Exploring visual-auditory coherence in virtual environments: a perceptual study

Simon Fargeot, Mitsuko Aramaki, Richard Kronland-Martinet

Audio Engineering Society

Conference Paper

Presented at the AES 5th International Conference on
Audio for Virtual and Augmented Reality
2024 August 19–21, Redmond, WA, USA

Exploring visual-auditory coherence in virtual environments: a perceptual study

Simon Fargeot[1], Mitsuko Aramaki[1], and Richard Kronland-Martinet[1]

[1]*Aix-Marseille Univ., CNRS, PRISM, Marseille France*

Correspondence should be addressed to Simon Fargeot (fargeot@prism.cnrs.fr)

ABSTRACT

The study presented here explores the perceived visual-auditory (VA) coherence between acoustic environments and their visual representations, addressing two main questions: are we able to associate room acoustics with their own visual representations, and does the type of visual rendering affect this judgment? A virtual reality (VR) experiment was conducted with five acoustic environments auralized through a Higher Order Ambisonics (HOA) multichannel system, each presented along with three types of visual representations: 360 photos, textured point clouds, and simple 3D models, rendered through a VR headset. For each VA condition, participants were asked to answer a simple question: "Do you think the sound you are hearing was produced in the environment you are seeing? Y/N." Results indicated that congruent situations were not consistently perceived as coherent, and the type of visual rendering did not influence judgments. These findings have implications for applications in cinema, video games, architectural simulation, and virtual reality.

1 Introduction

The study of indoor spaces perception has been an important research theme in psychophysics since the second half of the 20th century. It is most often approached either through the lens of auditory perception or vision. However, as noted by Gibson in his ecological approach to visual perception [1], our perception of the environment is inherently multimodal. Since then, numerous studies have revealed interactions between auditory and visual modalities in the perceptual process [2, 3, 4]. Spatial perception is no exception to this rule. One of the most famous effects of this interaction in space is the "ventriloquist" effect [5], widely used in cinema. It describes the fusion in space of a sound source with a visual source when their content is coherent (i.e. perceived as coming from the same event and leading to a single percept), even when these two sources are spatially separated.

1.1 Visual-Auditory Perception of Rooms

The study of multimodal perception of acoustic environments, however, is a relatively new subject, facilitated by the development of virtual technologies since the late 1990s. Numerous research works have been conducted since. Maempel and Jentsch, for example, conducted a study aimed at quantifying the impact of visual and auditory modalities and visual-auditory interactions on the judgment of source-listener distance

and room size [6]. For this purpose, an experimental design including unimodal, audio, and visual conditions, as well as bimodal (VA) congruent and non-congruent conditions, was implemented. The authors concluded that the source-listener distance is primarily evaluated based on acoustic criteria, and the assessment of room size is rather guided by visual information. However, they could not conclude on the presence of an interaction between the two modalities.

An influence of visual feedback on the perception of source-listener distance has been revealed in several studies since then [7, 8, 9]. Valente et al. reported an effect of the visual environment on the perception of apparent source width (ASW) and the sense of envelopment (LEV) in reverberant environments [10]. In 2020, in a master's thesis, Greif attempted to answer the question: "Can you hear the shape of a concert hall?" through a multimodal experiment [11]. An acoustic and visual modeling of 4 concert halls with different shapes ("Shoebox", "Horseshoe", "Fan", and "Vineyard"), with equivalent geometric (volume) and acoustic (reverberation time) characteristics, was conducted. The study shows that, on average, subjects were not able to associate the different acoustics with their visual representation. However, it reveals that with learning, the association scores are slightly better, although statistically still close to chance. According to these results, humans are not auditorily sensitive to the shape of rooms.

As part of the SEACEN consortium [1], Maempel also endeavored to define a methodology for studying VA perception [13]. In this document, he notes that in the field of multimodal perception studies, there is a great diversity of experimental protocols, sometimes making studies difficult to compare with each others. He proposes a theoretical framework to better conduct such studies and presents a platform dedicated to these experiments: "the Virtual Concert Hall".

1.2 Influence of Virtual Environments on Room Perception

Efforts have been made to construct audiovisual virtual environments conducive to research on VA perception [14, 15]. However, the use of virtualization technologies is not without bias. In 2001, Swedish researchers Larsson, Västfjäll, and Kleiner showed that audiovisual virtual environments greatly differ from real conditions in terms of perceived room size and perceived source distance [7]. In a test comparing the perception of these two dimensions in four different conditions: Audio only, Audio + photograph, Audio + Virtual Environment (VE), Audio + Real Environment, the authors demonstrated that the Audio + VE condition is more comparable to the audio-only and audio + photograph conditions. These results indicate that, at that time, the visual rendering in VR of distance and room size was not convincing, and participants in the virtual condition primarily relied on acoustic characteristics to evaluate these dimensions. More recently, Maempel and Horn published an extensive study titled "Audiovisual perception of real and virtual rooms" [16]. The study's results showed a significant difference between real and virtual conditions in the evaluation of bimodal geometric criteria (perceived source distance, apparent source size, perceived room size), particularly for the "visual only" and "visual-auditory" conditions. The authors concluded that their optical virtual system was not suitable for studying these geometric dimensions. However, the use of the VA virtual device is valid for studying monomodal criteria (acoustic and visual), such as reverberation (acoustic) or room brightness (visual).

Most studies on the VA perception of acoustic environments use similar visual restitution technologies (TV, curved or U-shaped screen) [10, 9, 17, 16]. However, the dominant technology in terms of optical virtual environments is the virtual reality headset. An article from 2020 comparing results from different studies on visual distance perception in virtual and real environments reveals that recent studies show little to no distortion of perceived distance in VR compared to reality [18]. This result suggests that the performance of VR headsets in terms of geometry restitution (notably distance and depth) has significantly improved over the past ten years, making them favorable for use in studying the VA perception of acoustic environments. This observation is also shared by [19] in an article presenting an overview of VR for architectural acoustics and by [20], in a 2019 study on source localization performance in different VA conditions, both real and virtual.

[1] The SEACEN consortium (Simulation and Evaluation of Acoustical ENvironments) has brought together researchers from various German universities since 2011 around the themes of capture, simulation, reproduction, and perception of acoustic environments. Hans-Joachim Maempel is responsible for research on the VA perception of acoustic environments [12].

1.3 Motivations and Issues

Numerous studies over the past 10 years have revealed the importance of studying room perception in its multimodal aspect. Most of the time, these studies focus on the influence of visual stimulation on the perception of auditory-related dimensions (e.g., room acoustic parameters, sound source parameters). However, the overall coherence between an acoustic environment and its visual representation has been little studied, despite its major interest, particularly in the fields of video games, cinema, and architectural simulation. Moreover, the visual stimuli used in these studies are either re-projected real images (360 photos and videos) or virtual 3D room models. The amount of important information such as depth, texture, and level of detail varies greatly from one rendering to another. It thus seems important to us to understand the influence of the choice of these representations on the VA perception of the environment. Therefore, the experiment we present here is motivated by two issues:

- Are we able to associate room acoustics to their own visual representations?
- Does the type of visual rendering influence our ability to perform this VA association?

2 Methods

To answer these two questions we conducted a perceptual experiment at PRISM laboratory.

2.1 Participants

21 normal-hearing participants (14 men and 7 women) with an average age of 30 years (std: 10.1 years) took part in this experiment. The subjects' expertise in room acoustics varied and was not a factor in the analysis of collected data.

2.2 Auditory and visual stimuli

The acoustic and visual environments under study here have been selected from the "Sesames" corpus[2]. One of the strengths of this corpus is that, for each building (called indivudual subsequently in the article), visual data (360 photos at the positions of acoustic measurements and point clouds of the interior of the buildings) were collected in parallel with 4th order HOA SRIRs, measured with mh-acoustics em32 spherical microphone array, making it perfectly suited for studying VA perception.

Room selection

A selection of 5 chapels from the corpus was made based on their reverberation times (RT) and volumes (V). As illustrated in Figure 1, the 4 extreme individuals of the corpus in terms of RT and volume were chosen (Peyrl, StMaPa, Bras, Esp). A 5th environment (TvNDSalt) from the same corpus was also selected because it stood out from the others due to its complex architecture although its reverberation time and volume are comparable to that of Esp. One SRIR per chapel was selected, corresponding to a source positioned at the center of the nave and listener 5.6m from the source.

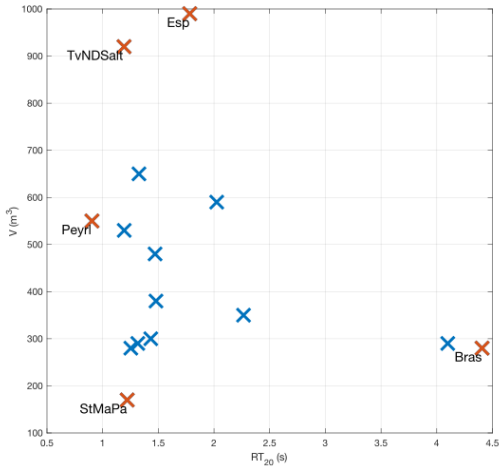

Fig. 1: Visualization of the 5 environments selected for this study (in red), based on their RT_{20} and Volume V.

The reverberation times and volumes of the 5 selected locations are given in Table 1.

Type of visual rendering

To quantify the influence of visual rendering on the perceived audio-visual coherence of acoustic environments, 3 types of renderings were chosen (*cf* Figure

[2]Sesames ANR project gathered acousticians and researchers with architectural and patrimonial expertise, aiming to characterise a corpus of 15 small and medium chapels from the country-side of southern France. A measurement campaign was conducted in these 15 buildings between 2019 and 2020, including metric measurements, 360 photos and 3D point cloud capturing of the indoor of these buildings on the one hand and spatial room impulse responses (SRIR) measurements at different positions of the buildings, on the other hand. More information on the Sesames corpus, measurement campaign and data collection can be found in [21].

Environments	RT_{20} (sec)	V (m^3)
Bras	4.41	280
Esp	1.78	990
StMaPa	1.22	170
TvNDSalt	1.19	920
Peyr	0.90	550

Table 1: Reverberation times and volumes of the 5 environments under study.

2). Each has different characteristics regarding depth, texture, and level of detail:

- **360 Photos**: captured at the measurement position of the selected impulse responses. These panoramic photos are not stereoscopic; they are rendered by spherical projection. They correctly represent texture and color information but struggle to convey the sensation of depth.
- **Point Clouds** (textured): obtained from a photogrammetric capture of the environments. They contain texture and color information (sparser than the 360 photos) but also depth information (3D representation).
- **Simple Models**: representing the studied environments in a simplified polygonal form with uniform texture. These models provide a rough representation of the geometry of the places but do not contain texture information. They are derived from the point-clouds using the PolyFit method [22].

Sound Excerpts

Two sound stimuli were chosen, consistent with the common use of the studied locations (small and medium rural chapels in the southern France region): a short speech excerpt (5 sec) followed by a few seconds of silence, and a classical guitar excerpt (3 min), both recorded in an anechoic condition.

Summary

For auditory modality, the two sound excerpts are convolved with the SRIRs of the 5 studied environments, resulting in a total of 10 different auditory conditions. For the visual modality, the 3 types of visualization presented in section 2.2 are tested (360 photos, point clouds, and simple models) for each of the 5 selected visual environments, yielding 15 different visual conditions.

In the experiment, a multimodal condition corresponds to the presentation of one of the 10 sound stimuli with one of the 15 visual stimuli. The experimental design thus obtained is a factorial design with 4 factors, with a total of 150 test trials:

- Acoustic environment (5),
- Visual environment (5),
- Type of visualization (3),
- Sound excerpt (2),

with the number of levels of each factor indicated in parentheses.

2.3 Apparatus

For this experiment, participants are placed at the center of PRISM-laboratory's multichannel spatialization setup (a sphere of 42 speakers, placed in a semi-anechoic chamber). Room acoustics are rendered in 4th-order HOA using the tools from the spat5 library for Max8. The test interface is a virtual reality interface, developed in the Unity game engine and rendered via a virtual reality headset (Oculus Quest). The experiment's progress, the presentation of visual environments, and user responses are managed in Unity while audio management is done in Max8. An OSC communication between the two softwares allows the transmission of auditory condition information from Unity to Max8.

2.4 Procedure

The 150 test conditions were divided into three sessions (one session per type of visual representation). In each session, the 50 audio-visual conditions were presented randomly, and the order of the sessions was also randomized to avoid presentation biases. At the beginning of the experiment, a learning phase of about five minutes was planned to allow the subjects to familiarize themselves with the experimental setup and the task to be performed. For each trial, participants were asked to judge the coherence between the presented acoustics and visuals. To do this, they had to answer

(a) 360 Photo **(b)** Point Cloud **(c)** Simple Model

Fig. 2: Illustration of the 3 types of renderings proposed for the experiment. Example for the "Bras" individual.

"yes" or "no" to the following question: "Do you think the sound you hear was produced in this place?". The response was given via the user interface in the virtual environment, as illustrated in Figure 3. The sound stimuli were played in a loop until the participant responded to the question. They were free to turn 360° without time constraints, before making their judgment. Each session lasted an average of 15 minutes. Between sessions, a minimum 10-minute break was imposed to limit fatigue related to the virtual reality setup. Participants were also invited to take a break or stop the experiment at any time if they wished.

Fig. 3: Capture of the user interface in the VR headset. Using a controller, the user can select the desired response to the question displayed on the screen.

2.5 Data Analysis

For each test trial, the user's response was recorded in a text file as a binary response: 1 for "yes" and 0 for "no". The responses of the 21 participants were summed for each of the 150 test conditions and stored in a 4-dimensional matrix corresponding to the four experimental factors. For each cell of the table, the positivity rate was calculated, corresponding to the percentage of positive responses relative to the number of repetitions of the considered condition.

A first analysis aimed at evaluating the level of perceived coherence between acoustic and visual environments was conducted. This analysis addressed the question, "Are we able to associate an acoustic environment with its visual representation?". To measure a statistical effect, these results were compared to the theoretical distribution of chance in a binary response test. To determine if the participants responded significantly differently from chance, a confidence interval associated with the hypothesis H_0 "Participants responded randomly" was calculated using the following formula:

$$\left[\hat{p} - z_{\alpha/2} * \sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}; \hat{p} + z_{\alpha/2} * \sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}\right]$$

with:

- $\hat{p} = 0.5$: probability of chance for a binary test,
- $z_{\alpha/2} = 1.96$: standard score for a 0.95 confidence test,
- n: sample size, corresponding here to the number of repetitions of a condition.

Positivity rate values outside this confidence interval do not respect hypothesis H_0 and were thus considered statistically different from chance ($p_{val} < 0.05$).

Furthermore, the influence of the "sound exceprt" and "type of visualization" factors was studied. To measure the effect of a factor, the responses according to the other factor were summed, resulting in a 3D table of dimensions "Acoustic Environment" $\times$ "Visual Environment" $\times$ "Studied Factor". A χ^2 independence test was then conducted between the levels of the factor of interest.

3 Results and Discussions

Results and discussions are addressed in two distinct sections. The first section describes the perceived audio-visual coherence qualitatively. The second section examines the effects of "sound excerpt" and "type of visualization".

3.1 Study of Perceived Coherence between Acoustic and Visual Environments

Perceived audio-visual coherence is quantified by the positivity rate τ_+ (ratio of positive responses to total trials in the condition). Table 2 represents the audio-visual coherence for each acoustic/visual pair, without considering sound stimulus and visualization type.

In 8 out of 25 cases, the obtained responses are statistically close to chance. Conversely, in 68% of cases (17 conditions out of 25), the positivity rates are statistically different from those expected by random guessing. Specifically, among these 17 conditions, 8 were judged as strongly coherent (in green) and 9 were judged as incoherent (in red). The minimum and maximum coherence values are 7.94% for the "Bras-StMaPa" audio-visual condition and 70.63% for the "Bras-Esp" condition.

Discussions

By observing the positivity rates obtained in the diagonal of table 2, corresponding to congruent audio-visual conditions, it is possible to address the question "Are we capable of associating an acoustic environment with its visual representation?" It is noted that none of the congruent conditions were judged by participants as coherent. In some cases (such as Bras and TvNDSalt), they were even judged as strongly incoherent, leading to the conclusion that even under favorable test conditions (3D audio-visual immersion), participants were unable to associate different acoustics with their visual representation. These observations are comparable to those from a multimodal study recently conducted at TU Berlin, titled "Can you hear the shape of concert halls? An audiovisual test in simulated 3D environments" [11]. However, this result does not imply that participants were unable to judge the relevance between acoustic and visual environments. On the contrary, in 68% of cases, the judgment of VA coherence is significant, indicating that overall, subjects relied on specific acoustic and visual attributes to evaluate this coherence.

It is also noteworthy that the positivity rate does not exceed 70.63%. Conversely, responses were more unanimous in judging the lack of coherence between acoustic and visual environments. Indeed, 6 out of the 9 judged non-coherent responses have a positivity rate below 20%, with the minimum positivity rate obtained being 7.94%. There are multiple explanations for these observations. Even though the locations were chosen to represent the diversity of the corpus from which they are extracted - a corpus of small to medium-sized chapels - the group formed is relatively heterogeneous and not representative of all environments encountered in everyday life. It can be noted, in particular from Figure 1, that the largest environments are relatively matte, and the most reverberant environment is of small volume. However, the corpus does not include large-volume environments with long reverberation times. Furthermore, the virtual listening and visualization conditions may have played a role in capping the positivity rate at 70%. It can be assumed that the question of VA coherence is closely linked to that of realism. However, it has been shown in [23] that virtual listening conditions based on HOA auralization methods can present several biases compared to real listening conditions. Moreover, the visualizations types proposed here all have certain distortions compared to reality. It should also be noted that the source is not represented in the visual interface. This absence of a visual reference was noted by some subjects and may have had a negative impact on the sensation of realism during the experiment. Lastly, it seems plausible that the heterogeneous level of expertise among participants, in terms of room acoustics, also influenced these results.

3.2 Influences of Sound Excerpt and Type of Visualization on Visual-Auditory Coherence

To measure the influence of sound stimulus and visual rendering type, a χ^2 independence test was performed

		V				
		Bras	Esp	StMaPa	Peyr	TvNDSalt
A	Bras	14,29%	70,63%	7,94%	23,81%	50,79%
	Esp	46,03%	53,97%	22,22%	52,38%	59,52%
	StMaPa	61,11%	37,30%	45,24%	61,90%	44,44%
	Peyr	62,70%	19,05%	64,29%	42,06%	16,67%
	TvNDSalt	59,52%	12,70%	64,29%	54,76%	19,84%

Table 2: Perceived audio-visual coherence (positivity rate τ_+) for pairs of acoustic (A) and visual (V) environments. Cells in color indicate results significantly different from chance. In red: conditions judged non-congruent, in green: conditions judged congruent. 95% confidence interval: $[0.41; 0.59]$.

for these two factors. Table 3 presents the results of these tests.

Factor	DF	χ^2	p
Sound Stimulus	24	35.6526	0.056
Rendering Type	48	46.7451	0.5243

Table 3: χ^2 independence test results for "Sound Stimulus" and "Rendering Type" factors.

For the "Rendering Type" factor, the p-value $p = 0.5243$ indicates no effect of visual rendering type. Despite being very different from each others, the three types of visualization (360 photo, point cloud, simple model) led to similar results. This finding indicates that the poor representation of volume in the case of 360 photos and the lack of detail and texture information in the case of simple models did not have a notable influence on the judgment of coherence. However, it is possible that for 360 photos and point clouds, containing texture information, participants perceived the texture of a poorly absorbing material. The choice of a plain, matte gray texture for simple models may have also evoked a material of the same type. The question of the influence of texture warrants further investigation by conducting an experiment using simple models with several texture conditions tested. Regarding the sound stimulus, the p-value $p = 0.056$ also does not allow for concluding a significant effect of the stimulus.

4 Summary

This experiment allowed us to explore the visual-auditory perception of virtual environments. Participants were unable to associate the acoustic characteristics of studied environments with their visual representation, consistent with the findings of [11]. Neither the type of visual rendering (360 photos, point clouds, and simple models) nor the sound stimulus (guitar excerpt, speech excerpt) statistically influenced participant responses. Although instructive, this study is primarily prospective, and the results obtained encourage further exploration. Initial findings suggest that visual-auditory coherence is evaluated based on specific acoustic and visual cues. At first glance, for a VA condition to be judged as coherent, a certain adequacy between reverberation time and perceived volume appears necessary. Future work will focus on developing a perceptual model of VA coherence using objective acoustic and visual data.

5 Acknowledgement

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-18-CE38-0009-01 (project Sesames: `http://anr-sesames.map.cnrs.fr/`). We also extend our gratitude to our colleagues from the MAP-CNRS laboratory, Jean-Yves Blaise, Iwona Dudek, Anthony Pamart, and Laurent Bergerot, for providing us with the visual data used in this experiment (360 photos and point clouds).

References

[1] Gibson, J. J., *The Ecological Approach to Visual Perception: Classic Edition*, Psychology Press, 1979.

[2] Regan, D. and Spekreijse, H., "Auditory—Visual Interactions and the Correspondence between Perceived Auditory Space and Perceived Visual Space," *Perception*, 6(2), pp. 133–138, 1977, ISSN 0301-0066, 1468-4233, doi:10.1068/p060133.

[3] Macdonald, J. and McGurk, H., "Visual Influences on Speech Perception Processes," *Perception & Psychophysics*, 24(3), pp. 253–257, 1978, ISSN 0031-5117, 1532-5962, doi:10.3758/BF03206096.

[4] Beerends, J. G. and De Caluwe, F. E., "The Influence of Video Quality on Perceived Audio Quality and Vice Versa," *Journal of the Audio Engineering Society*, 47(5), pp. 355–362, 1999.

[5] Thurlow, W. R. and Jack, C. E., "Certain Determinants of the "Ventriloquism Effect"," *Perceptual and motor skills*, 36(3_suppl), pp. 1171–1184, 1973.

[6] Maempel, H.-J. and Jentsch, M., "Auditory and Visual Contribution to Egocentric Distance and Room Size Perception," *Building Acoustics*, 20(4), pp. 383–401, 2013.

[7] Larsson, P., Västfjäll, D., and Kleiner, M., "Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments," in *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland*, pp. 245–249, 2001.

[8] Gorzel, M., Corrigan, D., Kearney, G., Squires, J., and Boland, F., "Distance Perception in Virtual Audio-Visual Environments," in *25th UK Conference of the Audio Engineering Society: Spatial Audio In Today's 3D World (2012)*, pp. 1–8, 2012.

[9] Postma, B. N. J. and Katz, B. F. G., "The Influence of Visual Distance on the Room-Acoustic Experience of Auralizations," *The Journal of the Acoustical Society of America*, 142(5), pp. 3035–3046, 2017, ISSN 0001-4966, doi:10.1121/1.5009554.

[10] Valente, D. L. and Braasch, J., "Subjective Scaling of Spatial Room Acoustic Parameters Influenced by Visual Environmental Cues," *The Journal of the Acoustical Society of America*, 128(4), pp. 1952–1964, 2010.

[11] Greif, J. and Ackermann, D., *Can you hear the shape of a concert hall? An audiovisual test in simulated 3D environments*, Master's thesis, Institut für Sprache und Kommunikation, Technische Universitat Berlin, Berlin, Germany, 2020.

[12] Lindau, A., Maempel, H.-J., and Horn, M., "SEACEN : P9 - Audio-visual Perception of Acoustical Environments," https://www.seacen.tu-berlin.de/subprojects/p9/parameter/en/, 2014.

[13] Maempel, H.-J., "Apples and Oranges: A Methodological Framework for Basic Research into Audiovisual Perception," 2017, retrieved from http://dx.doi.org/10.14279/depositonce-6424.

[14] Seeber, B. U., Kerber, S., and Hafter, E. R., "A System to Simulate and Reproduce Audio–Visual Environments for Spatial Hearing Research," *Hearing research*, 260(1-2), pp. 1–10, 2010.

[15] Maempel, H.-J., "The virtual concert hall—A research tool for the experimental investigation of audiovisual room perception," *International Journal on Stereo & Immersive Media*, 1(1), 2017.

[16] Maempel, H.-J. and Horn, M., "Audiovisual Perception of Real and Virtual Rooms," *Journal of Virtual Reality and Broadcasting*, 14(5), 2018.

[17] Postma, B. N. J. and Katz, B. F. G., "Influence of Visual Rendering on the Acoustic Judgements of a Theater Auralization," in *173rd Meeting of Acoustical Society of America and 8th Forum Acusticum*, p. 015008, Boston, Massachusetts, 2017, doi:10.1121/2.0000575.

[18] Feldstein, I. T., Kölsch, F. M., and Konrad, R., "Egocentric Distance Perception: A Comparative Study Investigating Differences Between Real and Virtual Environments," *Perception*, 49(9), pp. 940–967, 2020.

[19] Vorländer, M., Schröder, D., Pelzer, S., and Wefers, F., "Virtual Reality for Architectural Acoustics," *Journal of Building Performance Simulation*, 8(1), pp. 15–25, 2015.

[20] Ahrens, A., Lund, K. D., Marschall, M., and Dau, T., "Sound Source Localization with Varying Amount of Visual Information in Virtual Reality," *PLOS ONE*, 14(3), p. e0214603, 2019, ISSN 1932-6203, doi:10.1371/journal.pone.0214603.

[21] Blaise, J.-Y., Dudek, I., Pamart, A., Bergerot, L., Vidal, A., Fargeot, S., Aramaki, M., Ystad, S., and Kronland-Martinet, R., "Acquisition & integration of spatial and acoustic features: A workflow tailored to small-scale heritage architecture," *ACTA IMEKO*, 11(2 (2022)), pp. 1–14, 2022.

[22] Nan, L. and Wonka, P., "Polyfit: Polygonal Surface Reconstruction from Point Clouds," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2353–2361, 2017.

[23] Fargeot, S., Vidal, A., Aramaki, M., and Kronland-Martinet, R., "Perceptual evaluation of an ambisonic auralization system of measured 3D acoustics," *Acta Acustica*, 7, p. 56, 2023.