 7 microphones [21]. These measurements were then analyzed using the Spatial Decomposition Method [19] and Binaural RIRs (BRIR) were generated using the open source Matlab toolbox for binaural SDM rendering with RTMod time-frequency equalization [20]. The method has shown to be capable of generating binaural signals that are perceived as being equally plausible as real loudspeakers in direct comparisons. Note that in the present study we did not include all-pass equalization nor quantization of the spatial information, as opposed to [20]. The multi-channel RIRs were sampled at a sampling frequency $f_s = 48\,\text{kHz}$ and were bandpass filtered between 200 Hz and 8 kHz prior to the directional analysis. The SDM analysis window was 62 samples long. These parameters have shown to minimize the estimation error of the directional information [20].

In order to reduce the memory requirements of the rendering, we truncated the rendered BRIRs after 80 ms and rendered a direction independent reverberation tail. We chose a conservative truncation time to ensure that the direction independent late reverberation did not have an audible effect [22].

Real-time rendering was implemented using Max/MSP as an integrating framework for real-time convolution and signal processing. Objects from the Spat [23] and HISSTools [24] libraries were used to implement the signal processing operations. Anechoic signals were convolved dynamically with the early portion (0 to 80 ms) of the rendered BRIRs, switching filters for various head orientations. The grid resolution of the rendered BRIRs was 1° for azimuth and 5° for elevation. The late reverberation was convolved statically with the anechoic signals, thus resulting in direction independent reverberation.

The Head-Related Impulse Responses (HRIR) used for the binaural rendering corresponded to a KEMAR mannequin and were generated using Boundary Element Method (BEM) simulations. Headphone equalization was performed by convolution with filters generated from measurements on the same mannequin. Frontal equalization was also applied and filters were generated by spectral division of binaural measurements by the re-synthesized BRIRs. In order to allow direct comparison of virtual sounds presented over headphones with real loudspeakers we used non-occluding headphones (AKG K1000) during the test. Further details, along with instrumental and perceptual validation of the method can be found in [20], [21].

Given that during the experiment binaural renderings were compared to real loudspeakers placed at the same position, both the listener and the loudspeaker were tracked using an OptiTrack optical tracking system with 7 cameras. Tracking both source and listener allowed us to present the right direction of arrival of the direct sound at all times by selecting the BRIRs corresponding to the relative source-listener orientation, although only one single point-to-point measurement was used to render the final BRIRs. This allowed listeners to perform small translations, besides head rotations. This was done to avoid the presence of virtual source localization shifts due to translation that could affect the listener judgments during the test. However, note that the synthesized BRIRs were only aligned with the real room at the origin of coordinates. Similar to video renderings allowing small translations, we term this rendering approach 2DoF+. Repeated measurements on the real-time end-to-end pipeline reported a motion-to-sound latency of approximately 60 ms.

[1]In the present context, 2DoF+ refers to 2 degrees-of-freedom (2DoF) rendering with the ability of introducing small translations. The direct sound is rendered in 3DoF, and by tracking both source and listener and rotating the entire presented scene the correct direction of arrival (DOA) of the direct sound is reproduced regardless of the listener position in the room. Direct sound attenuation or amplification once listeners depart from the measured position is not included, thus only small translations are supported. The room information is rendered with 2DoF, corresponding to head yaw and pitch, in order to reduce memory requirements.

### B. Room Conditions

The experiment was conducted in a shoe-box shaped room with minimal furniture present. One single source and listener position was evaluated in this test. The room presents a reverberation time (T30) of approximately 0.55 s at mid frequencies. In order to test the effects of reverberation time mismatch on externalization we generated several versions of the BRIRs with varying reverberation time by multiplying the BRIRs by an exponential function, as detailed in [25]. We generated 7 BRIR variations, each of them with a different percentage of the original T30 - 50%, 75%, 90%, 100%, 110%, 125%, and 150%. The actual T30 of each condition and that of the real room are reported in Fig. 1. Note that slight deviations from the desired reverberation times are present, likely due to uncertainties in the T30 estimation process and due to the binaural nature of the responses, showing slightly different results for left and right ears. However, the averaged results are overall close to the desired T30 values.

To further compare the RT values of the rendered BRIRs against their theoretical value, the ratio between the goal RT and the rendered ones is presented in Fig. 2. It can be observed that there is a slight underestimation of the RT for the left ear and a slight overestimation for the right ear.

### C. Test procedure

The test procedure consisted of a two-alternative forced choice (2AFC) paradigm presenting pairwise comparisons including all the rendered conditions as well as a real loudspeaker in the room. The loudspeaker was hidden behind an acoustically transparent curtain to avoid using localization cues as a discerning element in the judgments (see Fig. 3).

Listeners were asked to report *'Which of the two sounds is better externalized?'*. Note that in each trial they could be asked to compare either two virtual sounds or a real loudspeaker and a virtual sound. They were encouraged to perform natural head rotations and were provided unlimited listening time and were able to switch between stimuli as much as desired. The interaction and test responses were conducted using a touchscreen, minimizing the interaction between the experimenter and the subjects. A screenshot of the test interface is shown in Fig. 4.

Two stimuli were used in the test: castanets and male speech. Castanets was selected due to their impulsive nature, allowing listeners to easily hear the decay properties of the sounds. The male speech was a sequence extracted from the Harvard Sentences. The total number of trials per subject were 112 (28 combinations of conditions $\times$ 2 stimuli $\times$ 2 repetitions).

A total of 5 expert listeners participated in the test. All of the subjects are familiar with binaural rendering, have participated in similar tests in the past, and none of them reported known hearing impairments.

## III. Results

The results of the pairwise test can be arranged as a decision matrix. Each element $a_{i,j}$ of the matrix contains the number of favorable judgments of stimulus $i$ over $j$. Thus, summing up the rows of the matrix results in the total number of times that stimulus $i$ has been favored over each other stimulus, i.e. how many times each stimulus has been chosen as being better externalized than another stimulus. Finally, an *Externalization Score* $E_i$ for each of the conditions can be obtained by normalizing the sum of each row by the total number of presentations of each stimulus

$$E_i = \frac{\sum_{j=1}^{N_s} a_{i,j}}{N_s - 1} \tag{1}$$

where $N_s$ refer to the total number of times that each stimulus was presented. The *Externalization Score* can be obtained for any arbitrary matrix, either for individual or grouped subjects, or for each signal or grouped signals. In the present analysis we run the analysis with independent matrices for each individual and signal, and thus $N_s$ = 14 (7 comparisons per stimulus $\times$ 2 repetitions). A value of $E_i = 1$ would refer to the stimulus $i$ always being rated as being more externalized than the other presented stimulus. Given that $E_i$ is derived from paired comparisons, it is important to note that its absolute value does not necessarily relate to the absolute degree of externalization of a stimulus. For instance, in a case in which all stimuli are equally and fully externalized, random perceptual judgements would result in null relative differences and absolute values well below 1. Thus, relative differences of each condition with regards to a baseline (real loudspeaker) should be analyzed.

The results, separated by stimulus, are presented in Fig. 5 and 6. The idea behind including a real loudspeaker was to obtain the aforementioned baseline judgment of externalization, with the assumption that it would be reliably externalized in all cases.

In the results for the *castanets* stimulus (Fig. 5), it can be observed that the condition of 100% RT (corresponding to the most convergent acoustic conditions) obtains practically the same rating as the real loudspeaker. All BRIRs between 90% and 125% of the original RT are similarly rated, with a slightly lower level. Externalization then degrades significantly and progressively for BRIRs presenting T30 lower than 90% and higher than 125% of those of the real room.

For the case of *male speech* (Fig 6), all the conditions below 110% of the original T30 obtain externalization scores comparable or higher to those of the real loudspeaker, while externalization degrades signifficantly for T30 longer than 110%. Note that in this case, the externalization scores for the real loudspeaker are in some cases lower than those of the virtual sounds. It is known that in situations of room divergence listeners experience adaptation to virtual sounds, resulting in an increase of externalization [14]. However, because of the small sample size in the test it is unclear whether this phenomenon is due to adaptation or a statistical artifact.

Despite the small sample size it seem reasonable to conclude that the divergence effect thresholds for the two tested stimuli

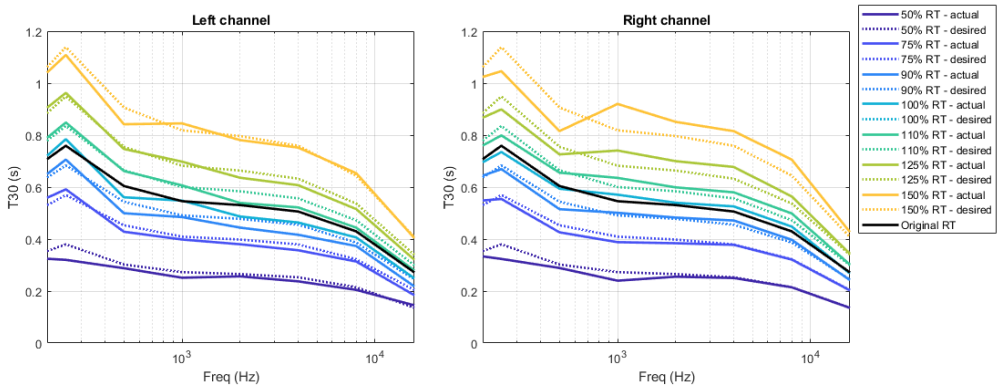


Fig. 1. Reverberation times of the pressure RIR (black), the goal RT (dashed lines) and the RT of the rendered BRIRs (solid lines).

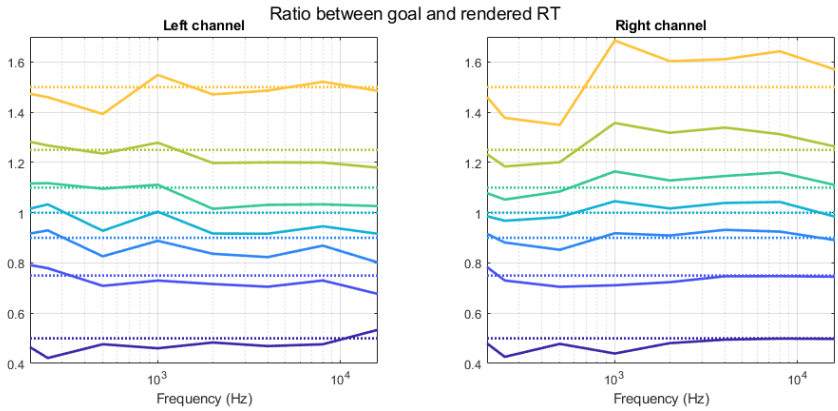


Fig. 2. Ratio between the goal RT and rendered BRIR RT.


Fig. 3. General view of the experiment room.

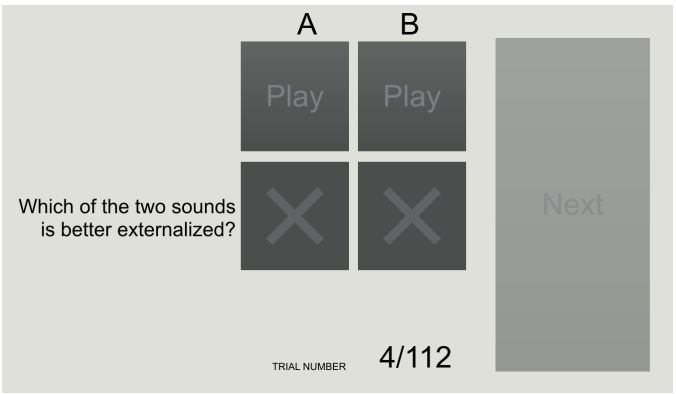


Fig. 4. GUI of the listening test.

are different. The acceptable thresholds for *castanets* trend towards longer tolerable T30, while *male speech* seems to degrade less when T30 shorter than the real room are presented.

## IV. Discussion

### A. Room divergence effect

Up until now, studies investigating the reasons behind the room divergence effect have mostly focused on the investigation of the direct-to-reverberant ratio (DRR) and how its manipulation can enhance perceived externalization [8], [26]. In the present work we have explored the effects of reverberation time mismatch on externalization, as this is one of the fundamental parameters in the characterization of the acoustics of a room. Additionally, by modifying the BRIR envelope, the temporal and spatial properties of the BRIR remain unchanged. This could be regarded as driving a computational model for sound propagation with correct geometry and incorrect materials. This is a plausible scenario in XR applications, as sound

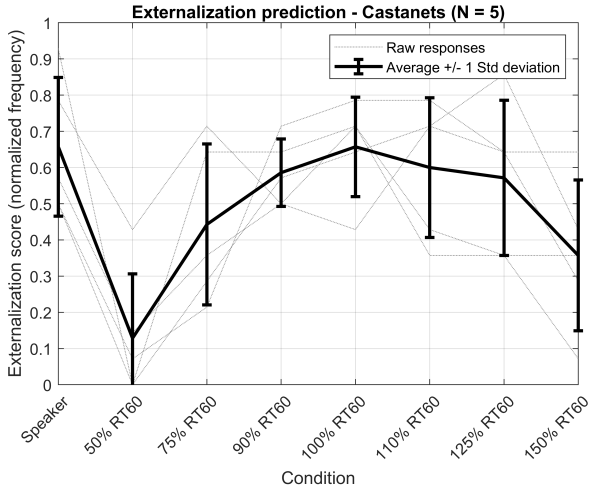


Fig. 5. Externalization Score for the *castanets* stimulus.

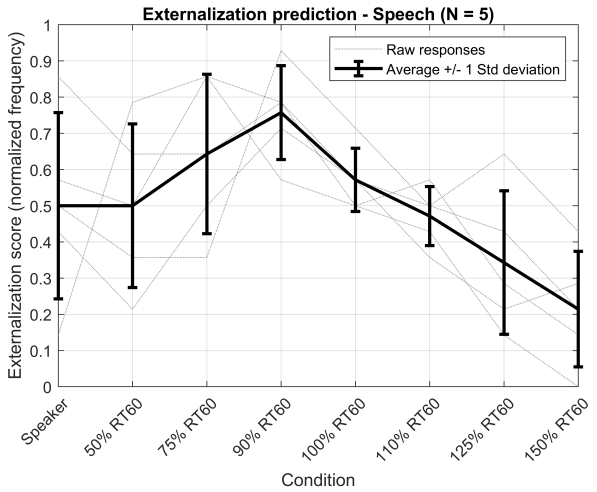


Fig. 6. Externalization Score for the *male speech* stimulus.

propagation models can be driven from 3D reconstruction from computer vision. However, currently the identification of room geometry is more robust than material estimation, potentially resulting in the diverging scenarios evaluated in the present work [27].

Note as well that by manipulating the reverberation time of the BRIRs, other perceptually relevant parameters, such as Center Time, Clarity, or the aforementioned DRR are affected as well. Given the known relationship between distance perception and DRR [28], it is worth considering the possibility of an interaction between perceived externalization and perceived distance for stimuli with shortened T30, which present a higher DRR. As we have discussed, perceived externalization of *Castanets* sounds tend to decrease strongly under reduced T30 coditions, which represent an increase of DRR, and thus could support the assumption of an interaction between T30, DRR, perceived distance, and externalization. However, the opposite is true for sounds with longer T30, which present lower DRR. In this case, although in the present experiment they result in decreased externalization, a lower DRR is generally associated with a larger distance. It is thus desirable to evaluate the interaction between multiple parameters in these situations of acoustical divergence and including a larger number of acoustical environments.

Additionally, a similar procedure could be implemented using BRIRs corresponding to different rooms and modifying the T30 to generate responses matching those listening room. This could help in determining whether the spatio-temporal properties of the BRIRs are relevant or affect the reported T30 thresholds significantly.

### *B. Externalization and plausibility*

During the post-experiment interview, one listener reported that judging externalization in this case was challenging, as most of the percepts were well externalized, regardless of the length of the reverberation. Instead, in some cases their judgment would be based on the plausibility of the percept. In this sense, it might be relevant to experiment with other test paradigms where an absolute scale for externalization is used [5], [8], [18] in order to isolate the judgment from other factors. However, it is worth highlighting the potential challenges of using a continuous scale, as externalization judgments could then morph into distance judgments, and the interrelation of externalization and distance is currently not well understood [2].

### *C. User adaptation*

As reported previously, the *male speech* stimulus presents lower externalization scores for the loudspeaker than for some of the BRIR renderings. It is known that in situations of room divergence, externalization can be enhanced by inducing adaptation effects and shifting listener expectations [14]. Given that the number of trials in which the real loudspeaker was presented was relatively low compared to the total number of trials (28 comparisons over 112 trials) it seems possible that some listeners experienced adaptation and thus shifted their expectations of the actual acoustics of the space. Then, only extreme cases in which audiovisual divergence is obvious (very long reverberation times) would present a strong mismatch with their expectations.

A potential approach to test this hypothesis would be to conduct the same test with the inclusion of explicit references to the natural acoustics of the room. For instance, instructing the user to periodically produce sound, e.g. claps, talking - or reproducing sound from a source in the room and explicitly informing the listener.

### *D. Ground truth and perceived reverberation*

As noted previously, there are some discrepancies between the T30 estimated from the binaural RIRs diverge slightly from the target values. In this study the method used for the estimation was compliant with the ISO 3382, and yet differences exceeding the Just Noticeable Differences (JND) of 5% can be observed. This introduces an uncertainty in

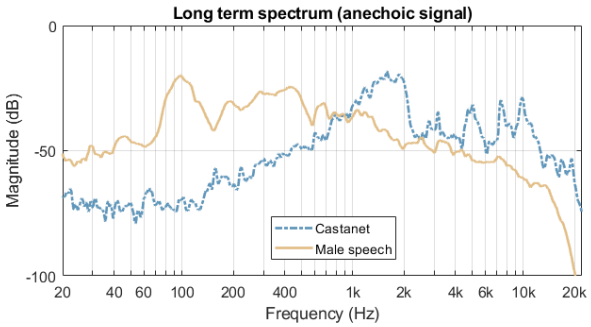


Fig. 7. Smoothed (1/12 octave) long term spectra of the anechoic signals.

the generation of BRIRs that can influence the results of the renderings.

In addition, although the reverberation times are scaled independent of frequency, the absolute T30 difference between renderings is frequency dependent. Thus, T30 differences between stimuli at low and mid frequencies are generally larger than at high frequencies. The fundamental frequency of the stimulus *speech* is around 100 Hz, and most of the energy is below 1 kHz, while *castanets* has most of its energy at mid frequencies, with strong harmonics at higher frequencies (see Fig. 7). Since the reverberation time is frequency dependent, we could expect that the perceived reverberation differences for the same percentage difference are in fact different for each stimuli, depending on their frequency content. In addition, the castanets stimulus is of impulsive nature, which could reveal further differences.

Although from the small sample size in this experiment it is not straightforward to draw strong conclusions, it is reasonable to assume that both spectral and temporal characteristics of the stimuli have an impact on audibility of reverberation changes and perceived externalization. We explored the topic of audibility of reverberation changes in another contribution to this conference [29].

## V. Conclusions

In this work we presented a perceptual test exploring the effects of reverberation mismatch between virtualized sounds and the real space on perceived externalization. The test was conducted using 2DoF+ binaural renderings generated using BSDM and generic HRIRs.

We concluded that for the *castanets* stimuli, externalization degrades significantly when the T30 of the renderings is outside of the range 90% to 125% of that of the real room. For *male speech*, externalization degrades significantly for sounds with reverberation greater than 110% of that of the real room.

We discussed the potential implications of listener adaptation during the test and alternatives to investigate this phenomenon. Additionally, we discussed potential explanations for the differences in perceptual thresholds for the evaluated stimuli.

Further work includes collecting data with a larger pool of subjects, as well as expanding the number of evaluated stimuli, source locations, and room conditions.

## References


[1] G. Plenge, "On the differences between localization and lateralization," *The Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 944–951, 1974. [Online]. Available: https://doi.org/10.1121/1.1903353

[2] V. Best, R. Baumgartner, M. Lavandier, P. Majdak, and N. Kopčo, "Sound externalization: A review of recent research," *Trends in Hearing*, vol. 24, p. 2331216520948390, 2020, pMID: 32914708. [Online]. Available: https://doi.org/10.1177/2331216520948390

[3] D. R. Begault and E. M. Wenzel, "Headphone localization of speech," *Human Factors*, vol. 35, no. 2, pp. 361–376, 1993, pMID: 8349292. [Online]. Available: https://doi.org/10.1177/001872089303500210

[4] W. O. Brimijoin, A. W. Boyd, and M. A. Akeroyd, "The contribution of head movement to the externalization and internalization of sounds," *PLOS ONE*, vol. 8, no. 12, p. null, 12 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0083068

[5] T. Leclère, M. Lavandier, and F. Perrin, "On the externalization of sound sources with headphones without reference to a real source," *The Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2309–2320, 2019. [Online]. Available: https://doi.org/10.1121/1.5128325

[6] S.-M. Kim and W. Choi, "On the externalization of virtual sound images in headphone reproduction: A wiener filter approach," *The Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3657–3665, 2005. [Online]. Available: https://doi.org/10.1121/1.1921548

[7] D. R. Begault, A. S. Lee, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, October 2001.

[8] S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg, "A summary on acoustic room divergence and its effect on externalization of auditory events," in *Eighth International Conference on Quality of Multimedia Experience (QoMEX 2016)*, June 2016, pp. 1–6.

[9] J. M. Kates, K. H. Arehart, R. K. Muralimanohar, and K. Sommerfeldt, "Externalization of remote microphone signals using a structural binaural model of the head and pinna," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2666–2677, 2018. [Online]. Available: https://doi.org/10.1121/1.5032326

[10] S. Li, R. Schlieper, and J. Peissig, "The effect of variation of reverberation parameters in contralateral versus ipsilateral ear signals on perceived externalization of a lateral sound source in a listening room," *The Journal of the Acoustical Society of America*, vol. 144, no. 2, pp. 966–980, 2018. [Online]. Available: https://doi.org/10.1121/1.5051632

[11] J. Catic, S. Santurette, and T. Dau, "The role of reverberation-related binaural cues in the externalization of speech," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 1154–1167, 2015. [Online]. Available: https://doi.org/10.1121/1.4928132

[12] H. G. Hassager, F. Gran, and T. Dau, "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2992–3000, 2016. [Online]. Available: https://doi.org/10.1121/1.4950847

[13] Z. Jiang, J. Sang, C. Zheng, and X. Li, "The effect of pinna filtering in binaural transfer functions on externalization in a reverberant environment," *Applied Acoustics*, vol. 164, p. 107257, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X19309223

[14] F. Klein, S. Werner, and T. Mayenfels, "Influences of training on externalization of binaural synthesis in situations of room divergence," *J. Audio Eng. Soc*, vol. 65, no. 3, pp. 178–187, March 2017.

[15] F. L. Wightman and D. J. Kistler, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2841–2853, 1999. [Online]. Available: https://doi.org/10.1121/1.426899

[16] W. O. Brimijoin and M. A. Akeroyd, "The role of head movements and signal spectrum in an auditory front/back illusion," *i-Perception*, vol. 3, no. 3, pp. 179–182, 2012, pMID: 23145279. [Online]. Available: https://doi.org/10.1068/i7173sas

[17] J. Udesen, T. Piechowiak, and F. Gran, "The effect of bision on psychoacoustic testing with headphone-based virtual sound," *Journal of the Audio Engineering Society*, vol. 63, no. 7/8, pp. 552–561, july 2015.

[18] J. C. Gil-Carvajal, J. Cubick, S. Santurette, and T. Dau, "Spatial hearing with incongruent visual or auditory room cues," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[19] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc*, vol. 61, no. 1/2, pp. 17–28, March 2013.

[20] S. V. Amengual Garí, J. M. Arend, P. T. Calamia, and P. W. Robinson, "Optimizations of the spatial decomposition method for binaural reproduction," *Journal of the Audio Engineering Society*, vol. 68, no. 12, pp. 959–976, 2021.

[21] S. V. Amengual Garí, W. O. Brimijoin, H. G. Hassager, and P. W. Robinson, "Flexible binaural resynthesis of room impulse responses for augmented reality research," in *EAA Spatial Audio Signal Processing Symposium*, Paris, France, Sep. 2019, pp. 161–166.

[22] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses," *J. Audio Eng. Soc*, vol. 60, no. 11, pp. 887–898, Nov. 2012.

[23] T. Carpentier, "A new implementation of Spat in Max," in *15th Sound and Music Computing Conference (SMC2018)*, Limassol, Cyprus, Jul. 2018, pp. 184 – 191. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02094499

[24] A. Harker and P. A. Tremblay, "The HISSTools Impulse Response Toolbox: Convolution for the Masses," in *ICMC 2012: Non-cochlear Sound*, M. Marolt, M. Kaltenbrunner, and M. Ciglar, Eds. The International Computer Music Association, July 2012, pp. 148–155. [Online]. Available: http://eprints.hud.ac.uk/id/eprint/14897/

[25] D. Cabrera, D. Lee, M. Yadav, and W. L. Martens, "Decay envelope manipulation of room impulse responses: Techniques for auralization and sonification," in *Proceedings of Acoustics 2011*, Gold Coast, Australia, 2011.

[26] T. Sporer, S. Werner, and F. Klein, "Adjustment of the direct-to-reverberant-energy-ratio to reach externalization within a binaural synthesis system," in *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2016.

[27] H. Kim, L. Remaggi, P. J. B. Jackson, and A. Hilton, *Immersive Virtual Reality Audio Rendering Adapted to the Listener and the Room*. Cham: Springer International Publishing, 2020, pp. 293–318. [Online]. Available: https://doi.org/10.1007/978-3-030-41816-8_13

[28] A. J. Kolarik, B. C. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, & Psychophysics*, vol. 78, no. 2, pp. 373–395, 2016.

[29] F. Klein, S. V. Amengual Gari, J. M. Arend, and P. Robinson, "Towards determining thresholds for room divergence: A pilot study on detection thresholds," in *to be published in Proceedings of International Conference on Immersive and 3D Audio*, Italy, September 2021.