forum acusticum 2026
8-12 September · Graz, Austria

A PERCEPTUAL COMPARISON OF MODAL SOUND FIELDS IN REAL AND VIRTUAL ENVIRONMENTS

Thomas M. Robotham, Albert G. Prinn, Emanuël A. P. Habets

International Audio Laboratories Erlangen (A joint institution of the Friedrich-Alexander-University and Fraunhofer IIS), Am Wolfsmantel 33, Erlangen, 91058, Germany

Abstract

Modal sound fields arise from wave interference in enclosed spaces and dominate the low-frequency acoustic response of a room. Capturing this behavior in virtual acoustic environments requires numerical wave methods, which remain computationally expensive for real-time interactive applications. Geometrical methods achieve this necessary efficiency but may not incorporate accurate modal information. This study examines the perceptual impact of omitting a modal sound field in an interactive virtual environment (IVE) through a loudness-matching task. Participants equalized the perceived loudness of a sound field at positions throughout an interactive real environment (IRE) and an IVE using controlled stimuli. Relative level differences of stimuli achieved over related positions in the IRE were analyzed and compared to the same evaluation in the IVE, which does not support a modal field. The matched levels throughout the different position and frequency combinations demonstrate the presence of perceived modal-field in the IRE, while the IVE does not produce comparable trends. Furthermore, listeners were markedly less consistent in the IVE than in the IRE. These results suggest that geometrical method-based auralization does not convey the perceived spatial structure of a modal sound field.

Keywords: *virtual acoustics, perception, room modes*

1. Introduction

Virtual reality technology offers a promising setting to conduct ecologically valid research in controlled interactive virtual environments (IVEs). To obtain results in an IVE that reflect the same experiment performed in an equivalent interactive *real* environment (IRE), accurate audio rendering is a critical component.

Corresponding author: *thomas.robotham@fau.de.*

However, many challenges remain in delivering an authentic virtual representation of an IRE sound field that can meet real-time rendering requirements [1]. Despite recent advances suggesting the applicability of virtual acoustics for hearing research [2], no literature exists, to the authors' knowledge, that explicitly investigates the perception of modal sound fields in virtually simulated spaces compared to real rooms.
The physical and perceptual effects of room modes have been a long-standing topic of research in the field of loudspeaker reproduction in critical listening spaces [3]. However, the perception of room modes is often overlooked when evaluating virtual acoustics. For an IVE with six-degrees-of-freedom interactivity, the perceptual effect of salient room modes only becomes more prevalent given the fluctuations in perceived loudness [4] when moving through the environment. Hence, as we gravitate toward creating acoustically authentic IVEs, there is a need to understand the perceived differences when reproducing modal fields.
This work presents the results from an initial study investigating the impact of room modes on the spatially dependent perceived loudness in real and simulated acoustic fields. A novel loudness-matching task was performed in an IRE and an IVE replica, requiring subjects to equalize the perceived loudness of a test stimulus against that of a reference stimulus. When repeated at different spatial positions, the results present a clear difference between the acoustic environments.

2. Background

2.1. Room Modes

Room modes are the result of the constructive and destructive interference of sound waves that undergo multiple reflections from the boundaries of a room. At low frequencies, the sparsity of room modes in the frequency domain results in an uneven distribution of sound pressure in the room, often referred to as the modal region. As the frequency increases, so too does the number of modes and the modal overlap. At high frequencies, a high number of room modes are present, and the sound field is best described statistically. The transition between modal and statistical regions is qualitatively identified by the Schroeder frequency [5]. The perceptual evaluation of room modes has been studied primarily in the context of small-room acoustics [3]. Fazenda et al. [6] measured perceptual thresholds for modal interference as a function of modal decay time using both artificial and natural music stimuli. Their results showed significantly higher thresholds for music than for pure tones. Jun et al. [7] examined the audibility of modal damping differences in small rehearsal rooms. Their results showed that transient signals (e.g., a kick drum) reveal modal decay differences that broadband noise stimuli do not, and that these differences manifest in a wide range of perceived timbral characteristics.

2.2. Virtual Acoustics

Geometrical methods, such as the image source and ray-tracing methods, primarily assume that sound propagates as particles along specific ray paths. See [8] for an overview of geometric acoustic methods. This assumption precludes wave interaction effects, making the models suitable for simulating sound at frequencies above the Schroeder frequency, but, in principle, unsuited to the modal region below it. Nonetheless, their efficiency has made them the standard choice in interactive virtual acoustics across the full audible spectrum. Comparisons of established geometrical simulators note that results below the Schroeder frequency may be inaccurate for such methods [9].

2.3. Real vs. Virtual Acoustic Fields

Driven by the goal of perceptually authentic rendering, more studies have emerged directly evaluating the perceptual differences between IVEs and IREs, addressing the physical and methodological challenges. Blau et al. performed a perceptual comparison between IVE and IRE settings by allowing participants to switch between headphone and loudspeaker reproduction *in-situ* [10]. Stärz et al. performed a similar quality evaluation-style test using a multi-stimulus rating for five attributes [11]. The conditions included stimuli convolved with simulated binaural room impulse responses, in addition to a real loudspeaker playback as a hidden reference. In both studies, using speech as the stimulus, subjects' quality ratings suggest virtual acoustics methods can yield convincing perceptual results with respect to an IRE.

To evaluate perceptual differences across several IREs, Fargeot et al. performed a localization test in three real rooms and in their IVE counterparts [12]. To render the virtual sources, spatial room impulse responses were taken at the real listening positions, convolved with the raw test stimuli (speech, guitar, and noise), and played back over loudspeakers using Ambisonics. Their results showed that distance perception for far-field simulated sources is consistent with that of a real source, but localization blur increases due to the nature of Ambisonics reproduction. While these studies advance our understanding of perceived virtual acoustics against their real-world counterpart, there is little emphasis on the perceived reproduction of a modal field.

3. Experimental Design

3.1. Loudness-matching Task

To compare the perception of modal sound fields between an IRE and an IVE representation of a real room, we employed a fixed-reference loudness-matching task [13] to test the following hypotheses:

- $\mathcal{H}_1$ Subjects provide matched levels consistent with their own benchmark of perceptually aligned loudness, in both the IVE and IRE.
- $\mathcal{H}_2$ Room modes in an IRE cause spatially- and frequency-dependent changes in relative levels of a loudness-matching task.
- $\mathcal{H}_3$ An IVE with an acoustic environment modeled using geometrical acoustics does not produce trends comparable to $\mathcal{H}_2$.

As the perception of room modes is spatially and frequency-dependent, the task's methodology was adapted to address the challenges of modal excitation paired with perceptual comparisons.

A fixed-reference loudness-matching task involves matching the perceived loudness of a test stimulus to that of a reference stimulus. Participants compare levels by switching playback between the two stimuli at will [13]. For this study, loudness matching raises two challenges regarding i) the frequency response of the reproduction system and ii) the spectral content of both stimuli.

Concerning the first, a direct loudness comparison between real and virtual acoustic conditions conflates the perceptual effect of room modes with differing spectral transfer functions of the two reproduction systems. For example, head-related transfer functions (HRTFs) and headphones used in binaural rendering would spectrally color the stimuli differently from free-field loudspeaker reproduction. In such cases, an absolute loudness comparison between conditions becomes less suitable.

Second, stimuli with similar spectral profiles are typically preferred for conventional loudness-matching tasks. This mitigates the frequency-dependent equal loudness contour of the human auditory system [14]. However, at a given position within the same IRE acoustic field, wave interference will act equally on both stimuli. Consequently, no differential loudness effect attributable to a modal field can be observed.

This study adapts the fixed-reference loudness-matching task to evaluate the perception of room modes through *relative* differences. Specifically, at each receiver position, a reference stimulus is presented at a fixed playback level, calibrated to achieve a consistent target sound pressure level (SPL) at each position. Participants must then adjust the playback level of a test stimulus to the same perceived loudness as that of the reference stimulus. The test stimuli contain frequency content that targets specific modes, while the reference stimulus contains higher-frequency content, thereby avoiding frequency-dependent overlap in the modal region of interest.

3.2. IRE and IVE Room Description

A classroom at the International Audio Laboratories Erlangen was selected to serve as the room for the IRE and an IVE replica. The selection was based on: i) the ability to model the room's geometry accurately, ii) the room's physical construction properties and geometry permitting the presence of room modes, and iii) the room's dimensions mean it may be considered a 'small room'. Specifically, the x, y, z dimensions are approximately 4.962 m wide, 2.826 m tall, and 9.846 m long, respectively, with a total volume of $\approx 138.1\,\mathrm{m}^3$. Consequently, several room modes can be audibly induced at an appropriate frequency.

For the IVE, an instance of the real room was replicated using Blender[1] 3D modeling software, and realized in virtual reality using Unity[2]. The final IVE is shown in Fig. 1. The Meta XR SDK[3] was employed as the audio rendering engine that utilizes a geometrical acoustic ray-tracing approach. Acoustic geometry was assigned to all 3D assets in the room, and subsequently prescribed acoustic materials appropriate for the IRE objects (e.g., glass for windows, plastered brick for walls, carpet for floor, etc.).

3.3. Test Stimuli and Spatial Positioning

Artificial test tones were chosen as stimuli to excite the modal sound field at specific frequencies. This was done to highlight real-life issues with resonant frequencies in small rooms and to gain a precise depiction of the spatial dependency on modal perception in IREs, in contrast to IVEs modeled using current geometrical acoustic solutions.

Three tones were chosen at 51, 76, and 90 Hz (f_{51}, f_{76}, and f_{90}, respectively) based on their excitation of modes throughout the room, and validated through *in-situ* listening sessions. To capture the spatial dependency of different frequency-dependent targeted-node locations, three broader locations (L_1, L_2, L_3), each with three offsets (O_a, O_b, O_c), were included, resulting in nine positions. Positions $\mathsf{L}_1\mathsf{O}_a$, $\mathsf{L}_2\mathsf{O}_a$, and $\mathsf{L}_3\mathsf{O}_a$, were situated at the location of low SPL generated by f_{51}, f_{76}, and f_{90} frequencies, respectively, illustrated by Fig. 1[4]. Positions $\mathsf{L}_1\mathsf{O}_b$ and $\mathsf{L}_1\mathsf{O}_c$ were then spatially offset from $\mathsf{L}_1\mathsf{O}_a$ to underscore local transition out from the targeted node. Similar transitional offsets were chosen for $\mathsf{L}_2\mathsf{O}_b$ and $\mathsf{L}_2\mathsf{O}_c$ from $\mathsf{L}_2\mathsf{O}_a$, and $\mathsf{L}_3\mathsf{O}_b$ and $\mathsf{L}_3\mathsf{O}_c$ from $\mathsf{L}_3\mathsf{O}_a$. For all nine location-offset positions, all three test frequencies were evaluated, resulting in 27 test items in total and enabling a granular linear mixed-effects model analysis at a subject-frequency-position level.

3.4. Reference Stimulus and Calibration

The reference stimulus, against which subjects matched the perceived loudness of the test stimuli, consisted of band-passed pink noise with a center frequency of 800 Hz and a $Q = 0.4$, resulting in a -3 dB bandwidth of 2 kHz. This broad filter reduces the spectral magnitude by approximately 9 dB at the Schroeder frequency (approximately 120 Hz in this room), placing the majority of the reference signal's spectral content above it. For both IRE and IVE auralizations, the reference stimulus was calibrated using a Cortex M1 head and torso simulator (HATS) to achieve an average left and right response of 74 dB(A) SPL. In doing so, we minimize perceived loudness fluctuations arising from the level-dependent shape of the equal-loudness contour.

For the IRE, the HATS was physically placed at every location-offset position to measure the required compensation gains for the reference signal. A Brüel & Kjær omnidirectional loudspeaker was employed as the source to reduce perceived loudness changes from off-axis coloration. For the IVE, Beyerdynamic DT 770 Pro headphones were placed over the ears of the HATS, and the receiver in Unity was placed at each matching virtual location-offset position. The virtual source was configured as a point source with no directivity. In both calibration procedures, the HATS was diffuse-field equalized.

4. Experimental Procedure

Subjects were given a verbal description of the task, requiring them to listen to the pink noise signal, declared as the reference, and to increase the level of a tone until the perceived loudness matched that of the reference. The initial gain of the test signal was set to -40 dBFS, which could then be adjusted using a USB rotary dial ($-\Delta° =$ quieter, $+\Delta° =$ louder). Pressing the rotary dial confirmed the level and stored the result. Auditory feedback was given to the subjects via a short fade-out/-in transition of the reference signal, in addition to the gain of the test tone being reset.

For the IRE setting, the location-offset positions were marked and labeled on the floor with bright tape. Subjects performed the experiment seated on a wheeled stool (cf. Fig. 1), allowing them to easily move between locations. For the IVE setting, subjects performed the experiment in a separate room, experiencing the IVE over the Meta Quest 3 head-mounted display for visuals, and the same headphones used for calibration for the binauralized audio. Instead of the subject physically moving, the virtual position was automatically switched once a response was confirmed. For both IRE and IVE auralizations, a short training stage was conducted, enabling subjects to understand the controls, practice the loudness-matching task, and ask any clarifying questions.

Overall, $N = 20$ subjects participated in the experi-

[1] blender.org

[2] unity.com

[3] developers.meta.com/horizon/documentation/unity

[4] Note that while these positions intentionally target one frequency, some positions overlap closely with more than one node, e.g., $\mathsf{L}_2\mathsf{O}_a \rightarrow f_{76}$ and f_{90}

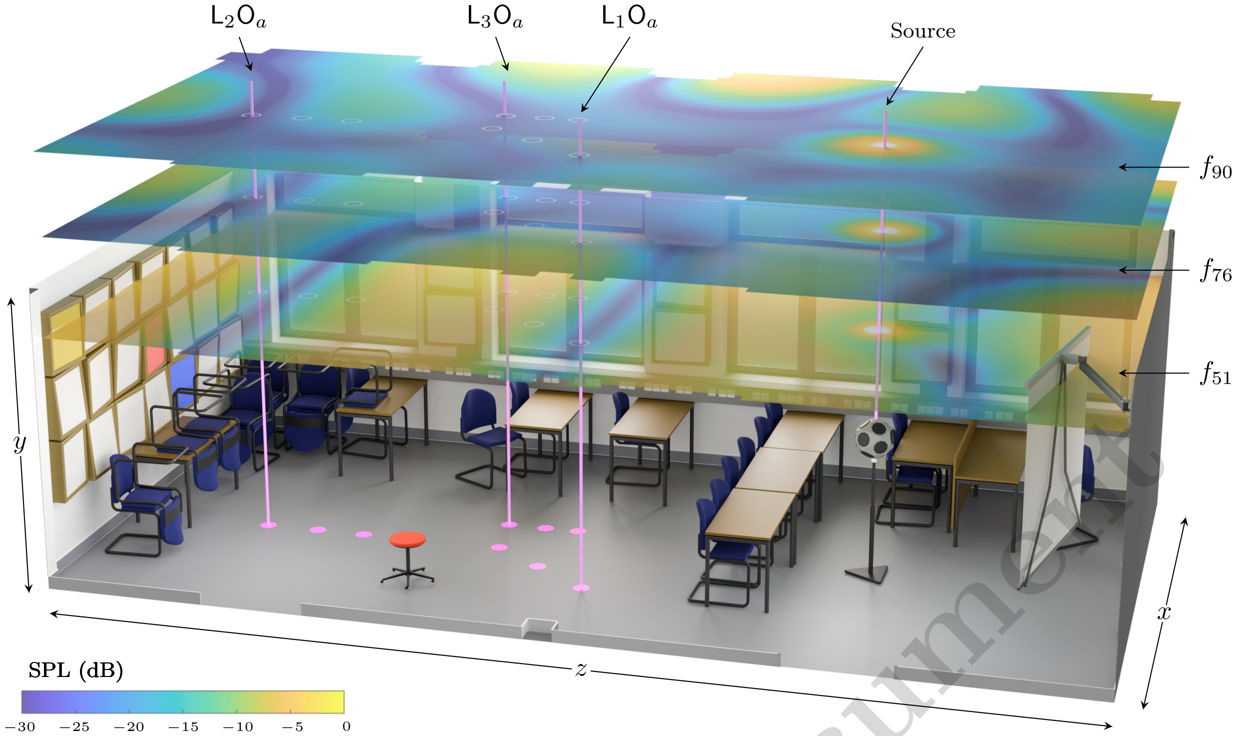

Figure 1: IVE replica of the IRE environment. Nine evaluation positions are illustrated by the solid (pink) circles, comprising three locations ($\mathsf{L} \in \{1, 2, 3\}$), each containing three offsets ($\mathsf{O} \in \{a, b, c\}$). For each location, O_a is spatially targeted to lie at a low-pressure node for a particular frequency. Specifically, $\mathsf{L}_1\mathsf{O}_a \rightarrow f_{51}$, $\mathsf{L}_2\mathsf{O}_a \rightarrow f_{76}$, and $\mathsf{L}_3\mathsf{O}_a \rightarrow f_{90}$. Three SPL heatmaps (generated using the finite element method) depict a modeled estimation of the IRE's acoustic field when excited by the three frequencies from the indicated omnidirectional source position, at a listener position height of 1.2 m (vertically offset to improve readability). *Note*: The acoustic SPL estimates are based on rudimentary geometry and global impedance values for surfaces, resulting in a slight mismatch between the listening positions and node locations.

ment, split evenly between the IRE and IVE reproductions. None of the subjects were screened for listening experience, but all reported normal hearing. The total test time was under 10 minutes.

5. Results

5.1. Response Consistency

Statistical evaluation of $\mathcal{H}_2$ and $\mathcal{H}_3$ depends on the consistency of subjects' responses within the IRE and IVE. Therefore, we first assess $\mathcal{H}_1$ to analyze the within-subject standard deviation and between-subject bias of the raw matched levels, shown in Fig. 2. In the IRE, listeners were highly self-consistent, resulting in a within-subject SD = 0.44 dB and between-subject SD = 0.85 dB. In the IVE, within-subject SD = 2.09 dB and between-subject SD = 5.02 dB. Results from a Levene's test report that variances between groups were not equal ($p < 0.001$). Hence, although subjects for the IRE were highly consistent, the same was not true for the IVE, and we reject $\mathcal{H}_1$.

5.2. Modal Effect

The results for subjects' matched levels are shown in Fig. 3, split into two rows for the IRE and IVE acoustic fields, and three columns for the three locations. For each panel, offsets O_a, O_b, O_c are shown along the x-axis, and the mean and 95% bootstrapped confidence intervals of the relative matched level (in dB) along the y-axis. Matched levels are normalized per subject to the O_c offset at each location. Hence, differences across offsets depict the relative differences by each subject, thereby avoiding between-subject variations in the absolute perceptually aligned level.

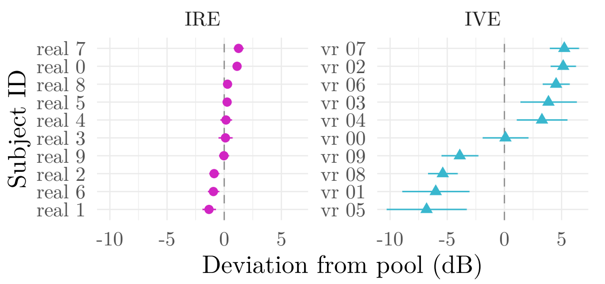

Figure 2: Mean and standard deviation of each subject's individual ratings, and deviation from the pool mean (dashed line), for both IRE and IVE.

Given that the residual variance differs strongly between IRE and IVE, testing of $\mathcal{H}_2$ and $\mathcal{H}_3$ requires analyzing the results independently without direct comparison between environments. Consequently, a

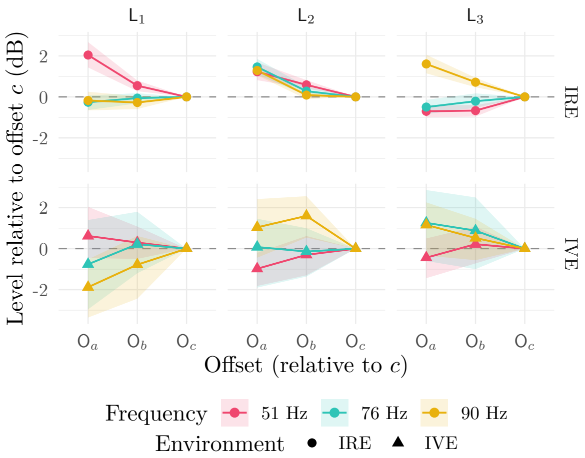

Figure 3: Mean and 95% bootstrapped confidence intervals of the relative matched level across L_1–L_3. Differences for offsets O_a and O_b are normalized to offset O_c.

linear mixed-effects model [15], given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{1}$$

was fitted separately to the data collected in the IRE and IVE using restricted maximum likelihood estimation. Here, $\mathbf{Y}$ is the raw response variable (the matched level, in dB). The fixed-effects design matrix $\mathbf{X}$ encodes the location L, frequency F, and offset O together with all of their interactions, and $\boldsymbol{\beta}$ is the corresponding vector of fixed-effect coefficients. The random-effects design matrix $\mathbf{Z}$ is a by-subject indicator, and $\mathbf{u}$ is the vector of by-subject random intercepts, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2\mathbf{I}_j)$; the residual errors are $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_k)$. The identity matrices $\mathbf{I}_j$ and $\mathbf{I}_k$ encode two standard assumptions: i) every subject intercept is equally variable and independently drawn (between-subject variance σ_u^2), and ii) every residual is equally variable and independent (within-subject variance σ^2)[5].

Omnibus F-tests[6] for the IRE results showed a significant interaction of $\mathsf{L} \times \mathsf{F} \times \mathsf{O}$ ($F_{(8,234)} = 18.3$, $p < 0.001$). This indicates the matched level was not uniform but depended jointly on location, frequency, and offset, providing support for $\mathcal{H}_2$. For each location L and frequency F, post hoc estimated marginal means were used to express the modal contrast $\mathsf{O}_a - \frac{1}{2}(\mathsf{O}_b + \mathsf{O}_c)$. The results reflect the data in Fig. 3, showing the design-targeted node positions required a significantly higher matched level than their neighbors, resulting in estimated modal contrasts of $+1.8$ dB for f_{51} at L_1 ($p < 0.001$, $d = 4.09$) and $+1.3$ dB for f_{90} at L_3 ($p < 0.001$, $d = 2.95$)[7].

[5] Given the rejection of $\mathcal{H}_1$, these assumptions are the primary reason why the model must be fitted independently for the IRE and IVE environments.

[6] Type III, with Satterthwaite-approximated degrees of freedom calculated using the lmerTest R package.

[7] Both reportings of Cohen's d are derived from $d = \hat{\delta}/\hat{\sigma}$, where $\hat{\delta}$ is the estimated modal contrast and $\hat{\sigma} = 0.44$ dB is the within-subject SD established in $\mathcal{H}_1$

Location L_2 exhibited a broadband elevation across all three tones, consistent with its placement near an overlap of modes, as seen in Fig. 1. For the IVE matched levels, the same omnibus F-tests showed no significant interaction effect $\mathsf{L} \times \mathsf{F} \times \mathsf{O}$ ($F_{(8,234)} = 1.05$, $p = 0.40$). Furthermore, the same post hoc estimated marginal means yielded no positive modal contrast for a given location L and frequency F combination. These results are reflected in Fig. 3, which depicts no clear distinction between matched level adjustments for a given frequency. Hence, the same trends seen in the IRE are not reproduced in the IVE, providing support for $\mathcal{H}_3$.

6. Discussion

The location- and frequency-specific trends observed in Fig. 3 for the IRE confirm that the loudness-matching task is sensitive to the modal field, providing a reasonable baseline for comparison for the IVE. The absence of the same trends in the IVE is expected, in principle, given that the geometrical acoustics approach does not reproduce wave interference. These results highlight perceptual limitations of these virtual acoustics approaches if the goal is to authentically reproduce a room's modal field. However, two aspects of the results warrant further explanation: i) the difference in response variance between the IRE and IVE, and ii) the magnitude of gain adjustment needed for a perceptually matched level.

A smooth, mode-free field should, theoretically, make loudness matching easier and yield more consistent responses. Instead, within- and between-subject variability was considerably greater for the IVE setting. Between-subject variation could potentially be explained by differences in the frequency-magnitude response of subjects' individual HRTFs vs. the generic HRTF used for binauralization. However, this does not explain larger within-subject variability, reflected by the larger error bars in Fig. 2. Instead, subtle acoustic or even haptic cues from a loudspeaker driven to high gain (present in node locations) may cause harmonics or vibrations that provide subjects with subtle cues that 'landmark' a certain level.

Given that the same energy is not required to reach the same SPL over headphones, and that the physical differences of the headphone transducers are different from those of the omnidirectional loudspeaker, these cues will also be different (or even absent). Such a 'landmark', that provides additional information beyond the SPL of a single tone, may also explain the difference between the theoretical SPL differences over offset locations (> 20 dB) seen in Fig. 1, and the level differences seen in Fig. 3 at the same locations in the IRE. However, while such cues would shape the sensitivity of the matched-level differences, they do not account for the location- and frequency-specific effect, which remains attributable to the modal field.

The results of this study are also bound by the limited sample size of $n = 10$ per environment. Further investigation into the perception of modal fields using a similar methodology should employ a larger sample size in a within-subject test design to increase the statistical power. Employing individualized HRTFs combined with audiogram screenings and impulse response measurements of the listening position would minimize potential confounding factors and provide more avenues for rigorous analysis.

7. Conclusion

This article reports on a novel study investigating the perceptual comparison of a modal sound field between a real room (IRE), and its geometrically modeled virtual replica (IVE), using an adapted fixed-reference loudness-matching task. The results revealed a spatially- and frequency-dependent modal structure in a classroom IRE up to 1.8 dB compared to neighboring positions, which was not reproduced in the IVE. These findings show that the geometrical method-based auralization tool used in this work, despite its efficiency, does not perceptually reproduce the spatially dependent modal field that governs low-frequency loudness in a real room. The IVE also elicited notably less consistent responses from subjects, posing further challenges for the perceptual comparison of modal fields between real and virtual settings.

8. Acknowledgments

The authors wish to thank Thomas Mayenfels for his technical assistance. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the QoEVAVE and EVOLVE-QoE projects (number 444832250) as part of the DFG Priority Program SPP 2236.

References

[1] A. G. Prinn, "A review of finite element methods for room acoustics," *Acoustics*, vol. 5, no. 2, pp. 367–395, 2023. DOI: 10.3390/acoustics5020022.

[2] H. Mali and N. K. Venkateswaran, "Auditory spatial assessment in real and virtual environments: A scoping review of current evidence and future directions," *The Egyptian J. Otolaryngology*, vol. 42, no. 1, p. 74, 2026. DOI: 10.1186/s43163-026-01045-z.

[3] F. Toole, *Sound reproduction: Loudspeakers and Rooms.* Oxford, UK: Elsevier Ltd., 2008.

[4] A. Neidhardt, C. Schneiderwind, and F. Klein, "Perceptual matching of room acoustics for auditory augmented reality in small rooms - literature review and theoretical framework," *Trends Hear.*, vol. 26, pp. 1–22, 2022. DOI: 10.1177/23312165221092919.

[5] M. R. Schroeder and K. H. Kuttruff, "On frequency response curves in rooms. comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima," *J. Acoust. Soc. Am.*, vol. 34, no. 1, pp. 76–80, Jan. 1962. DOI: 10.1121/1.1909022.

[6] B. M. Fazenda, M. Stephenson, and A. Goldberg, "Perceptual thresholds for the effects of room modes as a function of modal decay," *J. Acoust. Soc. Am.*, vol. 137, no. 3, pp. 1088–1098, 2015. DOI: 10.1121/1.4908217.

[7] D. Jun, C. Glorieux, Z. Fisarova, J. Plasek, and M. Rychtarikova, "Perception of room mode dampening from a musical perspective," in *Proc. Forum Acusticum*, Turin, Italy, 2023, pp. 1–6.

[8] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 708–730, 2015. DOI: 10.1121/1.4926438.

[9] J. C. B. Torres, L. Aspöck, and M. Vorländer, "Comparative study of two geometrical acoustic simulation models," *J. Braz. Soc. Mech. Sci.*, vol. 40, no. 300, pp. 1–15, 2018. DOI: 10.1007/s40430-018-1226-1.

[10] M. Blau, A. Budnik, M. Fallahi, H. Steffen, S. D. Ewert, and S. van de Par, "Toward realistic binaural auralizations – perceptual comparison between measurement and simulation-based auralizations and the real room for a classroom scenario," *Acta Acust.*, vol. 5, no. 8, pp. 1–13, 2021. DOI: 10.1051/aacus/2020034.

[11] F. Stärz, L. O. H. Kroczek, S. Roßkopf, A. Mühlberger, S. van de Par, and M. Blau, "Comparing room acoustical ratings in an interactive virtual environment to those in the real room," in *Proc. Forum Acusticum*, Turin, Italy, 2023, pp. 5009–5016. DOI: 10.61782/fa.2023.0525.

[12] S. Fargeot, A. Vidal, M. Aramaki, and R. Kronland-Martinet, "Perceptual evaluation of an ambisonic auralization system of measured 3d acoustics," *Acta Acust.*, vol. 7, p. 56, 2023. DOI: 10.1051/aacus/2023052.

[13] E. Skovenborg and S. H. Nielsen, "Evaluation of designs for loudness-matching experiments," in *Proc. Intl. Conf. Subjective & Objective Assessment of Sound*, Poznań, Poland, 2004, pp. 1–16.

[14] Y. Suzuki and H. Takeshima, "Equal-loudness-level contours for pure tones," *J. Acoust. Soc. Am.*, vol. 116, no. 2, pp. 918–933, 2004, ISSN: 0001-4966. DOI: 10.1121/1.1763601.

[15] A. Field, J. Miles, and Z. Field, *Discovering Statistics Using R.* London, UK: SAGE, 2012.