Perception of changes in source orientation-dependent reverberation for a visible virtual vocal source

Pablo Abehsera-Morell, David Poirier-Quinot, Brian F. G. Katz

Audio Engineering Society

Convention Express Paper 248

Presented at the 156th Convention
2024 June 15–17, Madrid, Spain

Perception of changes in source orientation-dependent reverberation for a visible virtual vocal source

Pablo Abehsera-Morell, David Poirier-Quinot, and Brian F.G. Katz

Sorbonne Université, CNRS, Institut Jean Le Rond ∂'Alembert, UMR 7190, Paris, France

Correspondence should be addressed to Pablo Abehsera-Morell
(pablo.abehsera_morell@sorbonne-universite.fr)

ABSTRACT

In extended reality experiences, dynamic rendering updates the auralisation as changes happen in the scene. In this study, we are interested in how the reverberant part of a directional vocal sound source is perceived as it rotates and whether it contributes to the perceived naturalness of the scene's acoustics. An experiment was conducted where participants were immersed in a virtual scene containing a visible virtual person source that would vocalise. The virtual source was auralised with a range of dynamic rendering approaches. Participants were asked to rate the perceived naturalness of the acoustics of the scene when listening to the different auralisations. The results of the perceptual test provide insights into how the changes in the reverberation pattern were perceived and how those changes affected perceived naturalness.

1 Introduction

When creating experiences in extended reality (XR), auralisation is used to assign desired acoustic characteristics to virtual sound sources. It is generally understood that, for a virtual source to be perceived as plausible, its acoustic characteristics must resemble those that it would have if it physically existed in the space where it is rendered [1, 2]. If the source or receiver are not static, the auralisation must be able to adapt and reflect any changes happening in the scene. Dynamic rendering is used to update the auralisation as such changes happen.

Directivity is one characteristic of sound sources that may require the use of dynamic rendering. It refers to how a sound source radiates energy as a function of direction and frequency. If a source is directional, its orientation relative to the listener will have an impact on how it is perceived. If the source's orientation changes, this will cause an acoustically perceivable change. In a reverberant environment, the directivity of a source ultimately affects the intensity of each of the reflections that arrive at a listener's ears from various directions, resulting in audible spectral and spatial cue differences [3].

Previous work has investigated the rendering of directivity in binaural auralisation of reverberant spaces. Steffens et al. [4] provides a good summary of the cues that must be present to plausibly render directivity.

The work of Postma et al. [5], Blau et al. [6] and Sloma et al. [7] investigated the impact of including source directivity rendering in auralisations. Overall, it is indicated that including source directivity in the rendering leads to higher plausibility of virtual sources.

It is not yet clear how accurate the directivity rendering needs to be when other modalities associated to XR, such as visuals, are present. In Postma and Katz [8], participants rated the plausibility of virtual sound sources with an accompanying visual reference. No significant differences were observed between a static directivity rendering and a dynamic directivity rendering, contrasting with the significant differences reported in a previous experiment where there was no visual reference [5]. Furthermore, in a study by Robotham et al. [9], it is suggested that 6-degrees-of-freedom (6DoF) listening may increase just-noticeable-difference (JND) values related to perception of source directivity. These works could imply that simplified directivity rendering may be used in XR without negatively impacting plausibility or perceived naturalness.

In this study, we are interested in how different degrees of directivity rendering are perceived by listeners when the virtual source is visible. It could be expected that an omnidirectional (or *static*) auralisation of a source that is visibly rotating may be perceived as less natural than one with directivity processing (or *dynamic*), as the acoustics would not be congruent with the visual; or, on the other hand, watching the source rotate could trick the listener into perceiving an acoustical change that is not happening, similarly to the way the ventriloquist effect affects sound localisation [10]. Moreover, it could also be of interest to observe how a partial implementation of directivity, or a completely dynamic but partially wrong implementation would compare to each extreme.

With this question in mind, we have conducted an experiment. Participants were immersed in virtual scenes, using a head-mounted display (HMD) and headphones, and rated the naturalness of a series of auralisations of a visible virtual source. The source was always a recording of a person, who would either tell a story or do other sorts of vocalisations (such as singing, whistling or beat-boxing) while facing different directions.

The results of this experiment could help to inform what degree of directivity processing to use in binaural auralisation of visible virtual sound sources in XR applications.

2 Methods

A listening test was prepared, employing static binaural audio rendering in a virtual reality application. This section breaks down the design of the experiment.

2.1 Acoustic and visual environments

Binaural room impulse responses (BRIRs) were captured in a meeting/library room (l 12 m, w 9 m, h 3 m; $T30_{mid}$ = 0.47 s). A directional loudspeaker (Genelec 8331A) was used as source and a binaural dummy head (Neumann KU-80) fitted with microphones (DPA 4060) at the entrance to the ear canals was used as receiver. Two source-receiver configurations were measured: one in the centre of the room ($\mathbf{Pos_{center}}$), and one where source and receiver were 1 m away from a wall with windows, forming a line parallel to it ($\mathbf{Pos_{wall}}$). In both configurations, source and receiver were 2 m apart and at a height of 1.5 m. For each configuration, BRIRs were captured for the source at 36 different horizontal orientation values: from 0° to 350° in 10° steps.

The room was visually recorded at 30 fps using a 360° camera (Ricoh Theta X). The camera was placed in each receiver position, facing straight at where the source was positioned. Video captured by the camera can be played back through a HMD, allowing the user visual 3DoF.

2.2 Acoustic and visual sources

Recordings for two different speakers were captured in an anechoic chamber using a head-mounted microphone (DPA 4088). The speakers, recorded individually, would vocalise while facing different directions,

Fig. 1: Measurement setup in $\mathbf{Pos_{center}}$ (L) and $\mathbf{Pos_{wall}}$ (R) of the test room.

as one may do when telling a story or singing. Clips lasting approximately 12 s were recorded.

The speakers were visually captured using the 360° camera, aimed straight at them. The camera was positioned 2 m away from the speakers, standing in a green-screen setup. For each captured clip, the green-screen was later removed using the keying feature (Adobe Premiere Pro), and the clip was then overlaid over each position configuration in the virtual room. Colour correction, brightness and contrast adjustments were applied to the video recordings to ensure they blended in as seamlessly as possible.

Orientation data to represent where the speakers were facing at each point in time was generated using Blender. An avatar was generated, and its orientation data was written, using automation, to match that of the speaker during each specific video clip. When playing back the video, it could be observed that the avatar was correctly mimicking the real-time orientation of the speaker. The orientation animations, saved at a rate of 30 Hz, were exported as text files for each clip.

2.3 Auralisation methods

Four acoustic conditions were devised to be rated in the listening test. In the first one ($\mathbf{Ac_{real}}$), each clip was auralised with the measured BRIRs. Using the generated orientation data, overlap-add convolution was used to convolve each frame with the BRIR corresponding to the measured orientation nearest to that of the speaker, at a rate of 30 Hz. Hanning windowing was used for smooth transition between frames. In the second condition ($\mathbf{Ac_{direct}}$), a modified version of the BRIR dataset was used, where the reverberant part of each BRIR was substituted with that of the BRIR corresponding to the source facing at the camera (azimuth 0°). This meant correct direct path directivity but identical reverberation across all orientations. In the third condition ($\mathbf{Ac_{static}}$), the BRIR of the source at azimuth 0° was used regardless of the source orientation. In the final condition ($\mathbf{Ac_{invert}}$), another modified BRIR dataset was used, where the direct sound was again left untouched but the reverberant part of each BRIR was substituted for that of the BRIR corresponding to a 180° shift. This meant correct direct path directivity and dynamic but incorrect reverberation, non-congruent with the visual orientation of the source.

Auralisations were created for each condition, for both source-receiver configurations. In total, nine clips were auralised: 3 *natural* speech clips ($\mathbf{Stim_{natural}}$), where the speaker told a short story at a normal intensity; 3 *forced* clips ($\mathbf{Stim_{forced}}$), consisting of **(i)** speech with a higher intensity, **(ii)** singing, and **(iii)** beat-boxing; and 3 *training* clips ($\mathbf{Stim_{training}}$) consisting of **(i)** the /tʃ/ phoneme (as in ***ch***op) repeatedly, **(ii)** whistling, and **(iii)** the word “pah" shouted repeatedly, with short pauses in between.

The motivation behind having different types of stimuli was to be able to assess the acoustic conditions for both ecologically valid scenarios ($\mathbf{Stim_{natural}}$) and scenarios that would better highlight the acoustics of the room ($\mathbf{Stim_{forced}}$, $\mathbf{Stim_{training}}$).

The goal when selecting which clips to use was to ensure that a variety of orientation change patterns were included in each of the three groups. For instance, in the $\mathbf{Stim_{natural}}$ group, one of the clips had a continuous anti-clockwise rotation, while the other two had more abrupt shifts in different directions.

2.4 Experimental setup

The listening test was carried out in a standalone application developed in Unity. Participants wore a HMD (Meta Quest 3) and headphones (Sennheiser HD-600) in an acoustically treated, semi-isolated studio room that ensured optimal listening conditions.

In the application, participants found themselves in the virtualised room, at one of the two captured positions. They could turn their head in 3DoF to look around the room, but their position was fixed. Participants were asked to stand during the experiment and to stay in the same spot during its duration.

As the binaural rendering was static, it was important for participants to face straight and not rotate their heads while listening to the auralisations. To ensure this, a movement threshold was implemented. If the participant did not face forward during the experiment, the audio would be muted, and an error popup would appear, reminding the participant that they should do so. A tolerance of 10° was allowed. It was verified before each participant began the experiment that the popup would not appear when they were facing straight in their natural posture to ensure that they would be comfortable during the experiment.

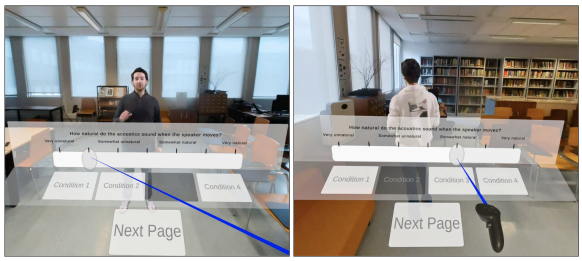

Fig. 2: Participant perspective for $\mathbf{Pos_{center}}$ (L) and $\mathbf{Pos_{wall}}$ (R).

2.5 Protocol

Figure 2 displays the visuals in the experiment. On each test page, participants were presented with one recorded clip at one of the two source-receiver positions. The speaker would look as if they were standing 2 m in front of the participants. A semi-transparent interface was presented in front of them, towards the floor at an angle, so as not to occlude the virtual speaker. To interact with the interface, participants used a laser pointer coming out of the controller and the frontal trigger to click.

Using the interface, participants could trigger and restart playback at will. They could choose between a number of acoustic conditions (2 during the training phase, 4 for the remainder of the experiment) to listen to while the clip played. On the top part of the interface, the question "*How natural do the acoustics sound when the speaker moves?*" was presented. A 7-point scale, with values labelled *Very unnatural* (**1**), *Somewhat unnatural* (**3**), *Somewhat natural* (**5**), and *Very natural* (**7**) was placed under the question. Ratings could be set using a slider. To avoid bias towards the initial value, the slider handle was invisible by default on each page, appearing only once a rating was first set. Participants could progress to the next page only once all conditions had been rated.

The listening test consisted of 15 scenes. The first 3 scenes, used for training, presented the 3 $\mathbf{Stim_{training}}$ stimuli at the $\mathbf{Pos_{wall}}$ configuration. Only the acoustic conditions $\mathbf{Ac_{real}}$ and $\mathbf{Ac_{static}}$ were presented during the training. The remaining 12 scenes presented all 4 acoustic conditions for the 6 $\mathbf{Stim_{natural}}$ and $\mathbf{Stim_{forced}}$ stimuli at both $\mathbf{Pos_{wall}}$ and $\mathbf{Pos_{center}}$ positions. Stimuli and position presentation order were randomised, as well as the order of the buttons used to trigger the acoustic conditions on the interface in each scene.

A pause scene was introduced in the middle of the 12-scene block, allowing participants to take a short break. During the break, the experimenter offered to remove the headset and allow them to sit down if they wished to.

After finishing the experiment, an informal interview was conducted with each participant. The aim of the interview was to complement the quantitative data provided by the ratings and help better understand what criteria participants used when rating the different conditions. They were asked to share their thoughts on the experiment and if they felt that any of the specific clips or one of the two positions made it easier or harder to distinguish between conditions.

2.6 Participants

A total of 17 people with self-reported normal hearing (4 female, 12 male, 1 non-binary) took part in the experiment. The mean age was 28.2 ± 3.5 years. All participants had previously been physically in the room that was virtualised. Based on their answers regarding experience with critical listening tasks and experience with HMDs, 8 participants were considered as experts and 9 were considered as naïve. The mean time taken to complete the experiment was 27 ± 7 minutes. Participants were compensated 15€.

3 Results

3.1 Analysis Method

To assess the impact of (**1**) *acoustic condition*, (**2**) *position in the room*, and (**3**) *stimulus* on participants' ratings, analyses of variances (ANOVAs) were conducted. The first-order interaction terms between the factors were also analysed. Statistical significance was determined for p-values below a 0.05 threshold. The notation $p < \varepsilon$ is adopted to indicate p-values below 10^{-3}. Post-hoc pairwise comparisons for significant factors were made with Tukey-Kramer adjusted p-values, or with Wilcoxon rank-sum p-values for unbalanced comparisons.

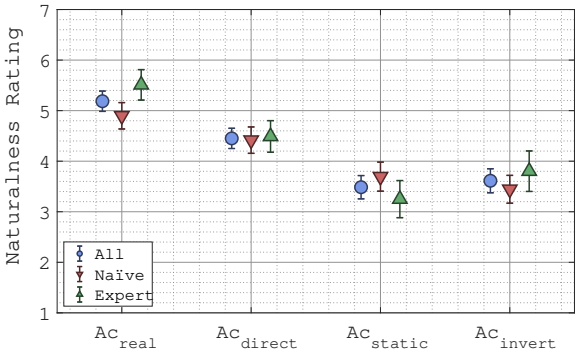

Fig. 3: Overall impact of acoustic condition on ratings. Mean and 95% confidence intervals (CI) displayed.

3.2 Direct impact of investigated factors

The acoustic condition had a significant impact on participants' ratings ($F = 51.7$, $p < \varepsilon$). $\mathbf{Ac_{real}}$ was rated above $\mathbf{Ac_{direct}}$ (5.2 ± 0.1 vs 4.5 ± 0.1, $p < \varepsilon$), itself rated above $\mathbf{Ac_{static}}$ (4.5 ± 0.1 vs 3.5 ± 0.1, $p < \varepsilon$). $\mathbf{Ac_{direct}}$ was also rated above $\mathbf{Ac_{invert}}$ (4.5 ± 0.1 vs 3.6 ± 0.1, $p < \varepsilon$). There was no significant difference between the $\mathbf{Ac_{static}}$ and $\mathbf{Ac_{invert}}$ ratings. The impact of acoustics on participants' ratings is depicted in Figure 3.

No significant difference was found between $\mathbf{Pos_{center}}$ and $\mathbf{Pos_{wall}}$ ratings. This can be observed in the similar rating trends for both participant groups, regardless of position, on Figure 4. Furthermore, there was no significant impact of stimulus category ($\mathbf{Stim_{natural}}$ or $\mathbf{Stim_{forced}}$) on ratings. Stimuli, on the other hand, had a significant impact on participants' ratings ($F = 3.1$, $p = 0.008$). Two of the three $\mathbf{Stim_{natural}}$ stimuli were rated above the $\mathbf{Stim_{forced}}$ beat-box stimulus (4.4 ± 0.1 and 4.3 ± 0.1 vs 3.7 ± 0.1, $p = 0.002$ and $p = 0.013$). Moreover, the singing stimulus was also rated significantly above the beat-box stimulus (4.4 ± 0.1 vs 3.7 ± 0.1, $p = 0.003$).

3.3 Further interactions

The analysis of variance did not reveal any significant first-order interaction between any of the investigated factors of acoustic condition, position, stimulus, and stimuli category. Focusing the analysis on individual acoustic conditions revealed that $\mathbf{Ac_{invert}}$ was rated significantly lower at $\mathbf{Pos_{wall}}$ than at $\mathbf{Pos_{center}}$ (4.0 ± 0.2 vs 3.3 ± 0.2, $p = 0.002$). No such difference was found for the other acoustic conditions.

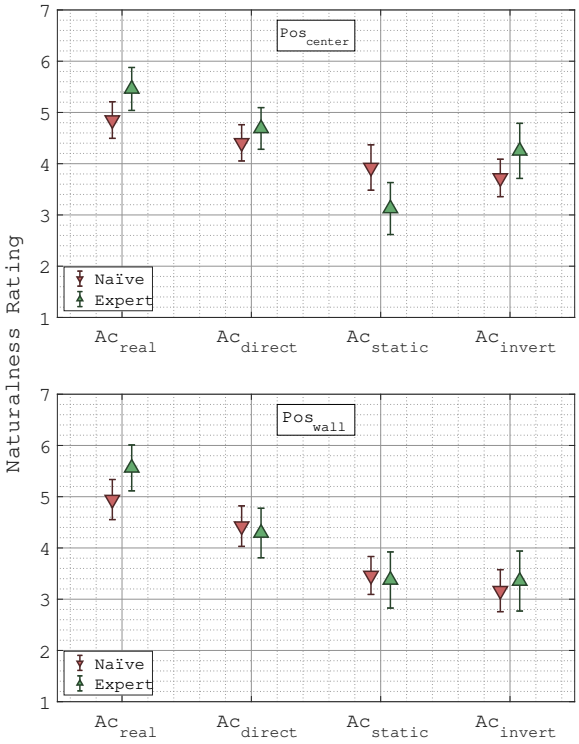

Fig. 4: Impact of acoustic condition on ratings for the two positions in the room. Mean and 95% CI displayed.

Interestingly, the group of expert participants rated the $\mathbf{Ac_{invert}}$ acoustic above $\mathbf{Ac_{static}}$ (3.8 ± 0.2 vs 3.2 ± 0.2, $p = 0.043$). For this group, $\mathbf{Ac_{invert}}$ actually received ratings similar to those of $\mathbf{Ac_{direct}}$ at the $\mathbf{Pos_{center}}$ position, significantly above the $\mathbf{Ac_{static}}$ condition (4.7 ± 0.2 and 4.2 ± 0.3 vs 3.1 ± 0.3, $p < \varepsilon$). This change was not observed at $\mathbf{Pos_{wall}}$. These results are illustrated in Figure 4.

3.4 Interviews

The most common idea shared by participants was that they felt most confident rating the $\mathbf{Stim_{natural}}$ auralisations. Non-experts usually claimed that, in the real room, they were used to hearing people talk, which helped them when comparing the sound of the virtual speaker to their internal reference of someone speaking in the room.

For other clips, such as the $\mathbf{Stim_{forced}}$ beat-box (which was universally identified as hardest to rate), they claimed that no internal reference existed because they were not used to hearing people beat-box, particularly in that type of room. According to them, this made it harder to rate how natural the auralisations felt. Some participants suggested that the way speakers moved during each stimulus also impacted their judgement. They argued that a consistent, progressive rotation was easier to judge than an abrupt one.

Around half of the participants felt more apt to appraise differences between acoustic conditions at the $\mathbf{Pos_{wall}}$ position. Most of the rest claimed that the position did not impact their ability to rate the conditions. Two participants said they preferred the $\mathbf{Pos_{center}}$ position.

Most participants alluded to noticing that one of the conditions was always static (or omnidirectional). Typically, they rated it as less natural when they identified it. Interestingly, a few participants claimed that they tended to prefer this condition, as they felt that the reverberation changes in the dynamic conditions were too drastic and, therefore, unrealistic.

4 Discussion

The observed impact of the acoustic conditions on ratings ($\mathbf{Ac_{real}} > \mathbf{Ac_{direct}} > \mathbf{Ac_{static}}$) was as expected, as naturalness was negatively impacted by the degradation of directivity in the auralisation. This is in line with the existing literature, which suggests that directivity rendering contributes to creating more realistic auralisations. In general, it could be said that the presence of the visual reference did not mask the acoustically perceivable differences between the ideal $\mathbf{Ac_{real}}$ and the degraded conditions. Interestingly, this contradicts the results of Postma and Katz [5, 8], where static rendering was rated as highly as dynamic rendering when there was a visual reference, but not when the experiment was audio-only.

It is interesting to see that the $\mathbf{Ac_{invert}}$ and $\mathbf{Ac_{static}}$ conditions were rated similarly. This could imply that having some sort of acoustical change when a visible source rotates is not enough by itself to produce a more natural experience, but rather, that the change has to be congruent with the visual to be perceived as natural.

For expert listeners, however, the $\mathbf{Ac_{invert}}$ condition was rated higher than the $\mathbf{Ac_{static}}$ condition overall, and it was even rated as highly as the $\mathbf{Ac_{direct}}$ condition at the $\mathbf{Pos_{center}}$. This implies that, depending on the acoustic configuration, a significantly wrong (but dynamic) rendering of directivity may be able to create a more natural auralisation than a rendering with no directivity at all.

Both naïve and expert listeners rated the $\mathbf{Ac_{invert}}$ condition significantly lower at the $\mathbf{Pos_{wall}}$ than at the $\mathbf{Pos_{center}}$. This could be explained by the fact that there is a bigger difference in the acoustic cues produced by the speaker facing left or right in $\mathbf{Pos_{wall}}$, with the noticeable presence of the close wall, than in $\mathbf{Pos_{center}}$. This implies that correct directivity rendering is particularly important when the acoustic configuration may be prone to producing particularly strong acoustic cue differences.

The impact of the stimuli on participants' ability to discern between the acoustic conditions was smaller than expected. Initially, it was expected that the $\mathbf{Stim_{forced}}$ would be better for revealing differences between acoustic conditions due to the more intense dynamics and variety of spectral content across clips. However, this was not the case, and it actually appeared that the clips did more harm than good – as highlighted by the interviews, where most participants said that the task was easier when assessing the $\mathbf{Stim_{natural}}$.

As some participants pointed out, other factors (such as the nature of the orientation shifts, as seen in Figure 5 and Figure 6) may have a bigger impact on the perceived naturalness of the auralisation. The results suggest that $\mathbf{Ac_{real}}$ was perceived as being significantly more natural than the other conditions when the orientation change was consistent and gradual, and that it was harder to perceive differences when the changes were abrupt.

5 Future Work

A clear next step to follow from this work would be to further investigate the impact of directivity rendering on the perceived naturalness of a visible virtual source, with a bigger focus on early reflections. Our results indicate that participants could tell the correct directivity rendering apart from a simplified approach and a partially wrong approach. Recent works such as [11] seek perceptually motivated rendering, aiming to enable plausible auralisation without needing to conduct extensive measurements or simulations. To contribute to this line of research, a study could be conducted

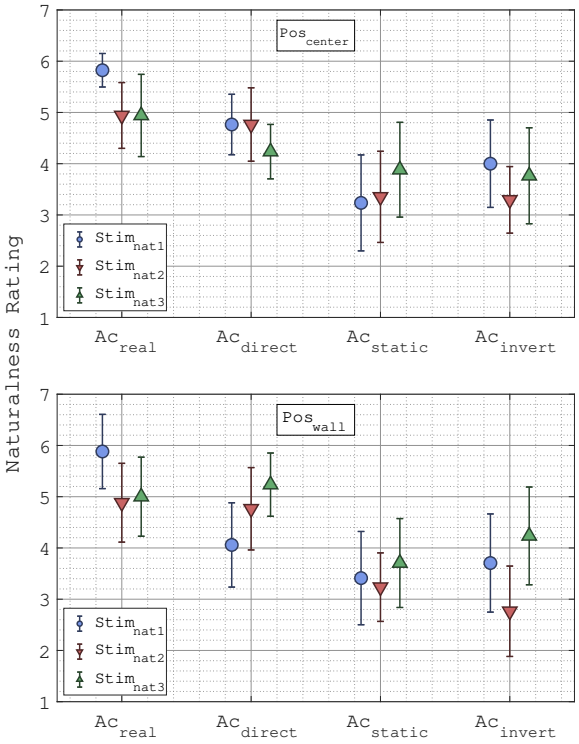

Fig. 5: Impact of the **Stim$_{\text{natural}}$** stimuli on ratings for the two positions in the room (all participants). Mean and 95% CI displayed.

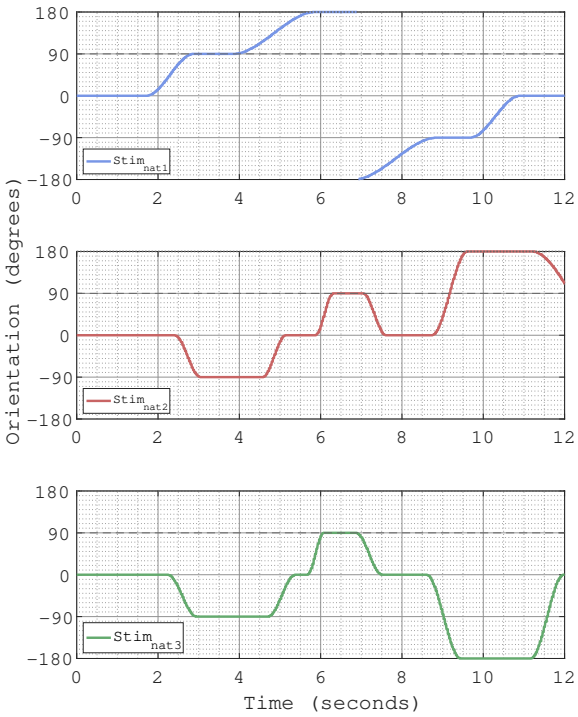

Fig. 6: Orientation data for the three **Stim$_{\text{natural}}$** stimuli. Positive 90° means facing towards the left of the source.

where different degrees of directivity-related processing in the early reflections are assessed for this kind of context.

Another path that could be followed would be to modify the task to become more ecologically valid. While the results of this work are valuable, it could be of great interest to properly assess the impact of directivity rendering on the realism of XR experiences without making participants actively focus on the acoustics, utilising paradigms more akin to those used in [12] or [13]. For instance, having an experimental setup where two participants talk to each other (instead of having single participants watch a recording) would allow to better understand the importance of directivity processing for having natural interactions in XR.

6 Summary

This paper reported the results of an experiment designed to assess the impact of directivity-related reverberation rendering on the perceived naturalness of a visible virtual vocal source. In this experiment, participants were immersed in a VR scene where a person spoke to them while facing in different directions.

In the VR experiment, a series of auralisation strategies were assessed. Auralisations made with realistic directivity rendering in the reverberant part were generally perceived as more natural than those made with static reverberation or partially incorrect directivity rendering. However, a partial implementation of directivity was better rated than no implementation at all and, depending on the position in the room, a more complete but partially wrong implementation. The influence of other factors on the perceived naturalness was also revealed. Listening expertise and the nature of the source's rotating motion may also play an important role.

Therefore, it can be concluded that the presence of a visual stimulus meant that the auralisations benefited from correct directivity rendering that was congruent with the visual. Instead of tricking participants into

perceiving wrong directivity rendering as correct, the visual somewhat highlighted the incongruities caused by auralisations that did not properly reflect what was being displayed.

7 Acknowledgements

We would like to thank the participants who took part in the listening experiment. This work was carried out in the context of the SONICOM project (`www.sonicom.eu`) that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017743.

References

[1] Brandenburg, K., Klein, F., Neidhardt, A., Sloma, U., and Werner, S., "Creating Auditory Illusions with Binaural Technology," in J. Blauert and J. Braasch, editors, **The Technology of Binaural Understanding**, pp. 623–663, Springer International Publishing, 2020, doi:10.1007/978-3-030-00386-9_21.

[2] Neidhardt, A., Schneiderwind, C., and Klein, F., "Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms - Literature Review and Theoretical Framework," **Trends in Hearing**, 26, 2022, doi:10.1177/23312165221092919.

[3] Bradley, J., Sato, H., and Picard, M., "On the importance of early reflections for speech in rooms," **J. of the Acoust. Soc. of America**, 113, pp. 3233–3244, 2003, doi:10.1121/1.1570439.

[4] Steffens, H., van de Par, S., and Ewert, S. D., "The role of early and late reflections on perception of source orientation," **J. of the Acoust. Soc. of America**, 149(4), pp. 2255–2269, 2021, doi:10.1121/10.0003823.

[5] Postma, B. N. J., Demontis, H., and Katz, B. F. G., "Subjective Evaluation of Dynamic Voice Directivity for Auralizations," **Acta Acustica united with Acustica**, 103(2), pp. 181–184, 2017, doi:10.3813/AAA.919045.

[6] Blau, M., Budnik, A., Fallahi, M., Steffens, H., Ewert, S. D., and Van de Par, S., "Toward realistic binaural auralizations–perceptual comparison between measurement and simulation-based auralizations and the real room for a classroom scenario," **Acta Acustica**, 5(8), 2021, doi:10.1051/aacus/2020034.

[7] Sloma, U., Werner, S., Klein, F., and Pappachan, T., "Synthesis of binaural room impulse responses for different listening positions considering the source directivity," in **147th Aud. Eng. Soc. Conv.**, 10237, 2019.

[8] Postma, B. N. J. and Katz, B. F. G., "Influence of visual rendering on the acoustic judgements of a theater auralization," in **Proc. Mtgs. Acoust.**, 015008, Boston, Massachusetts, 2017, doi:10.1121/2.0000575.

[9] Robotham, T., Rummukainen, O., Habets, E., and others, "Towards the Perception of Sound Source Directivity Inside Six-Degrees-of-Freedom Virtual Reality," in **Aud. for Virtual, Augmented and Mixed Realities: Proc. of 5th Int. Conf. on Spatial Aud.**, pp. 71–78, 2019.

[10] Howard, I. P. and Templeton, W. B., **Human spatial orientation**, John Wiley & Sons, London, 1966.

[11] Abehsera-Morell, P., Poirier-Quinot, D., and Katz, B. F. G., "Projecting source directivity variations onto an existing binaural room impulse response," in **Aud. Eng. Soc. Int. Conf. on Spatial and Immersive Aud.**, 37, Huddersfield, UK, 2023.

[12] 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 **Proc. of the 19th ACM Int. Conf. on Intell. Virtual Agents**, pp. 130–132, ACM, Paris France, 2019, doi:10.1145/3308532.3329434.

[13] Immohr, F., Rendle, G., Lammert, A., Neidhardt, A., Heyde, V., Froehlich, B., and Raake, A., "Evaluating the Effect of Binaural Auralization on Audiovisual Plausibility and Communication Behavior in Virtual Reality," in **2024 IEEE Conf. Virtual Reality and 3D User Interfaces (VR)**, pp. 849–858, IEEE Computer Society, Los Alamitos, CA, USA, 2024, doi:10.1109/VR58804.2024.00104.