DOI: 10.61782/fa.2023.0745

INTERACTIVE PERCEPTUAL EVALUATION OF AURALIZED ACOUSTICS IN REHEARSAL SPACES FOR VOCAL MUSIC

Hannes Rosseel[1] **Toon van Waterschoot**[1]
[1] Dept. of Electrical Engineering (ESAT-STADIUS), KU Leuven, Leuven, Belgium

ABSTRACT

Rapid sensory analysis methods such as Flash Profiling, where individually elicited attributes are used to evaluate the perceptual differences among different sensory experiences, have successfully been applied to the perceptual evaluation of reproduced acoustics of concert halls, listening rooms, and car cabins. In this paper, a pilot experiment was conducted to identify the perceptual differences between four spatially reproduced sound fields in an interactive manner. The spatial sound fields were reproduced for a 24-loudspeaker setup, installed in a rehearsal space for vocal music, using a commercially available real-time auralization engine. An expert panel consisting of four assessors, who are all part of a musical ensemble specialized in vocal music, interacted with each sound field by performing pieces of plainchant recorded using a microphone and processed by a computational engine that delivered the auralized sound field to the assessors in real-time. The results showed that the rapid sensory analysis method, used interactively, could reveal the spatial and timbral characteristics of the reproduced sound field through agreement of the individually elicited attributes with all assessors. Moreover, it was shown that Flash Profiling is a valid method for evaluating interactive auralizations of spatial room acoustics.

Keywords: *perceptual evaluation, auralization, spatial room acoustics*

****Corresponding author***: *hannes.rosseel@esat.kuleuven.be.*

1. INTRODUCTION

The acoustical properties of a venue play an important role in the quality of vocal music performances. Several studies have shown that the performance of vocal chant is intrinsically related to the acoustical characteristics of the venue [1–3]. To deepen our understanding of this relationship, there is a need for efficient sensory analysis techniques that enable the direct comparison and identification of the underlying acoustical characteristics in the context of interactive auralization of spatial room acoustics.

Such sensory evaluation techniques have already been extensively used in listening tests to compare the perceptual differences of auralized sound fields in concert halls and auditorium acoustics [4, 5], and in automotive acoustics to identify the perceptual differences in the listening experience of car cabins [6, 7], as well as small-room acoustics [8]. Another study utilized sensory evaluation techniques to investigate the perceived cathedral ceiling height while listening to Gregorian chant [9].

To date, the research on interactive perceptual analysis for auralized spatial room acoustics seems to be limited. In existing literature regarding interactive auralization, one study examined the influence of interactive virtual room acoustics on choir singing [10], but the analysis was based solely on objective parameters of the recorded chant. In another study, [11] described a real-time acoustic reproduction setup for musical instruments, but no perceptual analysis was performed.

The aim of this study is to address the current research gap by conducting a pilot experiment that aims to identify the perceptual differences among four spatially reproduced sound fields through interaction. To achieve this, the rapid sensory evaluation protocol called Flash Profiling (FP) [12, 13], previously used in the aforementioned listening tests [5–8], was tested on interactive perceptual evaluation.

The paper is structured as follows: Section 2 describes the real-time acoustic reproduction setup that was used to reproduce the virtual room acoustic stimuli in the experiment. The general procedure of the pilot experiment is explained in Section 3, after which the results of the experiment are analyzed and interpreted in Section 4. Finally, the conclusions of the paper are presented in Section 5.

2. METHODS

2.1 Interactive Auralization setup

The real-time synthesis of different acoustic environments was performed in the Alamire Interactive Laboratory (AIL), which is located in Saint Norbert's Gate at the Park Abbey site in Leuven, Belgium. The AIL, as seen in Figure 1, is an interactive spatial auralization lab specifically designed for real-time auralization of virtual acoustics in relation to early music performance. The primary goal of the lab is to provide early music performers with a tool to investigate the connection between early music performances and the acoustics of the venue in which they were originally conceived.

The audio setup of the lab consists of a multichannel loudspeaker array containing 24 loudspeakers (Martin Audio CDD6) and six microphones (Audio Technica AT4051b), which are mounted on top of the loudspeaker array. The loudspeakers and microphones are connected to amplifiers (microphones: Shure-ANI4IN, loudspeakers: Powersoft Ottocanali 8K4 DSP+D), which connect to a Dante audio-over-Ethernet network developed by Audinate. Dante is a network protocol that is capable of transmitting uncompressed multichannel digital audio over a standard Ethernet network, with low and deterministic latency (< 1 ms), making it highly suitable for real-time interactive auralization setups. The whole setup is controlled by a commercially available audio rendering processor by Astro Spatial Audio, called SARA II [14], which is capable of rendering virtual room acoustic models in real-time. An initial calibration of the system was performed such that each loudspeaker and microphone was level-matched at the center point of the reproduction setup.

2.1.1 Limitations

Acoustic auralization setups are typically assumed to exhibit anechoic characteristics, ensuring that the auralized sound field remains unaffected by the acoustics of the reproduction environment. However, these assumptions do not often hold due to practical limitations in the implementation. While the AIL was acoustically treated with two soundproof drapes, each on one side of the reproduction area, acoustic reflections remained present in the environment. During the reproduction of virtual acoustics, this causes an effect known as the room-in-room effect [15, 16], where the perception of the auralized room acoustics is influenced by the reproduction room acoustics.

Figure 1: The Alamire Interactive Laboratory, located in Saint Norbert's Gate at the Park Abbey site in Leuven, Belgium.

Moreover, due to the presence of microphones positioned in proximity to the reproduction loudspeakers, the reproduction setup is prone to the multichannel acoustic feedback problem, where a closed amplification loop could result in the unstable behavior of the reproduction setup. This limits the maximum available dynamic range in the reproduction of the virtual room acoustics and affects the perception of the reproduced acoustics.

2.2 Virtual rehearsal spaces

The SARA II rendering processor comes preinstalled with a software package called Room Simulation Software Module (RSM Pro) [14], capable of rendering interactive auralizations of virtual room acoustic models in real-time using a proprietary implementation developed by Fraunhofer IDMT called SpatialSound Wave (SSW) [17]. SSW originates from Wave Field Synthesis (WFS) [18] combined with additional psychoacoustic principles to spatialize monophonic sound sources for multichannel loudspeaker arrays. Each room acoustic model offers a virtual rehearsal space with a unique sound field that is suited for different types of musical performances. The system comes preinstalled with four preset models, which are called: *Dome*, *Church*, *Concert Hall*, and *Chamber Music*, and are sorted in order of reverberation time, from high to low. The virtual room acoustic models can be activated and controlled remotely from a web interface displayed on a computer located outside the rehearsal space. This enables seamless switching between virtual room acoustic models.

For every virtual room acoustic model, two sets of room impulse responses (RIRs) were measured. The initial set was measured while reproducing the room acoustic models inside the reproduction area. The second set was obtained by measuring the RIR of every room acoustic model directly from the rendering processor, bypassing the acoustics of the reproduction space. This enabled a comparison between the intended virtual room model and its reproduction. The measurements inside the reproduction area were conducted by placing a loudspeaker (Genelec 8030C) and an open microphone array (G.R.A.S. VI-50) at the center of the reproduction setup. The microphone array consists of six omnidirectional condenser microphones arranged in pairs along each Cartesian axis with a 50 mm spacing. The loudspeaker and microphone array were positioned one meter apart at a height of 170 cm. The RIR was captured by playing back an exponential sine-sweep [19] through the loudspeaker, with a duration of four seconds at a sampling rate of 192 kHz. The response of the room was captured using an RME FireFace UFX+ multichannel audio interface while the reproduction system was active. The captured RIR thus includes the auralized room acoustic model reproduced by the system. The magnitude responses for the reproduced and intended room acoustics are shown in Figure 2. It can be observed that the spectra of the reproduced acoustics deviate from the spectra of the intended acoustics. This can be explained due to the limited available dynamic range in the reproduction of the virtual room acoustics, which was limited to avoid acoustic feedback artifacts in the reproduction, as discussed in Section 2.1.1. The reproduced acoustics are still perceived inside the reproduction setup, but the spectral characteristics of the intended acoustics are not fully preserved, as they are masked by the spectral characteristics of the reproduction environment.

In Table 1, the Objective Room Acoustic Parameters (ORAP) are presented for the reproduced (top) and the intended (bottom) room acoustic models. The parameters for the reproduced room acoustic models were calculated from the RIR measured by the topmost microphone in the microphone array. The parameters were calculated in accordance with ISO:3382 [20, 21]. It can be seen that the ORAP of the reproduced room acoustic models do not match those of the intended room acoustic models. This can be explained from the limitations that were present in the reproduction setup, as discussed in Section 2.1.1. Since the room-in-room effect can be modeled as a convolution between the RIR of the reproduction area and the RIR of the intended room acoustic model, it not only affects the perception of the reproduced acoustics but also impacts ORAP such as Reverberation Time (RT30) and Early Decay Time (EDT) [16]. It is therefore advised to be cautious when interpreting the ORAP of reproduced acoustics in a non-anechoic environment.

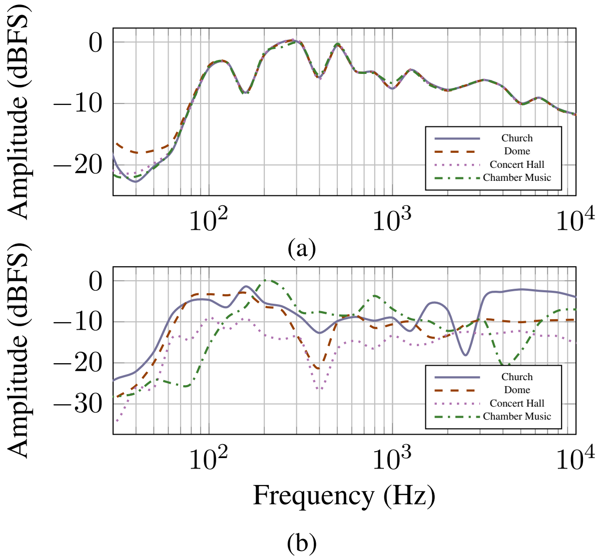

Figure 2: Magnitude responses for the (a) reproduced and (b) intended room acoustic models. These responses were calculated after removal of the direct path component. A 1/3 octave band averaging was performed for visualization purposes.

3. PILOT EXPERIMENT

3.1 Procedure

The experiment followed the general procedure and principles of FP [12, 13]. Prior to the experiment, the participants were briefed on the experimental methodology and given the opportunity to ask questions. The participants were encouraged to complete the presented tasks at their preferred pace and to take breaks when needed.

Table 1: Objective acoustic parameters of the reproduced (top), and intended (bottom) virtual room acoustic models. RT30 and EDT were averaged in the 500 Hz and 1 kHz octave bands (ISO:3382) [20, 21].

Reproduced model	$RT30_{0.5-1k}$ (s)	$EDT_{0.5-1k}$ (s)	C50 (dB)	D50 (dB)
Dome	1.14	0.23	18.32	0.99
Church	0.69	0.23	17.60	0.98
Concert Hall	0.58	0.23	18.14	0.98
Chamber Music	0.46	0.25	17.82	0.98
Intended model	$RT30_{0.5-1k}$ (s)	$EDT_{0.5-1k}$ (s)	C50 (dB)	D50 (dB)
Dome	1.87	1.90	2.36	0.63
Church	1.11	1.01	6.74	0.83
Concert Hall	0.97	1.05	4.61	0.74
Chamber Music	0.42	0.21	24.59	1.00

Even though the participants were required to perform vocal chant in group, they were instructed not to discuss their individual perception of the stimuli with one another throughout the duration of the experiment. During the experiment, the participants stood around a music stand that held the music score, which they used during their performance.

The experiment presented the virtual room acoustic models, further referred to as stimuli, in a randomized presentation order and assigned numerical identifiers to each stimulus. The participants were only given the numerical identifiers instead of the stimulus names to prevent possible bias. The following order was used: (1) *Concert Hall*, (2) *Church*, (3) *Dome*, (4) *Chamber Music*. During the experiment, each stimulus was activated by the experimenter using the ASTRO web interface. After activation, the participants were invited to engage with the stimulus in an interactive manner through vocal performance, such as performing a piece of polyphonic or Gregorian chant, within the reproduction setup. Participants had to observe the response of the virtual room acoustic stimulus while singing.

The pilot experiment was conducted in the interactive auralization setup, described in Section 2.1, and consisted of two consecutive phases, namely, the attribute elicitation phase, followed by the attribute ranking phase.

3.1.1 Elicitation Phase

During the elicitation phase, the participants were presented with the stimuli and immediately afterward had to elicit as many descriptive attributes as necessary to identify the perceived acoustic differences between the presented virtual room acoustic stimuli. The elicited attributes had to be scalar, unidimensional, and non-hedonic and had to be written on a questionnaire that was given to each participant before the start of the experiment. After the initial presentation of all stimuli, the participants could switch between the stimuli as desired by notifying the experimenter, who then changed the stimulus using the web interface described in Section 2.2. After the elicitation phase, a brief intermission was held to permit participants to rest their vocal cords and auditory system.

3.1.2 Ranking Phase

During the ranking phase, each participant received one questionnaire for each presented stimulus, on which they were instructed to rank each of the individually elicited perceptual attributes that they came up with based on the perceived intensity during each stimulus. The questionnaire was equipped with a continuous scale ranging from ”Low” to ”High” on which the participants could mark the perceived intensity of each perceptual attribute for each stimulus. During the ranking phase, participants could switch between stimuli as desired.

3.2 Participants

In this pilot experiment, four expert performers of plainchant volunteered as participants in the interactive perceptual assessment of the virtual room acoustic models. The participants were experienced in early music and chant, and had, due to their line of work, performed in a variety of halls and musical spaces, giving them expert knowledge of the perceptual characteristics of acoustics while performing musical chant.

4. EXPERIMENTAL DATA ANALYSIS

This section presents the results from the pilot experiment described in Section 3. The experiment employs an interactive version of FP, a rapid sensory evaluation technique, on four interactive virtual acoustic room models that were reproduced inside a reproduction area. Four expert performers of plainchant participated in this experiment and elicited a total of 24 perceptual attributes, which describe the perceptual differences among the acoustic characteristics of the models. Upon investigation of the data, one of the attributes, "Enjoyment" was found to be hedonic and therefore subject to individual interpretation. This attribute was excluded from further analysis.

The remaining perceptual attributes were analyzed with Multiple Factor Analysis (MFA) [22] using the *FactoMineR* package [23]. The goal of MFA is to construct a consensus space among the participants' chosen attributes and ratings by identifying common trends among groups of variables. First, the perceptual attributes are grouped per participant, after which Principal Component Analysis (PCA) is applied separately to each group of variables. The resulting PCA data are then normalized and brought together in a global matrix, on which a final PCA is performed. The outcome of the analysis positions the stimuli on a consensus space, which is called the factor map. This factor map is used to interpret the perceptual relationships between the different stimuli and allows for the identification of the most prominent perceptual characteristics that generate the variance in the data. These perceptual characteristics can provide insight into the most significant acoustical characteristics for performers of chant music.

The following sections outline the MFA data analysis. In Section 4.1, MFA is used to construct a consensus space from the sensory data obtained in the pilot experiment. In Section 4.2, the data is further analyzed by clustering the perceptual attributes that correlate significantly with each dimension in the consensus space. This analysis helps in gaining insight into the perceptual meaning of the dimensions in the consensus space.

4.1 Sensory data analysis using MFA

To minimize potential scaling effects, the attribute values were centered and mean normalized before applying MFA. Figure 3 shows the resulting factor map after MFA was performed. This map visualizes the position of the virtual room acoustic stimuli in the subspace spanned by the first two principal components of the consensus space.

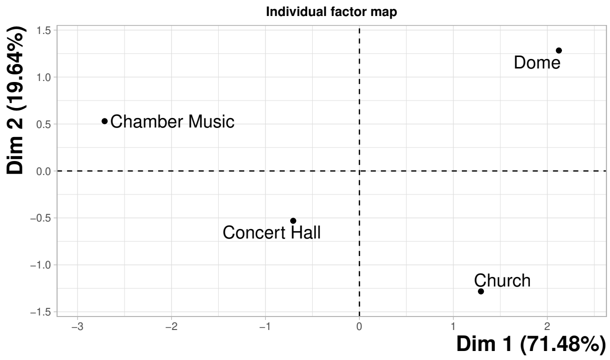

Figure 3: MFA factor map. The map visualizes the position of the stimuli on the first two components of the consensus space.

It can be seen that the stimuli are well separated in the first two principal components of the factor map. This indicates that the participants were able to clearly identify the perceptual differences between the presented stimuli. Moreover, the analysis found that the first two principal components of the consensus space jointly explain around 91% of the variance, and the entire variance of the consensus space can be explained from the first three principal components, as shown in Table 2. This shows that almost all perceptual differences between the presented stimuli can be explained from the first two dimensions of the factor map.

Table 2: Eigenvalues and variance percentages for the first three principal components of the MFA analysis.

PC	Eigenvalue	Variance (%)
1	3.51	71.48
2	0.96	19.64
3	0.44	8.88

When looking at the first dimension of the factor map, it can be observed that the stimuli are ranked according to their reverberation time (see Table 1), with *Chamber Music* and *Dome* presets situated at opposite ends of the factor map. In contrast, the second dimension does not exhibit a clear order of the stimuli.

In individual vocabulary methods like FP, each participant may use different vocabulary to describe the same sensory experience. This makes it difficult to directly compare the perceptual attributes among the participants, as semantic equivalence is not guaranteed. A clustering method is thus required to identify the common perceptual characteristics of the presented stimuli. This concept was first introduced in perceptual analysis techniques for concert halls [5], and later used in similar sensory evaluation experiments [6, 8].

4.2 Clustering of perceptual attributes

A clustering method called Agglomerative Hierarchical Clustering (AHC) was used to group the elicited perceptual attributes and reveal the hidden structure in the data. This was done based on Euclidean distances in combination with Ward's criterion [5]. Since AHC is performed separately from MFA, it applies equal weight to all attributes. Therefore, any individually elicited attribute that did not correlate well with the first two dimensions of the factor map ($|r| < 0.65$) was excluded from the clustering [8]. In this analysis, all attributes passed this threshold, and no attributes were excluded.

The dendrogram shown in Figure 4 illustrates the outcome of the AHC analysis, revealing three main clusters of elicited attributes. The perceptual attributes are categorized into two primary clusters, with the first cluster, located on the left-most side of the diagram, containing ten attributes that can be related to the characteristics of small-room acoustics, such as source presence and low-frequency modal behavior of the reproduced sound fields [24]. The second cluster further splits up into two clusters: a smaller subcluster in the middle containing four perceptual attributes that can be linked to the spectral perception of the reproduced sound fields, and a larger subcluster on the right containing nine perceptual attributes associated with diffuse or late reverberation. It was shown that these clusters can be used to identify the underlying perceptual constructs among the presented stimuli [6, 8]. The underlying constructs can be assigned to each cluster based on the common characteristics of the elicited attributes in each cluster. The constructs assigned to the clusters are *Presence* for the left-most cluster, *Spectral Content* for the middle subcluster, and *Reverberance* for the right-most subcluster.

It is important to note that the clustering of perceptual attributes is based on the similarity of attribute rankings across the stimuli rather than their semantic equivalence. This can be seen by looking at the right-most cluster of perceptual attributes. In this cluster, which contains mainly attributes associated with diffuse or late reverberation, the attribute *Dryness* is also present. This can be

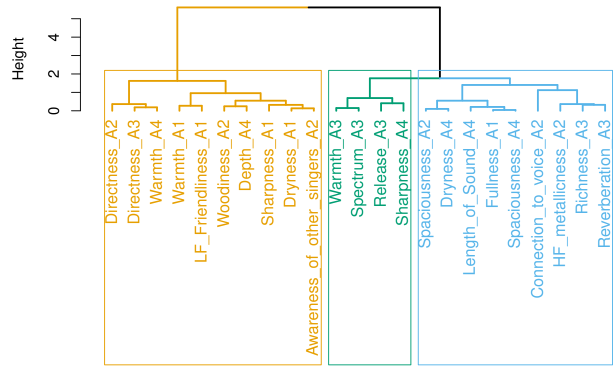

Figure 4: Dendrogram visualizing three main clusters of perceptual attributes. This is the result of Agglomerative Hierarchical Clustering based on Euclidean distances in combination with Ward's criterion. A(1-4) denotes which assessor elicited the attribute.

explained by the participant's reverse interpretation of the attribute during the ranking phase of the experiment, correlating the attribute with reverberation instead of dryness.

The clustered data containing the elicited attributes can provide a perceptual explanation for the variance within the dimensions of the factor map. By averaging the coordinates of the elicited attributes in the consensus space for each cluster, they can be projected onto the MFA consensus space. Figure 5 visualizes the factor map containing the projections, shown as vectors, of the averaged elicited attributes for each cluster. Each vector also shows the underlying perceptual construct of the cluster, allowing an interpretation of the variance in each dimension.

The first dimension of the factor map can be interpreted as relating to the overall spaciousness and reverberation of the reproduced room acoustic stimuli. Negative values on this dimension indicate a correlation with source presence and low-frequency modal behavior of the acoustic stimuli, whereas positive values are linked to the reverberance and late reverberation. This interpretation is reinforced by the arrangement of the stimuli in the first dimension of the factor map, which follows the order of reverberation time parameters, as shown in Table 1.

The second dimension of the factor map correlates best with the subcluster containing elicited attributes relating to the spectral content of the presented stimuli. This relationship is supported by the relative positions of the stimuli on the factor map, where the *Dome* and *Chamber Music* stimuli are well separated from the other two stim-

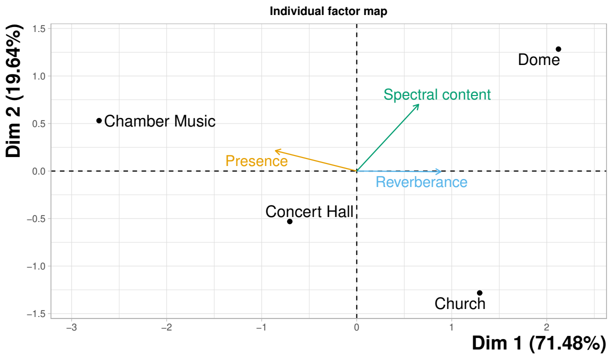

Figure 5: MFA factor map visualizing the position of the stimuli together with the perceptual constructs which explain the variance of factor map's dimensions.

uli in the second dimension. Further analysis of the magnitude responses in Figure 2 indicate that these two stimuli exhibit a higher energy in the lower frequencies compared to their high frequency energy, indicating a greater spectral imbalance in comparison with the other two stimuli.

5. CONCLUSIONS

In this paper, a pilot experiment was carried out to investigate the perceptual differences between four spatially reproduced room acoustic models in an interactive manner. The sound fields were reproduced using a commercially available real-time auralization engine, allowing for an interactive experience with minimal latency. During the experiment, four expert performers of chant music interacted with the four virtual room acoustic models. A rapid sensory evaluation technique called Flash Profiling was followed, where each assessor elicited perceptual attributes that described the differences among the presented stimuli. After this, a ranking phase took place where the assessors ranked the previously elicited attributes according to the perceived intensity of each stimulus. The results of this experiment were analyzed using Multiple Factor Analysis, in which the perceptual constructs underlying the stimuli were identified and validated against the physical properties of each acoustic model.

Our findings showed that reverberation or lack thereof, i.e., directness, and spectral content were the main perceptual constructs explaining the variance in the experimental data. This is consistent with previous studies that found that these perceptual characteristics contribute most to the performance of vocal musicians [1, 3, 10]. Moreover, to the best of the authors' knowledge, this is the first interactive perceptual experiment for auralized room acoustics that employed Flash Profiling as a rapid Sensory Evaluation Technique. It was shown that Flash Profiling is a valid method for the evaluation of interactive auralization of spatially reproduced room acoustics.

Future work could explore the impact of a broader range of acoustic room models in an interactive setting. This could involve designing room acoustic models with varying objective room acoustic parameters and evaluating their impact on perception during performances of vocal music. This could further identify the relationship between vocal music and performance spaces.

6. ACKNOWLEDGMENTS

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of FWO Large-scale research infrastructure ”The Library of Voices - Unlocking the Alamire Foundation's Music Heritage Resources Collection through Visual and Sound Technology” (I013218N) and FWO SBO Project ”The sound of music - Innovative research and valorization of plainchant through digital technology” (S005319N). The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

The authors would like to thank the expert assessors for offering their time to participate in the pilot experiment.

7. REFERENCES

[1] K. Ueno and H. Tachibana, “Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system,” *Acoust. Sci. Technol.*, vol. 24, pp. 130–138, 2003.

[2] A. Bassuet, “Acoustics of early music spaces from the 11th to 18th century: Rediscovery of the acoustical excellence of medium-sized rooms and new perspectives for modern concert hall design,” *J. Acoust. Soc. Am.*, vol. 115, no. 5, pp. 2582–2582, 2004.

[3] K. Ueno, K. Kato, and K. Kawai, “Effect of room acoustics on musicians' performance. Part I: Experimental investigation with a conceptual model,” *Acta Acust United Ac*, vol. 96, pp. 505–515, 2010.

[4] T. Lokki, H. Vertanen, A. Kuusinen, J. Pätynen, and S. Tervo, "Auditorium acoustics assessment with sensory evaluation methods," in *Proc. ISRA*, pp. 29–31, 2010.

[5] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes," *J. Acoust. Soc. Am.*, vol. 130, no. 2, pp. 835–849, 2011.

[6] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S. H. Jensen, "A rapid sensory analysis method for perceptual assessment of automotive audio," *J. Audio Eng. Soc.*, vol. 65, no. 1/2, pp. 130–146, 2017.

[7] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S. H. Jensen, "Perceptual aspects of reproduced sound in car cabin acoustics," *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 1459–1469, 2017.

[8] N. Kaplanis, S. Bech, T. Lokki, T. van Waterschoot, and S. Holdt Jensen, "Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction," *J. Acoust. Soc. Am.*, vol. 146, no. 5, pp. 3562–3576, 2019.

[9] P. Hüttenmeister and W. L. Martens, "Perceived cathedral ceiling height in a multichannel virtual acoustic rendering for gregorian chant," in *Proc. of ACOUSTICS*, vol. 1/2, pp. 587–595, 2016.

[10] T. Fischinger, K. Frieler, and J. Louhivuori, "Influence of virtual room acoustics on choir singing," *Psychomusicology: Music, Mind and Brain*, vol. 25, pp. 208–218, 2015.

[11] S. Amengual Garí, D. Eddy, M. Kob, and T. Lokki, "Real-time auralization of room acoustics for the study of live music performance," in *Jahrestagung für Akustik, DAGA*, pp. 1474–1477, 2016.

[12] V. Dairou and J.-M. Sieffermann, "A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile," *J. Food Sci.*, vol. 67, pp. 826–834, 2002.

[13] J. Delarue and J.-M. Sieffermann, "Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products," *Food Quality and Preference*, vol. 15, no. 4, pp. 383–392, 2004.

[14] Astro Spatial Tech. BV., "SARA II premium rendering engine." `https://www.astroaudio.eu/products/sara-2`. Accessed: 2023-04-3.

[15] R. J. Hughes, T. Cox, B. Shirley, and P. Power, "The room-in-room effect and its influence on perceived room size in spatial audio reproduction," in *Audio Eng. Soc. Convention 141*, 2016.

[16] A. Haeussler and S. van de Par, "Crispness, speech intelligibility, and coloration of reverberant recordings played back in another reverberant room (Room-In-Room)," *J. Acoust. Soc. Am.*, vol. 145, no. 2, pp. 931–944, 2019.

[17] K. Brandenburg, M. Schneider, A. Franck, W. Kellermann, and S. Brix, "Intelligent multichannel signal processing for future audio reproduction systems," in *Audio Eng. Soc. Conf.: 52nd Int. Conf.: Sound Field Control - Engineering and Perception*, 2013.

[18] M. M. Boone, E. N. G. Verheijen, and P. F. van Tol, "Spatial sound-field reproduction by wave-field synthesis," *J. Audio Eng. Soc*, vol. 43, no. 12, pp. 1003–1012, 1995.

[19] A. Novak, P. Lotton, and L. Simon, "Synchronized swept-sine: Theory, application, and implementation," *J. Audio Eng. Soc*, vol. 63, no. 10, pp. 786–798, 2015.

[20] International Organization for Standardization, "Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces," International Standard 3382-1, ISO, Geneva, Switzerland, 2009.

[21] International Organization for Standardization, "Acoustics - measurement of room acoustic parameters - part 2: Reverberation time in ordinary rooms," International Standard 3382-2, ISO, Geneva, Switzerland, 2008.

[22] B. Escofier and J. Pagès, "Multiple factor analysis (AFMULT package)," *Computational Statistics & Data Analysis*, vol. 18, no. 1, pp. 121–140, 1994.

[23] S. Lê, J. Josse, and F. Husson, "le2008factominer: An r package for multivariate analysis," *Journal of Statistical Software*, vol. 25, no. 1, pp. 1–18, 2008.

[24] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, "Perception of reverberation in small rooms: A literature study," in *Audio Eng. Soc. Conf.: 55th Int. Conf.: Spatial Audio*, 2014.