



Audio Engineering Society

Conference Paper 6

Presented at the 6th International Conference on Audio for Games
2024 April 27–29, Tokyo, Japan

This paper was peer-reviewed as a complete manuscript for presentation at this conference. 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.

Perceptual comparison of efficient real-time geometrical acoustics engines in Virtual Reality

Sebastia V. Amengual Garí¹, Carl Schissler¹, and Philip Robinson¹

¹Reality Labs Research, Meta

Correspondence should be addressed to Sebastia V. Amengual Garí (samengual@meta.com)

ABSTRACT

Interactive immersive experiences and games require the dynamic modelling of acoustical phenomena over large and complex geometrical environments. However, the emergence of mobile Virtual Reality (VR) platforms and the ever limited computational budget for audio processing imposes severe constraints on the simulation process. With this in mind, efficient geometrical acoustics (GA) real-time engines are an attractive alternative. In this work we present the results of a perceptual comparison between three geometrical acoustic engines suitable for VR environments: an engine based on an Image Source Model (ISM) of a shoebox of variable dimensions, a path tracing (PT) engine with arbitrary geometry and frequency dependent materials, and a bi-directional path tracing (BDPT) engine with perceptual optimization of the Head-Related Transfer Function. The tests were conducted using Meta Quest and Quest 2 headsets and 26 listeners provided perceptual ratings of six attributes (preference, realism/naturalness, reverb quality, localization, distance, spatial impression) of three different sources in 6 scenes. The results reveal that the BDPT engine is consistently rated higher than the other two in 4 of the perceptual attributes i.e. preference, realism/naturalness, reverberation quality, and spatial impression, particularly in large reverberant spaces. In small spaces, trends are less clear and ratings are more subject dependent. A Principal Component Analysis (PCA) revealed that only two perceptual dimensions account for more than 80% of the explained variance of the ratings.

1 Introduction

The user perceived importance of audio in video games has been historically low [1], with video quality historically being prioritized. However, an increasing body of recent anecdotal evidence and formal studies reveal that spatial audio in games has significantly positive effects in the experience and perceived player value. For instance, in First-Person Shooter (FPS) games, consistent audio cues provide a competitive advantage to experimented players [2], and the use of head tracking can fur-

ther improve performance in multiple video game genres [3]. The increase in immersion provided by audio is greater in VR than in monitor-based games [4]. Additionally, the increase in perceived immersion in virtual environments when including head-tracking and room acoustic rendering against monaural audio is comparable to a five fold increase of the video resolution [5].

Acoustical realism might not be indispensable to reap the benefits of spatial audio in immersive interactive experiences, however acoustical consistency is key in

aiding users in building navigable mental maps [6]. In turn, efficient real-time geometrical acoustics (GA) simulations allow listeners to navigate unknown complex environments in Virtual Reality (VR) without the need for training or learning those mental maps [7]. In virtual outdoor spaces, which are common in video games, wave based simulations also provide benefits over GA in the task of locating an active acoustic source [8].

All of this suggests that audio for immersive environments plays a critical role in video games, and real-time engines have evolved vastly in the last few decades. Traditional approaches consisted of the manual artistic design of reverb zones via reverberators and the imitation of acoustic phenomena such as occlusion, transmission, or air absorption with the use of parametric filters [9, 10]. However, this involves an amount of effort that can result in unpractical situations as the size of environments in games and interactive productions continues to grow. In this context, an increasing number of real-time audio propagation engines have emerged in the last decade or so, including both research/experimental engines [11, 12, 13, 8, 14] and commercial engines [15, 16, 17, 18, 19]. These engines aim at providing physically inspired sound propagation at run time while at the same time providing perceptually satisfactory results. By making use of scene geometry and assigning acoustic properties to materials, they are easily scalable to simulate large virtual environments and pose an attractive solution.

While perceptual evaluation of room acoustics and spatial audio in general is a very active topic, relatively little is known about the actual implementation requirements of real-time engines in ecologically valid and immersive settings. In this paper we expand a previous experiment [20] which evaluated two real-time engines in six VR scenes. In the present experiment we add a third engine to the experiment and investigate the perceptual dimensions that govern subject ratings. The paper is organized as follows: Section 2 presents a summary of the evaluated engines, Section 3 describes the scenes and procedure followed in the experiment, Section 4 presents the results of the study and evaluates the underlying relevant perceptual dimensions as well as the room and content dependency of the ratings, Section 5 discusses the implications of the results as well as potential design guidelines for acoustic VR simulations, and finally Section 6 summarizes the conclusions of the study.

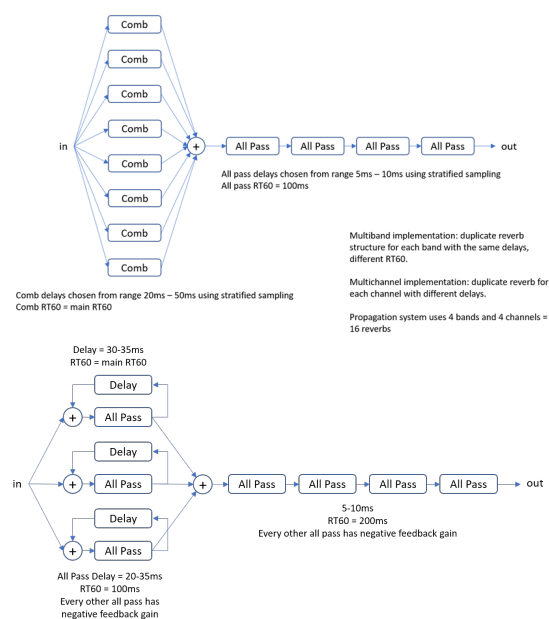


Fig. 1: Artificial reverberators used in PT (top) and BDPT (bottom). An independent instance of the reverberator is driven in each frequency band to generate frequency dependent results.

2 Engines

In this work we evaluate three engines: an Image Source Model (ISM) of a shoebox room of adjustable dimensions; a Path Tracing (PT) engine with reverberation rendering through an artificial reverberator; and a Bi-Directional Path Tracing (BDPT) engine based on the PT engine with improved Head-Related Transfer Function (HRTF) encoding using Magnitude Least Squares (MagLS) [21]. The initial HRTF in all cases is the same, based on a subject from the CIPIC database [22], although the processing is different in each engine. In this work all three engines were implemented as Unity plug-ins. In the following we expand on the characteristics of each engine. Additionally, we provide a summary table comparing the three engines in Tab. 1.

2.1 Image Source Model (ISM) Engine

The engine features an Image Source Model (ISM) [23] of a shoebox room for the early reflections in which the listener is centered in the middle of the simulated room, and directions of arrival of reflections are updated according to head rotations of the listener. The

Simulation aspect	ISM	PT	BDPT
HRTF	CIPIC subject 48 with tailored equalization. ITDs are extracted before SH conversion and reinserted later.	CIPIC subject 48 with equalization. ITDs are not extracted prior to SH conversion, resulting in a noticeable low-pass coloration.	CIPIC subject 48 with equalization. MagLS encoding into SH
Early reflections	Shoobox model with listener always fixed at the center	Raytracing with high diffusivity. The early reflections are not prominent.	Raytracing with high diffusivity and dedicated early reflections.
Late reverb	Sampled Room Impulse Responses (RIR). Static reverberation.	Raytracing, fully dynamic, Comb filter + series of All pass	Raytracing, fully dynamic, with BDPT+MIS, nested all pass + serie of all pass
Material properties	Broadband, one material per wall	Fully customizable in terms of geometry, limited to 4 frequency bands	Fully customizable in terms of geometry, limited to 4 frequency bands
Air absorption	No	Yes	Yes
Scattering	No	Yes, fully customizable in terms of geometry, limited to 4 frequency bands	Yes, fully customizable in terms of geometry, limited to 4 frequency bands
Occlusion	No	Yes	Yes

Table 1: Comparison of simulation aspects for each engine.

late reverberation is simulated via static sampled Room Impulse Response (RIR) convolution and faded out after 600 ms to restrict computational demands. In spite of the fade out, the decay during the first 600 ms of the RIR corresponds to the appropriate rate of decay for the specified Reverberation Time (RT). Several parameters are configurable, such as broadband reflection coefficients, room size, or reverb gain, among others. Occlusion, diffraction, or air absorption are not modeled. The processing is done in the Spherical Harmonics Domain (SHD) and downmixed to binaural, by convolving the simulated sound field and the HRTF dataset in the SHD. An instrumental evaluation of the engine is provided in [20] and is out of the scope of this paper.

2.2 Path Tracing (PT) Engine

The Path Tracing (PT) Engine features a full dynamic path-tracing simulation in 4 bands. The cutoff frequencies are logarithmically spaced frequency bands between 40 Hz and 15 kHz and the lowest and highest bands are modified to convert them to low and high-pass bands, respectively (0 Hz to 176 Hz, 176 Hz to 775 Hz, 775 Hz to 3408 Hz, and 3408 Hz to Nyquist frequency). The engine makes use of the game geometry and acoustic material parameters defined in the 4 mentioned bands to simulate absorption, scattering and transmission. The simulation is used to obtain time-energy profiles that are then used to drive a series of Schroeder reverberators [24] (see 1) to generate direction and frequency dependent reverberation. Note that an independent reverberator is needed for each

band, and thus the choice of 4 bands is a compromise between frequency resolution and computational cost. The outputs of the reverberators are processed in the SHD and, similarly to the ISM engine, the final binaural signals are obtained by convolving the simulated sound field and the HRTF in the SHD. Further details and an instrumental validation can be found in [7].

2.3 BDPT

The Bi-Directional Path Tracing (BDPT) engine is based on the PT engine, although it features several algorithmic improvements. The Bi-directional Path Tracing simulation [25] traces paths from both the source and the receiver and includes Multiple Importance Sampling (MIS) resulting in less noisy energy profiles and more stable simulations, increasing the robustness in edge cases. The reverberators in this engine are re-designed to feature a set of nested all-pass filters in parallel configuration followed by a cascade of all-pass filters, reducing the computational cost compared to the PT reverberator. The PT engine was found to deviate from a reference simulation [20] and the BDPT engine improves the results, especially on the early to late energy ratios of the generated RIRs. This engine also performs the rendering operations in the SH domain and downmixes to binaural by performing a convolution of the sound field and HRTF dataset in the SHD. However, in this case, the HRTFs have been encoded into the SHD by using MagLS, which improves the magnitude response of the encoded HRTFs


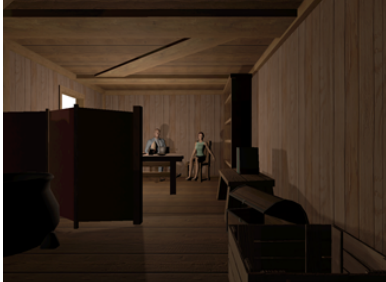

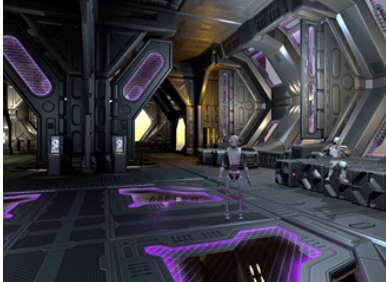

Living Room - T30 = 0.4 s	Cabin - T30 = 0.7 s	Lecture Room - T30 1.55 s
		
Shoe-box scene with plaster walls and ceiling. Big carpet on the floor and absorptive furniture.	Small wooden house with low absorption furniture. A highly absorptive room divider is placed between two of the sources and the listener, dampening potential early reflections from one side. The room has a second story and irregular ceiling height.	Large lecture room with wooden front wall and floor, plaster ceiling and curved back wall. Considerable amount of absorption at the audience area.
Warehouse - T30 = 1.7 s	Space Ship - T30 = 2.5 s	Church - T30 = 4.9 s
		
Large warehouse with many coupled rooms, shelves made out of steel, and concrete floor. The shelves act as coupled sub-rooms providing many late reverberation paths. The effective volume of the room is considerably reduced by the presence of the shelves.	Long room made out of steel with an opening at one end and inclined lateral walls.	Gothic church with cross shaped floor plan and a rectangular main room. The only furniture present are wooden benches. The material used to model the walls is "concrete rough".

Table 2: Description of the rooms featured in the listening test.

by disregarding phase information above a given cutoff frequency [21].

This engine has been recently used to generate real-time acoustic environments multiple Deep Learning tasks [26, 27, 28]. We refer the reader to [29] for further information on the instrumental evaluation of the engine.

3 Experiment

3.1 Evaluated scenes

The test was conducted in 6 scenes, the same rooms used in the study from [20], which aims at providing a wide range of variety in terms of room properties

and acoustics. The details of the rooms are included in Tab. 2.

In order to minimize confounding factors in the experiment, it is important to adjust the simulation parameters the engines to produce comparable RIRs in each scene. Following the procedure describe in [20], we first designed the rooms using the PT engine and iteratively modified the parameters of the ISM engine (shoebox size and absorption parameters) to match the room acoustical parameters of each scene i.e. Energy Decay Curve (EDC), Reverberation Time (via T20 estimation), Early Decay Time (EDT), and Clarity (C50) within approximately ± 2 JND for frequencies between 250 Hz and 4000 Hz. In the case of BDPT, since its underlying system and input data is based on PT, we used the same input data (geometry and material information).



Fig. 2: GUI of the listening test, represented as a hand-held virtual tablet.

3.2 Protocol

The test consisted of pairwise comparisons conducted in VR, where sounds generated by 2 engines were compared in each trial. Experiments were conducted in PC VR using a Meta Quest or Quest 2 and link cable, depending on availability. The listeners were asked to evaluate the engines using 6 perceptual attributes. These were selected based on their prevalent usage in a previous test with similar characteristics [20]. The definitions of each attribute were provided to the subjects and discussed with them before the experiment:

- **Overall Preference:** Your subjective preference. Might be based on any attribute, a combination of those asked later, or any other reason.
- **Realism/Naturalness:** The rendered sounds resembles the real expected sound of the presented room better and it sounds more natural than the other one.
- **Reverberation quality:** Overall perceived quality of the reverberation.
- **Localization:** The selected sound is localized closer to the true source position e.g. hearing that speech comes out of the mouth of a person.
- **Distance:** The distance of the presented sounds is closer to the visual distance.
- **Spatial impression:** Spatial properties of the renderer and their fit to the spatial visual properties of the room (presence of reflections, envelopment of the sound, direction of echoes...).

During each trial, listeners were provided with unlimited time to switch back and forth between the two engines by holding or releasing the trigger button on either controller. As all engines rendered sound in 6 degrees-of-freedom (6 DoF), participants were encouraged to freely rotate their heads and slightly translate around their listening position while trying to not penetrate any solid objects with their head. The ratings for each of the attributes were collected using a virtual hand held tablet and continuous sliders (see Fig. 2).

Only one source was active in each trial, with content of female speech, male speech, and solo trumpet music. All of the sources were represented by static avatars placed at different positions and were visible to the participants. This resulted in 54 trials without repetitions (3 comparisons x 3 sources x 6 scenes), which were presented randomly. A total of 31 subjects participated in the study. The self-reported gender distribution was 16.13% female and 83.87% male. The age distribution was 6.45% (under 25), 32.26% (26-34), 32.26% (35-44), 29.03% (45-54). None of the subjects reported known hearing impairments.

The final number of subjects included in the analysis was N=26. We decided to discard the data from a subset of participants based on incomplete datasets or technical problems reported during the conduction of the experiment. The experiments were conducted remotely, by distributing an executable build of the Unity project to the participants and a guide to ensure that the setup was uniform among them. The experimenters conducted a video call with each participant to explain the procedure and troubleshoot any potential issue. At the end of each session, participants were asked to complete a survey to document demographic information, equipment, and any potential issue encountered during the experiment session. All of the participants were highly familiar with VR hardware and immersive experiences.

The test levels were calibrated to resemble human speech levels using a Razer Blade FHD 15 (laptop) and Beyerdynamic DT990 headphones. However, differences in presentation level due to differing hardware and the assessment of background noise were not possible to control and participants were asked to adjust the reproduction level to a comfortable level. Table 3 contains a list of the headphone models used in the test.

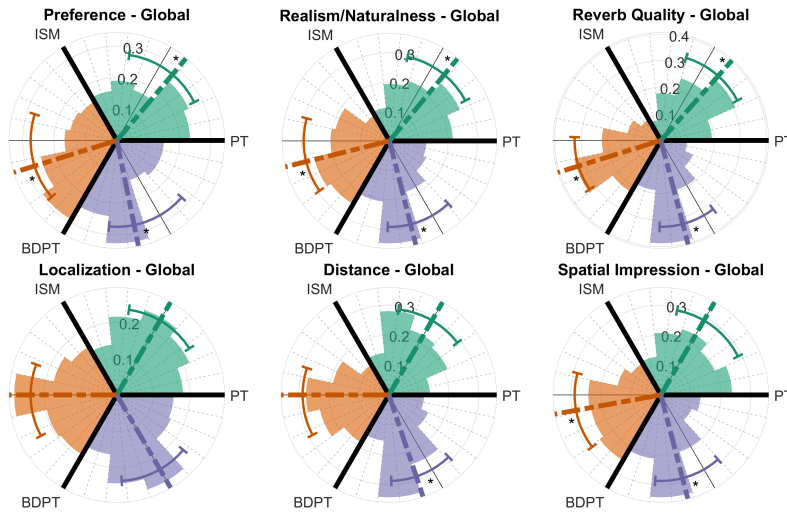


Fig. 3: Histograms of the ratings for the 6 evaluated attributes. Each colored sector refers to the comparison between two engines. Dashed colored lines represent the median rating of each engine comparison and colored error bars within each sector represent interquartile ranges of the ratings. The values of the radius represent the empirical probability of the histogram bins. Two-sided sign tests were performed to test the hypothesis of the data coming from a distribution with zero median. Statistical significance of the rejection of the null hypothesis is denoted by * and corresponds to $p < 0.01$.

Headphone model	# participants
Beyerdynamic DT770 Pro	5
Beyerdynamic DT990 Pro	4
Bose Quiet Comfort 35 II	3
Sony MDR-7506	2
Audio Technica ATH-M40fs	2
Sony WH-1000XM4	2
Others (over the ear)	4
Others (earbuds)	4

Table 3: Headphone models used in the test ($N = 26$).

3.3 Pilot test

In order to validate the setup and minimize potential disruptions in an uncontrolled environment, an onsite pilot test was conducted before the final experiment with $N = 7$ participants. The procedure, scenes, and rendering used for the pilot test were exactly the same as used for the final test. The results, albeit more noisy than those of the final test due to the smaller number of participants presented similar trends.

4 Results

In Fig. 3 we present histograms which show the grouped results including all sources and all rooms

for each perceptual attribute. Ratings favoring BDPT over both PT and ISM are apparent over most of the perceptual attributes except for localization, which seems to be neutral in all cases. Additionally, PT seems to be generally over ISM for the attributes of preference, realism/naturalness, and reverb quality, and neutral for localization, distance, and spatial impression.

4.1 Perceptual dimensions

To simplify the analysis of the data and interpretation of the results, we explored correlations between the ratings of each parameter. For this, we treated each of the comparisons as an independent dataset (ISM vs PT, ISM vs BDPT, PT vs BDPT) and conducted a Principal Component Analysis (PCA) analysis on each of the 3 datasets in order to construct perceptually relevant dimensions.

The PCA analysis reveals that only 2 perceptual dimensions can explain up to 85% of the variance seen in the perceptual results. In all 3 cases, the first dimension is composed of a roughly equal contribution of all parameters, with a slightly smaller contribution from Distance, Localization and Reverb Quality. The second dimension is dominated by Distance and Localization.

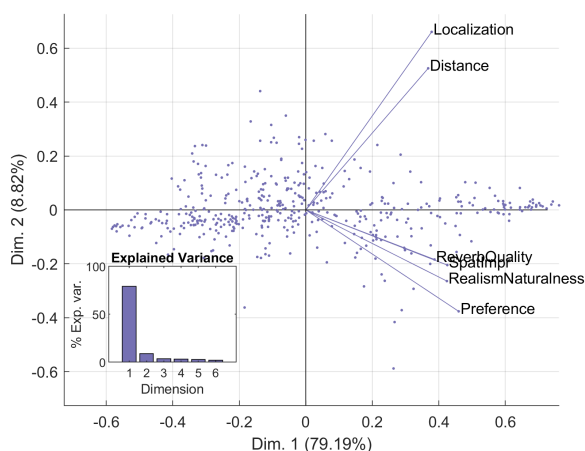


Fig. 4: PCA analysis of the perceptual attributes for the comparison of PT and BDPT. The other two paired comparisons (ISM and PT; ISM and BDPT) present very similar results and are not included due to space constraints.

	Dimension 1	Dimension 2
Preference	0.46	-0.31
Realism/Naturalness	0.42	-0.25
ReverbQuality	0.38	-0.31
SpatImpression	0.43	-0.19
Distance	0.37	0.56
Localization	0.38	0.62

Table 4: PCA loading factors averaged over the three datasets.

Averaged loading factors for the three datasets are provided in Tab. 4. The projected scores for dimensions 1 and 2 of the PT vs BDPT comparison is shown in Fig. 4.

4.2 Room Dependency

The variety of rooms used in the test varies from very dry environments (living room with a RT of 0.4 s) up to a church with almost 5 seconds of reverberation. It is then reasonable to assume that there might be a strong dependency on the room properties on the perceptual ratings. Fig. 5 shows a series of figures detailing the scores of Dimension 1 as a function of reverberation time. Indeed, a positive correlation between reverberation time and scores favoring BDPT is observed. A potential explanation for this effect is the increased audibility of the reverberation in larger (and more reverberant) spaces, drawing the subjects' attention towards

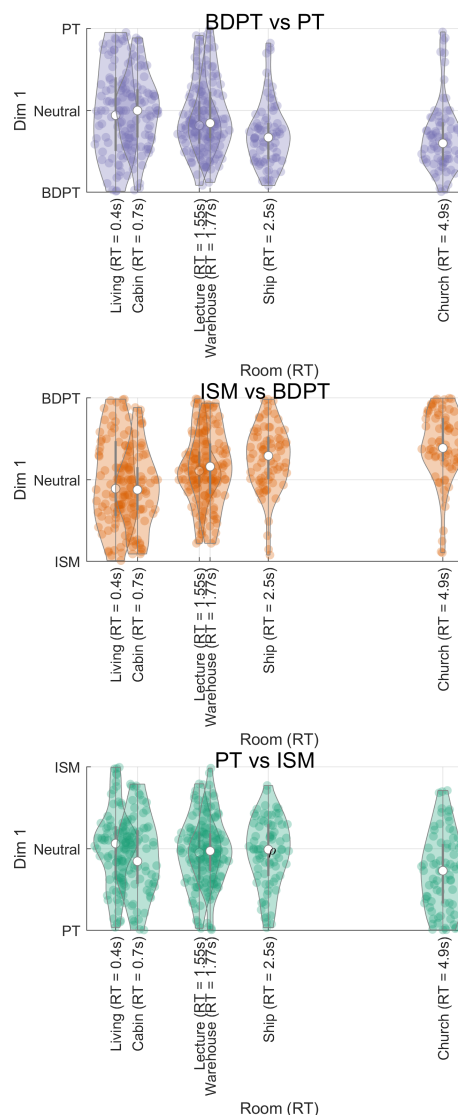


Fig. 5: Violin plots of the first PCA dimension as a function of Reverberation Time (RT).

the reverberation tail. In smaller spaces, scores present a much larger variance and are much more centered towards a neutral ratings. Scores for Dimension 2 (localization and distance) are largely neutral, suggesting that perceived distance and localization, which are the main contributors of this dimension, are not strong predictors of perceived differences. We do not include the corresponding graph due to space constraints.

Additionally, we include the raw histograms for the ratings of preference in Fig. 6. An interesting phe-

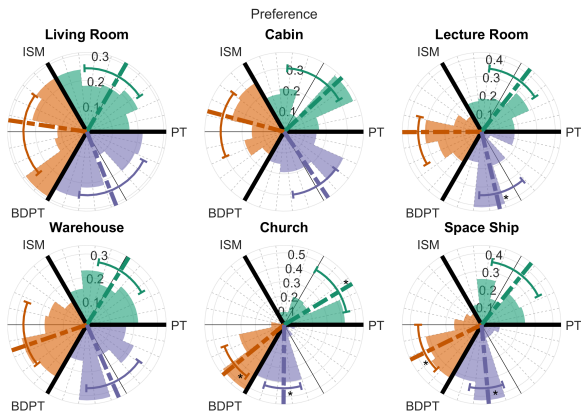


Fig. 6: Histograms of overall preference rating for each room. Each colored sector refers to the comparison between two engines. Dashed colored lines represent the median rating of each engine comparison and colored error bars within each sector represent interquartile ranges of the ratings. The values of the radius represent the empirical probability of the histogram bins. Two-sided sign tests were performed to test the hypothesis of the data coming from a distribution with zero median. Statistical significance of the rejection of the null hypothesis is denoted by * and corresponds to $p < 0.01$.

nomenon is that in the driest environment (living room) the preference judgments of BDPT present a bimodal distribution when comparing to the other engines. This suggests that listeners were not neutral, but rather divided. Additionally, for the cabin, which is a rather small and dry environment, we do not observe any statistically significant trend. The clearly strongest preference towards BDPT is in the larger and more reverberant spaces, and in the Lecture Room, Church, and Space Ship preference scores are significantly higher towards BDPT.

4.3 Source dependency

Similar to the analysis by rooms, we conducted an analysis investigating differences in ratings of each individual source. We conducted one-way ANOVA tests for both dimensions:

- Dimension 1: for ISM vs PT: $F(2, 465) = 3.49$, $p = 0.031$; for PT vs BDPT: $F(2, 465) = 3.05$, $p = 0.048$; for ISM vs BDPT: $F(2, 465) = 4.21$, $p = 0.015$.

- Dimension 2; for ISM vs PT: $F(2, 465) = 4.16$, $p = 0.016$; for PT vs BDPT: $F(2, 465) = 2.33$, $p = 0.098$; for ISM vs BDPT: $F(2, 465) = 0.62$, $p = 0.54$.

The results suggest that while it is possible that statistically significant differences exist in some cases, the trends are not clearly interpretable and the size effects are relatively small. Additionally, note that the sources were located in different positions in each room, and it is thus not possible to disentangle the effects due to source position to those caused by the actual source content.

5 Discussion

The gains of the overall best performing engine (BDPT) are not even across all the evaluated scenes, with the perceptual benefits vanishing in small rooms. This suggests that hybrid engines combining ISM with path tracing techniques, capable of producing strong coloration effects of early reflections in small rooms and long reverberation tails could pose a favorable solution [11]. However, it is unclear if the computation of high order image sources for multiple sources in environments of arbitrary geometry is feasible for real-time low cost applications on mobile VR devices. In these cases, strategies leveraging pre-computation of certain parts of the scene could be a promising avenue, as some engines are already doing [18].

One of the challenges of the study was to conduct the sessions remotely, while at the same time trying to ensure a uniformity in the setups and listening environments. However, this in fact raises several questions regarding the ecological validity of traditional listening experiments when evaluating immersive audio for practical applications, as tightly controlled laboratory environments do not translate to the conditions of final users. Although participants reported a wide variety of used headphones, this could indeed approximate a real world scenario.

In the present work we aimed at keeping the scenes relatively simple, without any moving objects and with only one active sound source. However, it is clear that in real applications scenes are generally dynamic and much more complex. It is then to be seen whether the results of the present experiment would generalize to more complex environments.

6 Conclusions

Three propagation engines (ISM, PT and BDPT) were compared in a formal listening test (N=26) featuring 6 scenes (living room, cabin, lecture room, warehouse, space ship, church) and judging 3 types of content (female speech, male speech, music). The tests were conducted using a setup composed of Quest/Quest 2 + Link and headphones. Listeners rated the various engines in pairwise comparisons using a bidirectional continuous scale, comparing them on 6 perceptual attributes: subjective preference, realism/naturalness, reverberation quality, spatial impression, perceived distance, and localization accuracy. The main findings are as follows:

- BDPT outperforms both PT and ISM on 4 attributes (preference, realism/naturalness, reverberation quality, spatial impression).
- The ratings towards BDPT are positively correlated with the reverberation time of the evaluated scenes. Thus, the benefits of BDPT are larger in more reverberant spaces.
- The variance of the responses is negatively correlated with the reverberation time, trending towards bimodal distributions in some cases. In other words, listeners could have strongly diverging preferences in dry environments.
- Localization and perceived distance are not strong predictors of perceived differences, as they present generally the neutral ratings.
- PCA analysis revealed that most of the variance (>80%) can be explained by two dimensions, which are linear combinations of the 6 rated attributes.
 - Dimension 1 (70 to 75% of explained variance) is composed of a roughly equal contribution of all attributes, although preference, realism/naturalness and spatial impression dominate.
 - Dimension 2 (5 to 10% of explained variance) is strongly dominated by distance and localization.

In further tests, both dynamic and more complex scenes should be evaluated, by progressively incorporating more elements into the experimental protocol.

References

- [1] Cunningham, S., Grout, V., and Hebblewhite, R., “Computer Game Audio: The Unappreciated Scholar of the Half-Life Generation,” in *Proceedings of Audio Mostly 2006*, 2006.
- [2] Johanson, C. and Mandryk, R. L., “Scaffolding Player Location Awareness through Audio Cues in First-Person Shooters,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 3450–3461, 2016, doi:10.1145/2858036.2858172.
- [3] Kulshreshth, A. and LaViola, J. J., “Evaluating performance benefits of head tracking in modern video games,” in *Proceedings of the 1st Symposium on Spatial User Interaction*, SUI '13, p. 53–60, 2013, doi:10.1145/2491367.2491376.
- [4] Rogers, K., Ribeiro, G., Wehbe, R. R., Weber, M., and Nacke, L. E., “Vanishing Importance: Studying Immersive Effects of Game Audio Perception on Player Experiences in Virtual Reality,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018, doi:10.1145/3173574.3173902.
- [5] Potter, T., Cvetkovic, Z., and De Sena, E., “On the Relative Importance of Visual and Spatial Audio Rendering on VR Immersion,” *Frontiers in Signal Processing*, 2, 2022, doi:10.3389/frsip.2022.904866.
- [6] Grimshaw, M., “Sound and immersion in the first-person shooter,” in *Proceedings of CGAMES'2007*, 2007.
- [7] Amengual Gari, S. V., Calamia, P., and Robinson, P., “Navigation of virtual mazes using acoustic cues,” in *Audio Engineering Society Convention 154*, 2023.
- [8] Mehra, R., Rungta, A., Golas, A., Lin, M., and Manocha, D., “WAVE: Interactive Wave-based Sound Propagation for Virtual Environments,” *IEEE Transactions on Visualization and Computer Graphics*, 21(4), pp. 434–442, 2015, doi:10.1109/TVCG.2015.2391858.
- [9] Hiebert, G., “OpenAL 1.1 Specification and Reference,” *Creative Labs, Inc.*, 2005.

- [10] 3D Working Group of the Interactive Audio Special Interest Group, “Interactive 3D Audio Rendering Guidelines Level 2.0,” *Technical report*, 1999.
- [11] Schröder, D. and Vorländer, M., “RAVEN: A real-time framework for the auralization of interactive virtual environments,” in *Forum acusticum*, pp. 1541–1546, Aalborg Denmark, 2011.
- [12] Wendt, T., Van De Par, S., and Ewert, S. D., “A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation,” *Journal of the Audio Engineering Society*, 62(11), pp. 748–766, 2014.
- [13] Raghuvanshi, N., Snyder, J., Mehra, R., Lin, M., and Govindaraju, N., “Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes,” in *SIGGRAPH 2010*, 2010, doi:10.1145/1833349.1778805.
- [14] Schissler, C. and Manocha, D., “Interactive sound propagation and rendering for large multi-source scenes,” *ACM Transactions on Graphics (TOG)*, 36(4), p. 1, 2016.
- [15] “Steam Audio,” <https://valvesoftware.github.io/steam-audio/>, Accessed: 2023-11-03.
- [16] “Audiokinetic Wwise,” <https://www.audiokinetic.com/en/products/wwise/>, Accessed: 2023-11-03.
- [17] “Oculus Audio Propagation (Beta),” https://developer.oculus.com/documentation/unity/audio-osp-unity-propagation/?locale=en_GB, Accessed: 2023-11-03.
- [18] “Microsoft: What is Project Acoustics?” <https://learn.microsoft.com/en-us/gaming/acoustics/what-is-acoustics>, Accessed: 2023-11-03.
- [19] “Resonance Audio: Multi-platform spatial audio at scale,” <https://blog.google/products/google-ar-vr/resonance-audio-multi-platform-spatial-audio-scale/>, Accessed: 2023-11-03.
- [20] Amengual Garí, S. V., Schissler, C., Mehra, R., Featherly, S., and Robinson, P., “Evaluation of real-time sound propagation engines in a virtual reality framework,” in *2019 AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.
- [21] Schörkhuber, C., Zaunschirm, M., and Höldrich, R., “Binaural rendering of ambisonic signals via magnitude least squares,” in *Proceedings of the DAGA*, volume 44, pp. 339–342, 2018.
- [22] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C., “The CIPIC HRTF database,” in *WASPAA 2001*, pp. 99–102, IEEE, 2001.
- [23] Allen, J. B. and Berkley, D. A., “Image method for efficiently simulating small-room acoustics,” *The Journal of the Acoustical Society of America*, 65(4), pp. 943–950, 1979.
- [24] Schroeder, M. R., “Natural sounding artificial reverberation,” in *Audio Engineering Society Convention 13*, Audio Engineering Society, 1961.
- [25] Cao, C., Ren, Z., Schissler, C., Manocha, D., and Zhou, K., “Interactive sound propagation with bidirectional path tracing,” *ACM Transactions on Graphics (TOG)*, 35(6), pp. 1–11, 2016.
- [26] Chen, C., Jain, U., Schissler, C., Gari, S. V. A., Al-Halah, Z., Ithapu, V. K., Robinson, P., and Grauman, K., “Soundspaces: Audio-visual navigation in 3d environments,” in *Computer Vision–ECCV 2020: 16th European Conference*, pp. 17–36, Springer, 2020.
- [27] Purushwalkam, S., Gari, S. V. A., Ithapu, V. K., Schissler, C., Robinson, P., Gupta, A., and Grauman, K., “Audio-visual floorplan reconstruction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1183–1192, 2021.
- [28] Gao, R., Chen, C., Al-Halah, Z., Schissler, C., and Grauman, K., “Visualechoes: Spatial image representation learning through echolocation,” in *Computer Vision–ECCV 2020: 16th European Conference*, pp. 658–676, Springer, 2020.
- [29] Chen, C., Schissler, C., Garg, S., Kobernik, P., Clegg, A., Calamia, P., Batra, D., Robinson, P.,

and Grauman, K., “Soundspaces 2.0: A simulation platform for visual-acoustic learning,” *Advances in Neural Information Processing Systems*, 35, pp. 8896–8911, 2022.