



Audio Engineering Society

Convention Express Paper 322

Presented at the 158th Convention

2025 May 22-24, Warsaw, Poland

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 The complete manuscript was not peer reviewed.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.

Comparing Artificially Created Acoustic Environments to Real Space Responses: Integrating Objective Metrics and Subjective Perceptual Listening Tests

Emilia Kalliri¹, Rigas Kotsakis², Nikolaos Vryzas¹, Charalampos Dimoulas¹, and George Kalliris¹

¹ Aristotle University of Thessaloniki, Greece

² International Hellenic University, Greece

Correspondence should be addressed to Nikolaos Vryzas: nvryzas@auth.gr

ABSTRACT

This study evaluates the effectiveness of artificial reverberation algorithms that are used to create simulated acoustic environments by comparing them to the acoustic response of the real spaces. A mixed-methods approach, integrating objective and subjective measures, was employed to assess both the accuracy and perceptual quality of simulated acoustics. Real-world spaces, within the context of the SCENE research project, were selected for their varying sizes, functions, and acoustical properties. Objective acoustic measurements—such as Room Impulse Response (RIR), and extracted features i.e. Reverberation Time (RT60), Early Decay Time (EDT), Clarity index (C50, C80), and Definition (D50)—were conducted to establish baseline profiles. Simulated environments were created to replicate real-world conditions, incorporating source-receiver configurations, room geometries, and/or material properties. Objective metrics were extracted from these simulations for comparison with real-world data. After applying the artificial reverberation algorithm, the same objective measurements were re-recorded to assess its impact. Subjective listening tests were also conducted, with a diverse panel of listeners rating the perceived clarity, intelligibility, comfort, and overall sound quality of both real and simulated spaces, using a double-blind procedure to mitigate bias.

1 Introduction

Artificially created acoustic environments serve a wide range of applications across various fields. The reliability of such simulations as compared to real room responses requires objective and subjective evaluation such as acoustic measurements and listening tests. Listening tests are critical tools for assessing how individuals perceive and interact with auditory stimuli in various acoustic environments. These tests, widely used in psychoacoustics, room

acoustics, and sound design, help optimize acoustic spaces and audio systems by examining both objective and subjective parameters. Objective tests focus on quantifiable sound attributes like loudness, clarity, and reverberation, often using paired comparisons or descriptive analysis methodologies. Subjective tests, on the other hand, evaluate listener preferences and emotional responses through integrative scaling methods, such as Likert scales and semantic differential scales. Combining

subjective and qualitative methods, including interviews and focus groups, allows researchers to explore cultural and contextual influences on sound perception, enhancing acoustic design. In his 1962 book, *Music, Acoustics, and Architecture* [1], pioneering acoustician Leo Beranek, conducted extensive research to correlate objective measurements with subjective evaluations of concert hall acoustics. He analyzed and surveyed 54 concert halls and opera houses to identify physical parameters that influence perceived sound quality. Through interviews with musicians and conductors, he gathered subjective evaluations of these venues. In a 2003 study titled "Subjective Rank-Orderings and Acoustical Measurements for Fifty-Eight Concert Halls," Beranek expanded his analysis. One of Beranek's key findings is that a longer reverberation time (RT) of 1.8 to 2.2 seconds can create a sense of fullness and warmth in the sound. In contrast, a shorter early decay time (EDT) is associated with clarity and liveness. Another important measurement is the binaural quality index (BQI), which relates to the interaural cross-correlation coefficient (IACC) and is perceived as spaciousness. A lower IACC can enhance the listener's spatial impression and sense of envelopment. Additionally, the strength (G) measurement directly assesses sound presence and loudness, with higher G levels being unmistakably perceived as louder or more powerful. The bass ratio (BR) plays a significant role in creating warmth, with elevated levels producing a distinctly warmer auditory experience. The spaciousness of sound is effectively captured by the lateral energy fraction (LEF). Lastly, the initial time delay gap (ITDG) is fundamentally correlated with the sense of intimacy; specifically, delays of 10 to 20 milliseconds create a strong feeling of closeness to the performers. Even though balance between these parameters is crucial, the most important are clarity and definition. C50 is used mainly for speech (50 ms cutoff), while C80 is utilised for music (80 ms cutoff). Lower values of C80 offer a more blended, lush sound but with less definition. Higher C80 values (+2 to +4 dB for music) give a more precise, articulated, and transparent musical sound. Moreover, definition (D50) is the percentage of sound energy arriving in the first 50 ms of the impulse response. Increased D50 values (typically above 50%) indicate better intelligibility. Higher D50 leads to precise rhythmic articulation.

The most crucial factor for implementing an effective experimental setup for listening tests/subjective comparisons, is the artificial reverberation

generation, which has been extensively investigated for improving audio realism and immersion. Recent research evaluated seven reverberation algorithms, including Feedback Delay Network (FDN) and Hybrid Moorer-Schroeder (HMS), using expert listening tests and the MUSHRA paradigm to analyze perceptual similarities between artificial and real reverberation [2]. Another study proposed a feedback delay network reverberator with a modified cascaded graphic equalizer as an attenuation filter, analyzing decay characteristics and reverberation times to assess accuracy [3]. The perceptual differences between real and artificial height channels in surround sound systems were also examined, highlighting challenges in achieving realistic spatial audio reproduction [4]. The perceived naturalness of various artificial reverberation algorithms was assessed through listening tests employing modified MUSHRA methods to rank their effectiveness in replicating real reverberation [5]. More recently, a parameter optimization approach was introduced to automatically adjust artificial reverberator settings for matching real room acoustics, improving replication accuracy [6]. Collectively, these studies provide diverse methodologies for evaluating perceptual and acoustic fidelity in artificial reverberation research.

Room acoustics with associated perceptual/sentiment analysis and spatial audio assessment procedures are vital in high fidelity sound applications, including cultural events and related sound streaming experiences [7]-[12]. Machine and deep learning techniques have been enormously elaborated, extending their use from typical sound classification, sentiment analysis and summarization applications to more sophisticated spatial audio design and generation [13]-[17].

2 Problem Definition

This study aims to provide an infrastructure for evaluating the ability of enhanced simulated acoustics and compare them with real-space responses, contributing to more effective sound design and sound processing techniques. Within the context of the SCENE research project a comparison of real space recordings to simulated ones needed to be evaluated by using objective measures and subjective evaluations. The main scope of SCENE project is to create a dynamic and constantly evolving audio dataset that would facilitate efficient transferability of room acoustics based on a restricted number of measurements of the corresponding physical space. Limitations, such as

the small sample size and potential biases in subjective ratings, will be considered in the interpretation of results.

3 Methodology

The current research is focused on the comparison of the reverberation characteristics between real and artificially created acoustic environments towards the evaluation of the effectiveness of room acoustic simulation algorithms. The first step involves selecting a range of real-world spaces that differ in size, function, and acoustical properties, Church of Panagia of Sintis in a 16th century orthodox monastery in Cyprus, Olympia City Music Theatre "Maria Callas" in central Athens, and Church of Santa Caterina in Italy will serve as the baseline for comparison. In each real space, a series of acoustic measurements will be conducted, to obtain the Room Impulse Response (RIR), and extract Reverberation Time (RT60), Early Decay Time (EDT), Clarity index (C50, C80), and Definition (D50). These objective metrics will provide a comprehensive understanding of the reverberation characteristics within each space. On the other hand, simulated spaces will be generated incorporating accurate room geometries, material properties, and source-receiver positions to replicate the real-world conditions as closely as possible. Once the virtual environments are set up, subjective listening tests will be conducted to evaluate the perceptual impact of the algorithm on reverberation. A listener panel, consisting of individuals with varying levels of experience in acoustics, will rate the perceived clarity, comfort, intelligibility, and overall sound quality of the real and simulated spaces. This will be done through a double-blind procedure to mitigate bias. This combined objective and subjective approach will provide a comprehensive evaluation of the algorithm's effectiveness in optimizing reverberation characteristics and improving the perceptual quality of simulated acoustic environments.

4 Objective Metrics

As already mentioned, to obtain the RIRs of the selected real-world spaces acoustic measurements were carried out. Using the REW software two logarithmic sine sweeps were used for comparison reasons i.e. a short (256k-5.5s) and a longer version (512k-10.9s). For each real space and for every listening position two RIRs were obtained and extracted from REW. Using AURORA plugins in Audacity the required ISO3382 acoustical

parameters for each RIR were obtained. According to ISO3382 acoustical parameters are calculated in 10 octave bands, and 2 overall, i.e. A-weighted and Linear. For the purpose of this work the A-weighted calculations of the acoustical parameters of each measured RIR were selected. A comparison of A-weighted acoustical parameters is shown on table 1 for two listening positions of Olympia Hall with both the short and long sweep measurements. From this comparison we find that EDT[s] is higher in position 2 so we expect higher Spaciousness / Envelopment ratings from the listening tests. In position 1, C50/D50 are higher than in position 2 so we expect higher Clarity / Intelligibility ratings for speech but as C80s are similar for both positions we expect similar ratings for music. Another observation is that the G parameter (strenGth) is different between short and long sweep measurements probably due to lower accuracy of the short sweep. This is expected to be reflected on the Overall Realism rating of the listening tests.

Acoustical Parameter	Unit	OL1-S	OL1-L	OL2-S	OL2-L
Signal	[dB]	84.07	81.75	81.89	80.39
Noise	[dB]	29.68	26.43	22.40	20.60
strenGth	[dB]	7.07	4.75	4.89	3.39
C50	[dB]	16.60	17.01	10.12	10.05
C80	[dB]	18.40	18.80	17.01	16.91
D50	[%]	97.86	98.05	91.14	91.01
Ts	[ms]	5.45	5.05	9.15	9.17
EDT	[s]	0.33	0.38	1.79	1.82
Tuser	[s]	0.28	0.27	--	--
T20	[s]	0.65	0.64	0.52	0.52
T30	[s]	0.78	0.78	0.67	0.67
Peakiness	[dB]	40.15	41.81	40.39	43.10
Millisecondness	[dB]	31.91	32.88	32.80	32.20
Impulsiveness	[dB]	18.70	18.70	18.29	18.30

Table 1. Comparison of A-weighted acoustical parameters for two listening positions, short and long sweep measurements.

5 Subjective Listening Tests

For the implementation of the subjective listening tests, four different acoustic environments were selected and for each environment one reference real space recording was provided and three simulations were asked to be evaluated. In total 16 sound files of 22 s duration were tested. A rating of Likert scale from 1 to 7 was used. As this first study was considered a pilot research that was sent to experts,

questions about the form of the test were also included such as, to evaluate the selected Likert scale, the duration of the sound files and to propose improvements. Four different groups of descriptors were chosen to be evaluated for each simulation (three in total) plus one hidden reference. In figure 1 a sample image of the online evaluation screen and the associated descriptors are depicted.

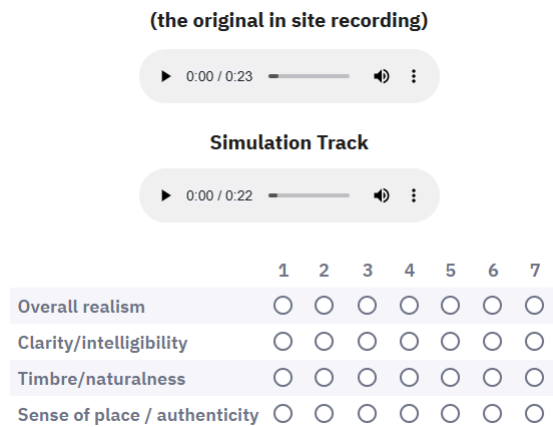


Figure 1. The online evaluation screen with the associated descriptors to be evaluated in a Likert 1-7 scale.

The evaluators were free to listen, each track as many times as they wanted, and they were also informed that they were able to stop the track before the end if they were ready to evaluate it. Headphones were asked to be used by the evaluators.

6 Results

Nine (9) experts were chosen to participate as evaluators and all of them responded to complete the listening test. Applying simple descriptive statistics, one key finding is that, as expected, in all four acoustic environments under test, the hidden reference obtained the best scores in all descriptors with the highest being "Overall realism" and "Sense of place / authenticity" (mean values from 5.5 to 5.8). Moreover, the hidden reference obtained the highest agreement between participants (minimum standard deviation values between 0.6 to 1.5). On the other hand, the lowest scores were found for both the short and long sweep simulations using the RIRs of the measurement in all environments. Especially for the two listening positions of Olympia Hall the objective metrics expectations were not verified by the listening tests probably because of the limitations of the experimental process that need

further investigation. Finally, it is important to note that the Jaura algorithm received very good scores (mean values 6.7) in "Clarity/intelligibility" and "Timbre/naturalness".

7 Conclusions – Future Work

This research within the SCENE project, provided the infrastructure for evaluating the ability of enhanced simulated acoustics and comparing them with real-space responses. Comparing artificial acoustic environments to real responses by integrating objective metrics and subjective perceptual listening tests is a very challenging task. There are several factors that need to be considered to draw reliable results. In this study, some results were expected but others were not. As already mentioned, participants were also asked to express their opinion about the elements of the listening test. They mostly agreed (mean value 3.1, standard dev 0,8) that the descriptors used as evaluation criteria within the test were neither very unclear nor very clear, that means a learning procedure may be needed. The Likert evaluation scale (1-7) was found to be a bit longer than expected (mean value 3.55). Finally, they all agreed (mean value 3) that the audio track's duration was appropriate. Future work should build on the findings of this research. Furthermore, statistical analyses should be performed including paired t-tests and correlation analysis, to assess the relationship between objective and subjective evaluations. This approach will provide a comprehensive framework for evaluating the algorithms' ability to enhance simulated acoustics and align them with real-world environments.

8 Acknowledgement



agreement No 101095303.

This research is part of the SCENE project and has received funding from the European Union's Horizon research and innovation programme under grant

References

- [1] L. L. Beranek, *Music, Acoustics & Architecture*. New York: Wiley, 1962.
- [2] M. Huan, G. Kearney and H. Daffern, "Perceptual Similarities between Artificial

- Reverberation Algorithms and Real Reverberation" *Applied Sciences* vol. 13, no. 2, pp 840 (2023).
- [3] K. Prawda, V. Välimäki and S. Serafin, "Evaluation of Accurate Artificial Reverberation Algorithm" in *Proceedings of the 17th Sound and Music Computing Conference*, Torino, June 24th – 26th, 2020, pp. 247-254.
- [4] R. L. King, B. Leonard, W. Howie and J. Kelly, "Real Rooms vs. Artificial Reverberation: An evaluation of actual source audio vs. artificial ambience" in *Proceedings of Meetings on Acoustics*, Vol. 29, no. 1, 2016.
- [5] S. Djordjevic, H. Hacıhabiboglu, Z. Cvetkovic and E. de Sena, "Evaluation of the perceived naturalness of artificial reverberation algorithms" in *Proceedings of 148th Audio Engineering Society International Convention 2020*, Vienna, Austria.
- [6] R. Bona, D. Fantini, G. Presti, M. Tiraboschi, J. I. E. Alonso-Martinez and F. Avanzini, "Automatic Parameters Tuning of Late Reverberation Algorithms for Audio Augmented Reality" in *Proceedings of AudioMostly 2022 (AM '22) Conference*, September 6–9, 2022, St. Pölten, Austria.
- [7] N. Vryzas, M. Matsiola, R. Kotsakis, C. Dimoulas and G. Kalliris, G., "Subjective evaluation of a speech emotion recognition interaction framework." In *Proceedings of the audio mostly 2018 on sound in immersion and emotion* (pp. 1-7), 2018.
- [8] G. Kalliris, M. Matsiola, C. Dimoulas and A. Veglis, "Emotional aspects and quality of experience for multifactor evaluation of audiovisual content." *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, 2(4), 40-61, 2014.
- [9] I. Thoidis, N. Vryzas, L. Vrysis, R. Kotsakis, G. Kalliris and C. Dimoulas, "Disentangled estimation of reverberation parameters using temporal convolutional networks." In *Audio Engineering Society Convention* (Vol. 152), 2022, May.
- [10] C. E. Vegiris, K. Avdelidis, C. A. Dimoulas and G.V. Papanikolaou, "Live broadcasting of high definition audiovisual content using HDTV over broadband IP networks", *International Journal of Digital Multimedia Broadcasting*, 2008(1), 250654, 2008.
- [11] R. Kotsakis, C. Dimoulas, G. Kalliris and A. Veglis, "Emotional prediction and content profile estimation in evaluating audiovisual mediated communication", *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, 2(4), 62-80, 2014.
- [12] N. Vryzas, M.E. Stamatiadou, L. Vrysis and C. Dimoulas, "Multichannel mobile audio recordings for spatial enhancements and ambisonics rendering", In *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)* (pp. 1-6). IEEE, 2023.
- [13] C. A. Dimoulas and G.M. Kalliris, "Investigation of wavelet approaches for joint temporal, spectral and cepstral features in audio semantics", In *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.
- [14] R. Kotsakis, M. Matsiola, G. Kalliris and C. Dimoulas, "Investigation of spoken-language detection and classification in broadcasted audio content", *Information*, 11(4), 211, 2020.
- [15] N. Vryzas, L. Vrysis, R. Kotsakis and C. Dimoulas, "A web crowdsourcing framework for transfer learning and personalized speech emotion recognition", *Machine Learning with Applications*, 6, 100132, 2021.
- [16] N. Vryzas, L. Vrysis, R. Kotsakis and C. Dimoulas, "Speech emotion recognition adapted to multimodal semantic repositories", In *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)* (pp. 31-35). IEEE, 2018.

- [17] M. Heydari, M. Souden, B. Conejo and J. Atkins, "ImmerseDiffusion: A Generative Spatial Audio Latent Diffusion Model" In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE, 2025.