Audio Engineering Society

Conference Paper

Presented at the AES 2026 International Conference on
Audio for Virtual and Augmented Reality and Immersive Games
2026 June 30 – July 3, Sorbonne University/d'Alembert & IRCAM/STMS, Paris, France

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Investigating the Perceptual Relevance of Voice Directivity in Virtual Vocal Instruction Environments

Eleni Tavelidou, Konstantinos Bakogiannis, and Areti Andreopoulou

Laboratory of Music Acoustics and Technology (LabMAT), Department of Music Studies,
National and Kapodistrian University of Athens, Panepistimioupoli, Ilisia, 15784, Athens, Greece

Correspondence should be addressed to Eleni Tavelidou (etavelidou@music.uoa.gr)

ABSTRACT

Given the limited research on the use of extended reality (XR) technologies in remote music instruction from the perspective of music tutors, this work examines the perceptual importance of voice directivity within a virtual reality (VR) environment. In particular, the perceptual ability to discriminate differences between a measured vocal directivity pattern and a slightly modified omnidirectional directivity pattern is investigated. Two listening tests were conducted to probe directivity perception under (i) static and (ii) dynamic listener conditions within a simulated music practice room, integrating 3rd order Ambisonics Room Impulse Responses (RIRs) and head-tracked binaural reproduction within an interactive Unity-based interface. The static test used an ABX discrimination model to assess directivity detectability as a function of location and stimulus content. The dynamic test involved free navigation around a virtual singer and the evaluation of the perceived directional plausibility, naturalness of the sound emission, and the adequacy of the experience for the assessment of the singer's vocal characteristics. The results suggest that while listeners can detect differences between vocal directivity patterns under controlled listening conditions, such differences may become less perceptually salient during dynamic interaction within a virtual environment. Nevertheless, the overall positive evaluations in the dynamic listening test indicate that the implemented spatial audio approach provides a plausible and effective auditory experience, supporting its potential use in XR-based applications for remote music instruction and performance evaluation.

1 Introduction

Remote music teaching has become an important part of contemporary education, yet the use of XR as a medium for remote music instruction, particularly from the perspective of the music tutor, remains unexplored. This is a critical gap, since music teaching relies strongly on the accurate perception of auditory qualities such as timbre, articulation, vocal quality, and expressiveness. [1]. In immersive virtual environments, these qualities may be affected by the spatial-audio rendering chain, including the room response, source-listener geometry, and the representation of source radiation/directivity.

Accurate auditory perception in complex scenes relies on multiple cues, such as interaural time and level differences, spectral filtering and room-related factors including reverberation and background noise [2]. In auralization, a realistic reproduction of these cues is

essential for a convincing acoustic display [3]. An attribute of significant importance relating to the above is source directivity. Real sources do not radiate sound uniformly in all directions; instead, their emitted spectrum varies with source orientation and this direction-dependent filtering can affect the perceived timbre and spatial impression of the source [4].

Previous studies have shown mixed findings regarding the perceptual importance of source directivity. Wang and Vigeant [5] found no significant perceptual differences between omnidirectional and realistically directional source models, possibly due to the limited directional octave-band data available, although listeners did detect differences between an artificially extreme beaming source and the former. On the other hand, Contreras-Saavedra and Viveros-Muñoz [2] reported that speech-source directivity affects distance perception with the largest errors occurring when the source faces away from the listener and when reverberation increases. However, no overall main effect of directivity on direction-of-arrival (DoA) accuracy was observed. Instead, DoA performance depended primarily on source angle, with lateral positions yielding the lowest errors, while the effect of directivity varied across directions. These results imply that source directivity may not uniformly affect all aspects of spatial perception; rather, its influence may be more prominent for some perceptual dimensions, such as distance, than for others, such as azimuth.

In singing voice research, Bakogiannis et al. [6] showed that small body-related variations during directivity measurements did not result in audible artifacts, however differences between omnidirectional and directional vocal radiation could still become perceptually discernible at lateral and rear positions. It should be noted, however, that the above studies formed solely listening-based investigations which did not involve the use of visual or XR means.

In virtual auditory environments these issues become even more complex. Visual context can influence auditory perception through cross-modal integration [7] and VR-specific factors may alter spatial judgments, including distance estimation [8]. Studies with embodied conversational agents have likewise produced inconclusive results. Wendt et al. [9] found no significant effect of speech directivity on realism or social presence, whereas Ehret et al. [10] showed that listeners could distinguish directional from omnidirectional speech, but not dynamic from static directivity. Listener movement appears to be important too. In this context, a distinction should be made between static directivity, where the source radiation pattern remains fixed during playback, and dynamic directivity, where the radiation pattern changes over time, for example as a function of phoneme, articulation, source orientation, or movement. A further distinction concerns the listening condition itself: in a static listening condition, the listener remains at a fixed position relative to the source, whereas in a dynamic listening condition, the listener is allowed to move through the virtual scene and therefore experiences continuously changing source–receiver angles. Sloma and Neidhardt [11] reported that source directivity was difficult to identify under static listening, but became more perceptible when listeners were allowed to move through the virtual scene.

In all, the literature suggests that the perceptual relevance of source directivity depends on several interacting factors, including the source type, room acoustics, listening geometry, movement availability, and visual context. However, little is known about these issues in immersive XR scenarios that focus on music teaching, where vocal radiation may be especially important.

The present study addresses this gap by investigating whether vocal directivity cues meaningfully influence perception in head-tracked binaural VR settings. More specifically, this work examines vocal directivity in immersive, head-tracked binaural VR through two complementary studies. Pilot study 1 focuses on controlled perceptual discrimination, testing whether listeners can detect differences between a measured human vocal directivity pattern and a slightly altered omnidirectional reference under selected source-listener angles in the form of an ABX test. Pilot study 2 investigates the same issue in an ecologically relevant dynamic listening XR scenario, where listeners experience the vocal source from the perspective of a music tutor, under conditions that resemble realistic music-teaching interactions and free movement within the virtual space whilst evaluating the conditions by means of 5-point Likert Scale questions. By combining those two pilot studies, this work aims to clarify not only whether vocal directivity is perceptually relevant, but also whether its modeling is justified for XR applications in remote music pedagogy.

2 Methods

The particulars of the experimental design as well as the apparatus used for the aforementioned studies are detailed in the section below.

2.1 Directivity Conditions

In order to examine the perceptual relevance of vocal directivity, two source radiation patterns were examined: a measured human vocal directivity intended to represent a realistic singer sound emission, and a slightly tailored omnidirectional reference pattern. The comparison of the aforementioned was designed to contrast an ecologically grounded vocal radiation to a more simplified off-the-shelf solution available in commercial acoustic software featuring customization/selection of this acoustical parameter (i.e., BERTA, CATT, ODEON).

The Measured Vocal Directivity condition, intended to represent average human vocal directivity, was based on laboratory directivity measurements conducted at LabMAT, averaged across ten professional singers (six male, four female) performing Greek-language singing material from classical, popular, and Byzantine traditions [12]. Those measurements were captured using a 29-microphone array across four elevation planes (-30°, 0°, $+30^\circ$, and $+90^\circ$). In that setup, directivity data for elevation planes below -30° were unavailable, hence these values were manually attenuated to a constant -30 dB.

The Omnidirectional directivity pattern (OmniCut30) condition was based on an omnidirectional directivity pattern, slightly modified in the vertical plane for angles below -30° to ensure compatibility with the angular limits of the measured dataset. This condition served as a simplified reference, which is widely used in current literature, rather than a realistic vocal radiation model [2, 5, 10].

2.2 Virtual Room and Acoustic Rendering

A small music practice room was modeled in CATT-Acoustic 9.1 with dimensions 4.5 x 4 x 3 m. The model was tuned to achieve RT 0.6 s (mid-frequency average between the 500Hz and 1kHz bands), consistent with recommended conditions for small practice rooms [13]. Absorption was applied evenly across surfaces to maintain a left–right room response symmetry relative to the singer and listening positions. Third Order Ambisonics (3OA) were generated in CATT-Acoustic for the selected source – listener positions for binaural reproduction. A virtual singer was placed at the center of the room. Across both experiments, the minimum source–receiver distance was set to ≈ 1 m, in order to reflect a realistic teacher–student setting.

2.3 Apparatus and Implementation

The virtual environments for the two studies were developed in Unity 2022.3.62f2, while the audio processing, routing, and playback were conducted in Max/MSP v8. Communication between the two software applications regarding head positional and rotational data was achieved via OSC messages. In the dynamic listening condition, the anechoically recorded stimuli were first convolved with Real-time interpolated 3OA RIRs using RoomZ [14]. The remainder of the processes were commonly approached in both experiments. The rotation of the scene, and thus the corresponding rotation of the RIRs was processed using the Scene Rotator available in the IEM Plug-in Suite [15], and the resulting 3OA signal was decoded to binaural using the Compass Binaural decoder available in the COMPASS plug-in suite [16].

In the absence of measured personalized HRIRs and whilst aiming for the optimization of the binaural listening experience, Interaural Time Difference (ITD) – individualization based on each participant's head circumference, was applied on the KEMAR HRTF set (AXD/SONICOM database [17, 18]), which features large pinnae and GRAS ear-simulator microphones. Participants were equipped with an Oculus Quest 3 head-mounted display, hand-tracked controllers, and open-back Sennheiser HD 650 headphones for binaural listening. Both experiments were carried out on a PC running 64-bit Windows 11, using a Steinberg UR22 USB audio interface.

2.4 Training and Participants

Each of the two pilot studies was preceded by a training session taking place in a virtual room different from the one developed for the main test. The purpose of undertaking these short training sessions was merely familiarization with the controls, locomotion, as well

as the elements and interaction with the virtual questionnaires. The total duration of both training tasks was 5 minutes, on average.

The training designed for the first pilot study was conducted in a virtual room comprising a number of teleportation areas and an interactive panel identical to the one included in the main experiment. A short description of the listening task was offered first followed, by a step-by-step guide to get accustomed to the locomotion system involving teleportation to specified areas, as well as the use of the interactive playback and response controls. Training in the second pilot study included a written description of the main listening task, familiarisation with the locomotion system, which involved free navigation in the virtual space, as well as the interaction with the response collection panel.

With regards to participants, fifteen adults (eight female, seven male), aged 23 to 52, M(37) and SD(8.5), took part in both listening tasks. The sample consisted entirely of individuals with a music-related background, including music students, choir conductors, vocal specialists, and sound engineers. Approximately two-thirds of the participants reported prior experience with VR equipment and/or similar experimental procedures.

3 Pilot Study 1: Static Listening Conditions

3.1 Overview

The purpose of this static listening experience was to examine perceivable differences in directivity patterns from selected listening positions in the perimeter of the emitting sound source (a virtual singer). The task involved two stimuli: 2 sec Pink Noise and 5 sec of Male voice singing (Baritone) [19]. The former was chosen as a control condition; it offers directivity cues across all frequencies and reduces the risk of the results being "stimulus-specific". The Baritone singing excerpt represents a median male voice. Using a male rather than a female voice allowed us to capture lower-frequency data, while maintaining adequate harmonic content through the 4 kHz octave, aligning perfectly with the limits of the directivity data under evaluation.

The azimuthal coordinates of the selected listening / evaluation positions were -165°, -45°, 15° and 90° (see Figure 1). The 15°, -45° and 90° angles are widely spread over the frontal hemisphere of the vocalist, serving a realistic scenario in vocal coaching activities. The $-165°$ angle, located behind the singer, was included as a rearward anchor condition. Performance at this angle was expected to reveal clearly perceivable differences between the two directivity patterns, due to the stronger front-back contrast in the measured vocal radiation.

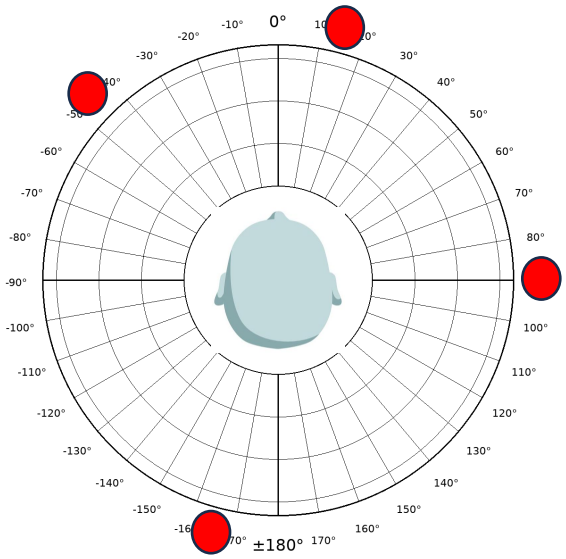


Fig. 1: Selected azimuthal angles used for the experimental design of Pilot Study 1

Table 1: Pilot Experiment 1: Experimental Design.

Factor	#	Elements
Directivity Pair	1	Human & OmniCut30
Angles	4	-165°, -45°, 15°, 90°
Stimuli	2	Baritone & Pink Noise
Permutations	4	ABA, ABB, BAA, BAB
Repetitions	3	Per combination

Based on an expected medium-size effect for this listening task, targeting high statistical power 80% as well as a minimum of 20 participants for the full-scale experiment in examination of our research questions, the experimental design of the pilot study was structured as shown in Table 1.



Fig. 2: Virtual music-room scene used in the Pilot listening study 1

3.2 Procedure

Each session starts with the participant's teleportation to the specified angle indicated in a virtual panel. Upon alignment, the participant listens to three sequential excerpts (ABX), of one of the two stimuli (either Baritone singing or Pink Noise), with excerpt X always being identical to either A or B. At the end of the playback the participant is asked to report whether the X audio excerpt is identical to A or B, submit their answer using the toggle buttons available and proceed to the next listening set (see Figure 2).

Trial sequences for each of the participants were generated using uniform random permutation sampling (Fisher–Yates shuffle implemented via the Mersenne Twister pseudo-random number generator in Python). The experimental design followed a fully crossed factorial structure (Stimulus × Angle × Directivity Pair × ABX Permutation), ensuring equal representation of all factor combinations prior to randomization. The ABX presentation order (A–B vs. B–A) and target identity (X=A vs. X=B) were fully counterbalanced to eliminate systematic order and response biases. To reduce unintended sequential dependencies, a constrained Monte Carlo optimization procedure was applied during randomization, minimizing consecutive trials sharing identical angle–stimulus–directivity configurations while preserving factorial balance. On average, participants completed the entire experimental session in 41 minutes (SD = 4 minutes).

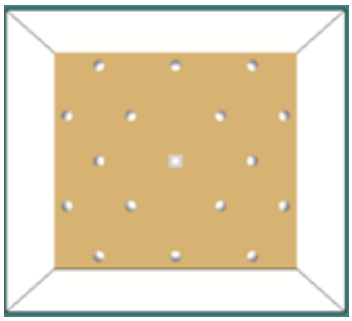


Fig. 3: Layout of the VR room showing the Impulse Responses generated in CATT for real-time interpolation during the free navigation within the VR space

4 Pilot Study 2: Dynamic Listening Conditions

4.1 Overview

The purpose of the dynamic listening experience, which entails free navigation around a virtual singer, was to evaluate and compare two acoustic conditions, differing solely in terms of the vocal directivity characteristics. These conditions were evaluated with respect to the perceived directivity emission, plausibility of the acoustic scene, as well as overall audio quality and adequacy of the experience for an educational setting.

Sixteen listening positions evenly distributed in the surrounding area of the virtual singer were accounted for in the generation of the 3OA RIRs of the acoustic scene (Figure 3). These were processed within the RoomZ plug-in as described in Section 2.3 for real-time interpolation of the auralized 3OA signal.

The experimental design comprised 2 conditions (one for each directivity pattern), 1 stimulus (baritone singing voice) and 3 repetitions per directivity. The duration of the playback was approximately 2.5 minutes per condition whereas the mean overall duration of the experiment was approximately 9 minutes (SD = 2.4 minutes). In this listening task 13 out of 15 participants were accounted for in the analysis, due to a technical error in the answers submitted.

4.2 Procedure

During the free exploration around the sound source in the Virtual environment participants were asked to

complete a 5-point Likert scale questionnaire. The question items were the following:

Q1: *How plausible does this acoustic experience seem given the characteristics of the room? (Could a singer realistically sound like this in such a space?)*

Q2: *How natural does the sound emitted by the singer seem during your navigation within the virtual room?*

Q3: *How convincingly does the sound appear to be emitted in the direction the singer is facing (i.e., forward)?*

Q4: *How adequate does the overall acoustic experience seem for evaluating the singer's vocal ability?*

More specifically, Question 1 targets the overall plausibility of the acoustic scene since it relates to room acoustics, level balance, and spatial impression in general. Question 2 is associated with the dynamic naturalness of the directivity emission. Question 3 focuses on the plausibility of the radiation and directional congruence matching the singer's facing direction, and Question 4 refers to functional quality and adequacy for the given purpose; the evaluation of the singer's vocal performance in a remote teaching scenario. The five points in the Likert scale were restricted to: Not at all, Slightly, Moderately, Very, and Extremely.

Prior to responding to the questionnaire in the virtual scene, participants were asked to complete a full circle around the virtual singer. It should also be noted that the questions appeared sequentially, so only after responding to one question participants could proceed to the next one. Revisiting a previous question for answer adjustments was also allowed. All questions were presented to participants prior to launching the listening test, so that they were familiar with the content and could actually focus on the acoustic experience while completing each question separately.

The trial order was randomized using a constrained permutation procedure. For each participant, the presented sequence was formed from a multiset of $N = 6$ trials comprising two conditions (Human directivity vs. OmniCut directivity), each repeated three times. Candidate orders were generated by drawing a (pseudo-)uniform random permutation of the full trial list (Fisher–Yates shuffle implemented via the Mersenne Twister pseudo-random number generator in Python).

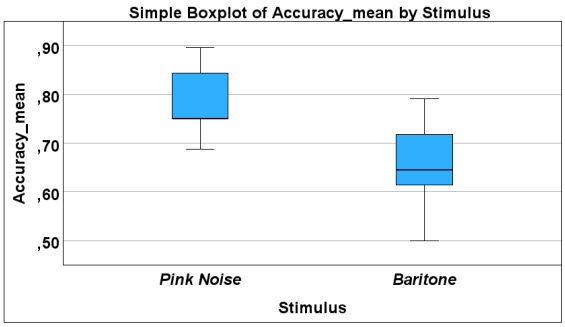


Fig. 4: Distribution of participant-level mean accuracy scores for the selected stimulus types (Pink Noise and Baritone)

5 Results

This section outlines the results for each of the two preliminary studies, followed by an overall interpretation of the findings. Overall, descriptive statistics were computed, and a Generalised Linear Mixed Model (GLMM) approach was employed to analyze the interaction effects between the experimental variables.

5.1 Pilot Study 1: data analysis and interpretation

A total of 1440 trials were collected from 15 participants. Across all experimental conditions, the mean accuracy in the ABX task was 73% (SD = 0.07). The distribution of the participants' performance, illustrated in Figure 4, indicates that they all performed above chance level (50%), suggesting that listeners were generally able to discriminate between the two tested directivity patterns within the virtual environment. With regard to the overall performance between the two stimulus types, it is shown that discrimination performance was consistently higher for pink noise (M=0.79) across participants, compared to the baritone stimulus (M=0.66). In contrast, the latter showed greater variability and overall lower identification accuracy.

The accuracy was further analyzed as a function of the position of the listener relative to the virtual singer as described in Section 3.1. Across both stimuli, the discrimination accuracy varied with the listener's angle. The highest performance was observed at the selected rearward angle -165°, followed by 90°, -45°, and finally 15°, which produced the lowest precision levels (see Figure 5). This pattern suggests that the contrast

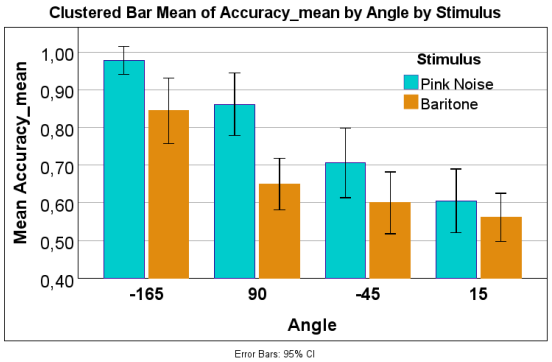


Fig. 5: Mean accuracy across listing angles and stimulus types

Table 2: Significance per Stimulus per Angle.

Angles	Stimuli	
	Pink Noise	Baritone
-165	p<0.05	p<0.05
90	p<0.05	p<0.05
-45	p<0.05	p=0.05
15	p=0.05	p=0.06

between the measured vocal directivity and the OmniCut30 condition was particularly salient at the sampled rearward position. However, because only one rearward angle was included, this finding should not be interpreted as a general assessment of rear-region discrimination or rear localization accuracy. Pairwise comparisons between angles indicated that performance differences were significant for most angle combinations, with the exception of the pair -45° and 15° where no statistically significant difference in accuracy was observed (p=0.097).

Higher accuracy toward the rear and lower toward the frontal listening positions was also observed when analyzing accuracy separately for each stimulus and angle. For both Pink Noise and Baritone stimuli, performance remained significantly above chance for all tested angles except the 15° angle for the former and the -45° and 15° angle for the latter (see Table 2), indicating that the perceptibility of directivity differences depends both on stimulus type and listener position.

These results suggest that the audibility of Human versus OmniCut30 directivity differences is cue-dependent. The stronger performance for pink noise is consistent with its broadband, stationary spectrum, which makes spectral coloration differences more prominent than in the time-varying Baritone excerpt. The marked advantage at -165°suggests that front–back radiation differences may have contributed to the discrimination performance at this selected position, whereas angles at the frontal area provide fewer discriminative cues and yield weaker evidence after correction for multiple testing.

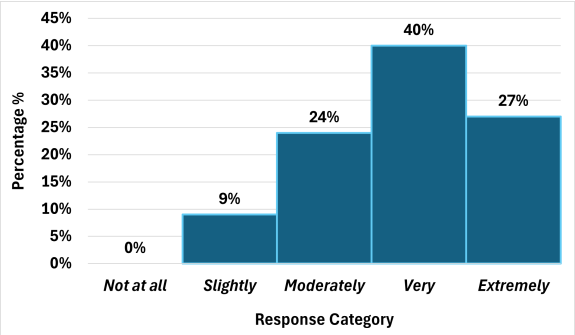


Fig. 6: Percentage distribution of questionnaire responses across the five-point scale

5.2 Pilot Study 2: data analysis and interpretation

Across all questions and conditions, participants tended to provide positive evaluations, with most responses falling in the categories “Very” and “Extremely” as illustrated in Figure 6. This suggests that the implemented spatial audio reproduction approach provided a convincing auditory experience during free navigation within the virtual space. The overall ratings in the dynamic listening environment indicate that, on average, the human directivity condition received higher scores than the OmniCut one. A Wilcoxon signed-rank test showed that the difference between the two conditions was statistically significant ($p < 0.05$) with a general preference for the measured human directivity pattern.

The mean ratings for the four questions ranged between 3.6 and 4.0, indicating generally positive evaluations of the acoustic experience. A Friedman test comparing all four questions showed no statistically significant differences among them ($F=6.3$, $p > 0.05$), suggesting that participants responded in a broadly consistent manner across the evaluated aspects of the acoustic experience.

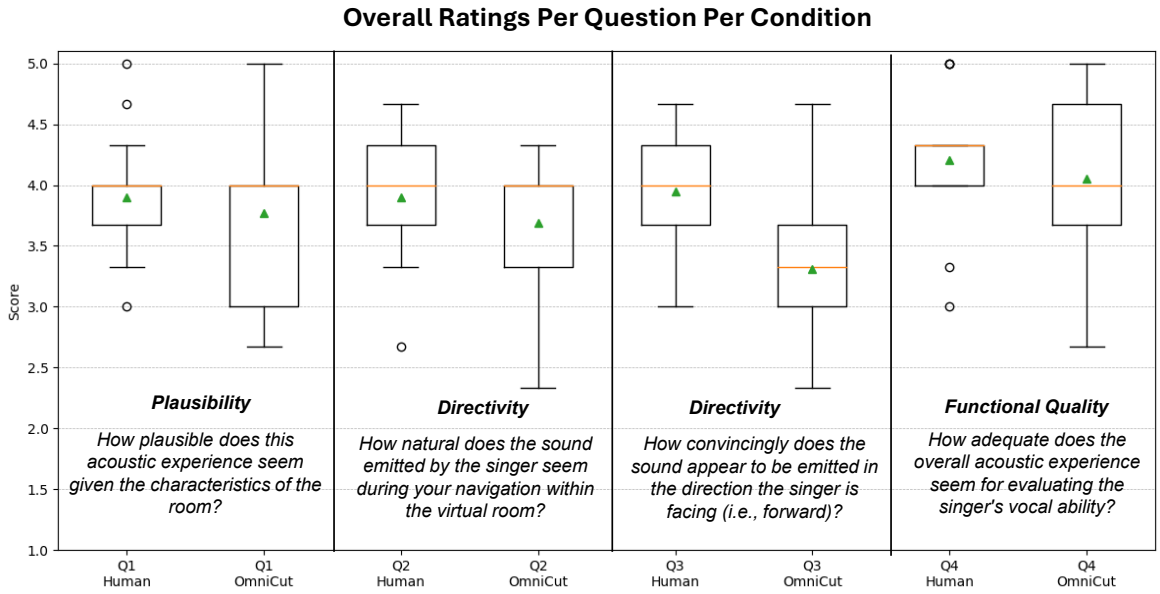


Fig. 7: Boxplots of participant ratings for the four questionnaire items under the Human and OmniCut conditions: Plausibility, Directivity–Naturalness of emission, Directivity–Perceived forward emission, and Functional Quality for performance evaluation

When responses were analyzed separately for each question and directivity condition, it was revealed that human directivity generally obtained higher ratings than the OmniCut one (Figure 7). However, the difference between conditions reached statistical significance only for Question 3 ($Z = -2.777$, $p = 0.005$), which assessed whether the sound appeared to be emitted from the direction the singer was facing. This result is consistent with expectations, as the omnidirectional pattern emits sound equally in all directions whereas the measured directivity pattern is a more realistic representation of the directional characteristics of a human voice, offering an enhanced alignment between the visual orientation of the singer and the emitted sound.

For the remaining questions (Q1, Q2, and Q4), the differences between the two directivity conditions were not statistically significant ($p > 0.05$), indicating that participants evaluated both acoustic conditions with respect to plausibility, naturalness, and functional adequacy of the experience similarly (Figure 7). The lack of significant differences in most evaluation criteria indicates that both directivity representations provided a similarly convincing acoustic experience during dynamic navigation.

6 Conclusions

This work investigated the perceptual relevance of vocal directivity in binaural listening scenarios within a virtual reality environment, as a first step toward evaluating the potential of utilizing XR technologies for remote teaching instruction from the perspective of the music tutor. The results of the two experiments provide complementary insights into the perceptual relevance of vocal directivity in virtual acoustic environments.

In the first pilot study, conducted under controlled static listening conditions, participants demonstrated an overall discrimination accuracy of approximately 73%, indicating that differences between the measured human vocal directivity and the omnidirectional condition were perceptually detectable. However, this ability was strongly dependent on listener position and stimulus characteristics, with differences becoming more evident at lateral and selected rearward listening positions, and when using broadband stimuli. These findings suggest that the perceptual salience of vocal directivity is context-dependent, emerging primarily under conditions where directional radiation cues are more pronounced and acoustically informative.

In contrast, the second pilot study, which examined user experience during dynamic navigation within the virtual environment, showed that both directivities received similarly positive evaluations of the acoustic experience, with participants rating the realism, naturalness, and functional adequacy of the system highly. Although the measured directivity was generally rated higher, statistically significant differences between the two conditions were observed only in the perceived directional congruence of the sound emission.

All things considered, these findings suggest that while listeners can detect differences between vocal directivity patterns under controlled listening conditions, such differences may become less perceptually salient during dynamic interaction within a virtual environment. In other words, the overall positive evaluations suggest that simplified directivity models may be sufficient for realistic and functional spatial audio reproduction in remote music interaction scenarios.

To build on these pilot findings, future work should involve a larger and more diverse participant pool, including individuals with specialized expertise in vocal performance and training, such as vocal students, professional singers, vocal coaches, and other expert listeners. This would allow for a more comprehensive assessment of the perceptual relevance of vocal directivity in ecologically valid music practice and teaching scenarios. In addition, although the selected baritone stimulus provided adequate harmonic content within the frequency range covered by the evaluated directivity dataset, future studies should investigate whether similar perceptual trends extend to female voices and other vocal registers.

It should also be noted that the static ABX study included only one rearward listening angle; therefore, the higher discrimination accuracy observed at -165° should be interpreted as specific to that sampled position rather than as evidence for rear-region perception more generally. Future studies should include additional rearward angles to assess whether this trend can be generalized across the rear hemisphere. Finally, although the ABX task used here followed a standard forced-choice format, future work could consider including confidence ratings or additional response options, such as “both” or “neither”, in order to capture cases where both alternatives are perceptually close to X or where listeners are uncertain in their judgment.

Overall, the findings indicate that vocal directivity can be perceptually relevant in controlled binaural listening, but that its practical importance in interactive VR music-instruction scenarios depends on the listening context, task, and level of perceptual detail required. As such, this work provides an initial step toward understanding when detailed vocal directivity modeling may be necessary, and when simplified source representations may be sufficient for plausible and functional XR-based remote music instruction.

7 Acknowledgments

This work was supported by SONICOM (www.sonicom.eu), a project that has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 101017743. Additionally, this work was funded and implemented in the framework of H.F.R.I call "3rd Call for H.F.R.I.’s Research Projects to Support Faculty Members & Researchers" (H.F.R.I. Project Number: 23677).

References

[1] McPherson, G., *The Oxford handbook of music performance*, volume 2 of *Oxford handbooks series*, Oxford University Press, 2022, ISBN 9780190058883, doi:10.1093/oxfordhb/9780190056285.001.0001.

[2] Contreras-Saavedra, C. E. and Viveros-Muñoz, R., “Does source directivity influence human perception of speech position?” in *Proceedings of the 11th Convention of the European Acoustics Association, Forum Acusticum 2025*, Málaga, Spain, 2025, doi:10.61782/fa.2025.0310.

[3] Vorländer, M., *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, RWTHedition, Springer-Verlag, Berlin, Heidelberg, 2008, ISBN 978-3-540-48829-3, doi:10.1007/978-3-540-48830-9.

[4] Rindel, J. H., Otondo, F., and Christensen, C. L., “Sound source representation for auralization,” in *International Symposium on Room Acoustics: Design and Science. Hyogo, Japan*, 2004.

[5] Wang, L. M. and Vigeant, M. C., "Evaluations of output from room acoustic computer modeling and auralization due to different sound source directionalities," *Applied Acoustics*, 69(12), pp. 1281–1293, 2008, doi:10.1016/j.apacoust.2007.09.004.

[6] Bakogiannis, K., Tavelidou, E., and Andreopoulou, A., "Estimation and assessment of vocal directivity variations as an effect of small body movements," in *Audio Engineering Society Convention 154*, Audio Engineering Society, 2023.

[7] Haverkamp, M., "Look at that sound! Visual aspects of auditory perception," in *III Congreso Intyernacional de Sinestesia, Granada*, volume 1, pp. 301–323, 2009.

[8] Roßkopf, S., Mühlberger, A., Stärz, F., Van De Par, S., Blau, M., and Kroczek, L. O., "Impact of Visual Virtual Scene and Localization Task on Auditory Distance Perception in Virtual Reality," *IEEE Transactions on Visualization and Computer Graphics*, 31(5), pp. 2464–2474, 2025, doi:10.1109/TVCG.2025.3549855.

[9] Wendt, J., Weyers, B., Stienen, J., Bönsch, A., Vorländer, M., and Kuhlen, T. W., "Influence of Directivity on the Perception of Embodied Conversational Agents' Speech," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, IVA '19, p. 130–132, Association for Computing Machinery, New York, NY, USA, 2019, doi:10.1145/3308532.3329434.

[10] Ehret, J., Stienen, J., Brozdowski, C., Bönsch, A., Mittelberg, I., Vorländer, M., and Kuhlen, T. W., "Evaluating the Influence of Phoneme-Dependent Dynamic Speaker Directivity of Embodied Conversational Agents' Speech," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, IVA '20, Association for Computing Machinery, New York, NY, USA, 2020, doi:10.1145/3383652.3423863.

[11] Sloma, U. and Neidhardt, A., "Investigations on the impact of listener movement on the perception of source directivity in virtual acoustic environments," in *Proc. of Annual German Conference on Acoustics (Munich, Germany, 2018)*, pp. 1433–1436, 2018.

[12] Bakogiannis, K. and Andreopoulou, A., "Spatial Directivity Characteristics of Greek-Language Singing," *Applied Sciences*, 16(4), 2026, ISSN 2076-3417, doi:10.3390/app16042014.

[13] Osman, R., "Designing small music practice rooms for sound quality," in *Proceedings of 20th International Congress on Acoustics, Sydney, Australia*, 2010.

[14] Poirier-Quinot, D., Stitt, P., and Katz, B. F., "RoomZ: Spatial panning plugin for dynamic RIR convolution auralisations," in *AES 2023 International Conference on Spatial and Immersive Audio*, 2023.

[15] Institute of Electronic Music and Acoustics, "IEM Plug-in Suite," `https://plugins.iem.at/`, 2025, accessed: 2026-03-24.

[16] McCormack, L. and Politis, A., "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.

[17] Engel, I., Daugintis, R., Vicente, T., Hogg, A. O. T., Pauwels, J., Tournier, A. J., and Picinali, L., "The SONICOM HRTF Dataset," *Journal of the Audio Engineering Society*, 71(5), pp. 241–253, 2023, doi:10.17743/jaes.2022.0066.

[18] Poole, K. C., Meyer, J., Martin, V., Daugintis, R., Marggraf-Turley, N., Webb, J., Pirard, L., La Magna, N., and Picinali, L., "The Extended SONICOM HRTF Dataset and Spatial Audio Metrics Toolbox," in *Proceedings of the 11th Convention of the European Acoustics Association, Forum Acusticum 2025*, pp. 6483–6486, Málaga, Spain, 2025, doi:10.61782/fa.2025.0864.

[19] D'Orazio, D., "Anechoic recordings of Italian opera," 2020, doi:10.5281/zenodo.3628246.