HAL open science

Creating navigable auralisations using RIR convolution: Impact of grid density and panning method on perceived source stability

Julien De Muynke, David Poirier-Quinot, Brian F. G. Katz

▶ **To cite this version:**

Julien De Muynke, David Poirier-Quinot, Brian F. G. Katz. Creating navigable auralisations using RIR convolution: Impact of grid density and panning method on perceived source stability. AES 2023 International Conference on Spatial and Immersive Audio, Audio Engineering Society, Aug 2023, Huddersfield, United Kingdom. hal-04259798

HAL Id: hal-04259798

https://hal.science/hal-04259798

Submitted on 26 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Audio Engineering Society

Conference Paper 49

Presented at the International Conference on Spatial and Immersive Audio
2023 August 23–25, Huddersfield, UK

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. 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.

Creating navigable auralisations using RIR convolution: Impact of grid density and panning method on perceived source stability

Julien De Muynke[1,2], David Poirier-Quinot[1], and Brian F.G. Katz[1]

[1]*Sorbonne Université, CNRS, Institut Jean Le Rond ∂'Alembert, UMR 7190, Paris, France*
[2]*Eurecat, Centre Tecnològic de Catalunya, Tecnologies Multimèdia, Barcelona, 08005, Spain*

Correspondence should be addressed to Julien De Muynke (`julien.de_muynke@sorbonne-universite.fr`)

ABSTRACT

Convolution with spatial room impulse responses (RIRs) is often used to create realistic auralisations. The technique can be combined with spatial interpolation to create navigable virtual environments. This paper reports the preliminary results of an experiment designed to assess the impact of various interpolation parameters on perceived auditory source stability under various auralisation conditions. Participants freely explored a virtual scene while listening to a 3rd order Ambisonic RIR auralisation over headphones equipped with a tracked head-mounted display. They were asked to rate source stability under various conditions of RIR grid density, interpolation panning method, and room acoustics. A preliminary analysis of the results is presented.

1 Introduction

In recent years, more and more auditory virtual/augmented reality (VR/AR) experiences have been offered to visitors of cultural heritage spaces as a new kind of audio-guided visit with immersive audio content. For example, the visit offered in Vaux-le-Vicomte castle[1] includes historically informed soundscapes and automatic transition of the reproduced audio content between rooms thanks to proximity sensors distributed around the exhibition space. In the Hôtel de la Marine[2], an innovative mediation tool was recently deployed that additionally includes a tracked pair of headphones for each visitor. It allows to trigger different audio content according to the visitor's head orientation, *e.g.* a character starting to talk when the visitor looks at his portrait. The position tracking accuracy offered by recently emerging indoor position tracking technologies [**?**] may further allow for actual auditory walk-through in six degrees-of-freedom (6DoF), in which the reproduced auditory scene is continuously adapted to the visitor's position and head orientation. In this context, the degree of immersion in the virtual auditory scene greatly depends on the authenticity of the auralised

[1]Immersive-visit-vaux-le-vicomte.pdf
[2]Hotel-de-la-marine.paris

scene, *i.e.* on the perceived similarity of its sound attributes with those of a natural scene.

In practice, these experiences often rely on Room Impulse Response (RIR) convolution to produce realistic auralisations. With 6DoF, the auralisation must continuously be adjusted to the user's position within the scene in real-time. This can be achieved by selecting and combining RIRs from different positions following the user trajectory in the auralised space. This process is referred to as spatial interpolation. Compared to a discrete rendering based on a unique RIR, such interpolations can result in audio rendering artefacts, potentially detrimental to the authenticity of the auralisation.

The aim of this study is to assess whether those interpolation artefacts have an impact on the **perceived source stability** during free exploration of an audiovisual virtual environment. The results of a perceptual test that compared three panning methods applied on three RIR grid density conditions in real-time during the navigation are reported. The test was conducted with RIRs simulated in two different room acoustic conditions in order to assess how a change in reverberation time would affect stability ratings. The intent is that the reported results will serve as a guideline for the design of realistic navigable auditory environments.

2 Previous work

Various techniques have been proposed to create realistic navigable auditory scenes. Tylka and Choueiri [1] and Patricio [2] proposed approaches to interpolate between Ambisonic recordings spatially distributed within the sound scene. These respectively achieved accurate localisation with minimal spectral colouration, and auditory image naturalness and smoothness during navigation.

Alternatively, several techniques rely on the interpolation between RIRs prior to convolution with a dry source signal for auralisation. McKenzie et al. [3] proposed a perceptually informed method that interpolates a sparse grid of RIRs through separate treatment of the direct sound, the early reflections, and the late reverberation, and is robust to changes in room acoustics between coupled rooms. Similarly, Kearney et al. [4] and Masterson et al. [5] performed time-warping of a sparse set of RIRs in order to time align early reflections prior to spatial upsampling to reduce spatial blur. This upsampling proved to benefit localisation accuracy for static listener scenarios (the method was not tested during navigation conditions).

Müller and Zotter [6] proposed another method for upsampling a grid of RIRs based on joint localisation of early reflection peaks across RIRs and adjustment of their temporal and directional characteristics prior to interpolation. According to tests conducted on pre-rendered listener trajectories, this method achieved higher localisation accuracy and better sound colouration than a more naive interpolation approach. Finally, Geldert et al. [7] proposed a RIR interpolation method that preserves the temporal fine structure of the early reflection components better than the linear combination and the nearest neighbour methods.

The impact of the RIR grid density in the context of navigable virtual reality auralisation was specifically studied by Neidhardt and Reif [8], Werner et al. [9]. The present study extends the scope of those previous studies by proposing a characterisation of how RIR grid density, panning method, and room acoustics interact on the perceived auralisation during navigation. Compared to Neidhardt and Reif [8], Werner et al. [9], the current study uses Ambisonic RIRs instead of binaural RIRs, enabling listener head rotation at the expense of spatial resolution, both of which might impact perceived source stability during navigation.

3 Materials and methods

3.1 Room impulse response grids creation

This study uses 3rd order Ambisonic RIRs simulated using a calibrated geometrical-acoustic model of the St. Elisabeth church in Paris, developed in CATT-Acoustic, illustrated in Figure 1.

An omnidirectional source was located in front of the altar at a height of 1.5 m with receivers distributed in the same horizontal plane along homogeneous grids of various spatial densities covering the navigation zone. This zone, covering a 2×2 m^2 square, was located in the central nave, on the symmetry axis of the church. The closest distance between all the receiver positions and the source was 1.5 m.

RIR grids of various spatial densities were generated, composed of equilateral triangular cells of varying edge lengths of 0.5 m, 1 m, and 2 m. As shown in Figure 2,

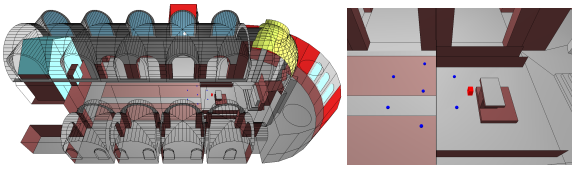

Fig. 1: (left) View of the CATT-Acoustic model of the St Elisabeth church. (right) Example RIR grid of receivers (blue spheres) for a given auditory source (red cube) position in the church, near the altar.

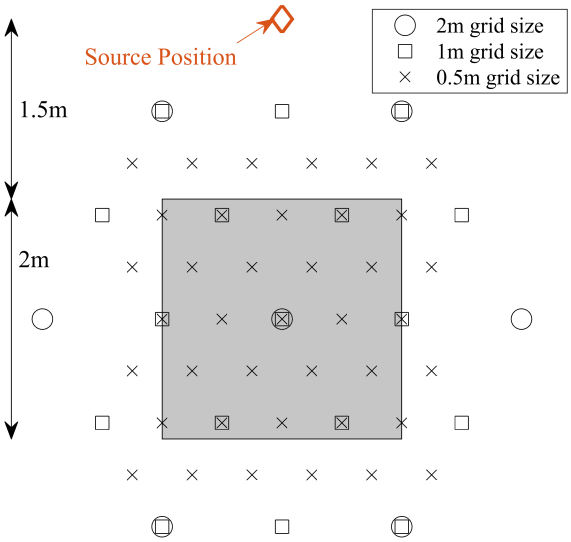

Fig. 2: Arrangement of the 2×2 m^2 navigation zone (grey area) and the simulated RIR positions for the 3 considered grid sizes.

all grids were aligned on the centre node of the experimental navigation zone. The number of RIRs required to cover the entire navigation zone varied with the grid spatial density: 39 RIRs for the 0.5 m grid, 17 RIRs for the 1 m grid, and 7 RIRs for the 2 m grid.

The two different room acoustics used in this study were generated with the same geometrical room model, built however with different acoustic materials. The "reverberant" acoustic condition corresponds to the acoustic of the actual church, calibrated based on measurements performed in St. Elisabeth. The acoustic calibration was done following the calibration procedure published in [10]. The "damped" acoustic condition

Octave band (Hz)	125	250	500	1k	2k	4k
Reverberant (s)	3.00	3.10	3.18	3.17	2.95	2.59
Damped (s)	1.13	1.16	1.19	1.19	1.14	1.04

Table 1: RT_{60} of St. Elisabeth model as a function of octave bands for the two considered acoustic conditions.

on the other hand was generated by using the same geometrical model with more absorbent materials. The RT_{60} reverberation time averaged over the listening zone is shown in Table 1 for both considered room acoustics. These two conditions were constructed so that the compared acoustics have the same temporal and spatial characteristics, differing only in energy and reverberation time.

A known issue with interpolating between spatially distributed RIRs is that of comb filtering effects due to slight differences in time of arrival of the direct sound and reflections. In order to avoid the most notable artefacts in the interpolated RIR, all generated RIRs were time-aligned on the direct sound by trimming the leading zeros corresponding to the propagation time.

3.2 Panning methods description

In the following, a cell comprised of RIR_i, RIR_j and RIR_k is denoted Cell_{ijk}. For any given target position contained in Cell_{ijk}, the panning method provides the amplitude gain-weights applied to RIR_i, RIR_j and RIR_k in the spatial interpolation. Three panning methods were compared in this study:

- **1NN**: only the single nearest neighbour RIR is selected, its gain is set to 1, regardless of position details.
- **3NN$_{\mathbf{dist}}$**: the 3 nearest neighbour RIRs are selected, each RIR gain is inversely proportional to its distance to the target position, denoted $\mathrm{d}_{i,i\in(1:3)}$ in Cell_{123} of Figure 3.
- **3NN$_{\mathbf{area}}$**: the 3 nearest neighbour RIRs are selected, each RIR gain is proportional to the surface area of the inner triangle formed by the two other selected RIRs and the target position, denoted $\mathcal{A}_{i,i\in(1:3)}$ in Cell_{123} of Figure 3.

For any given target position, $3NN_{dist}$ and $3NN_{area}$ each require three convolutions where 1NN only requires one. This means that $3NN_{dist}$ and $3NN_{area}$ require three times the CPU power compared to 1NN for a static listener position. As two adjacent cells always comprise two common RIRs and one unique RIR each, when the listener enters a new cell, any of the proposed panning methods temporarily performs one extra convolution with the RIR unique to the new cell. The signal convolved with the RIR unique to the old cell is crossfaded with the signal convolved with the RIR unique to the new cell to ensure a smooth transition. That is, when switching cells 1NN performs two convolutions while $3NN_{dist}$ and $3NN_{area}$ each performs four convolutions.

In $Cell_{ijk}$, the interpolation weights of RIR_i are denoted w_{dist_i} for $3NN_{dist}$ and $w_{\mathcal{A}_i}$ for $3NN_{area}$, and are calculated as follows:

$$w_{dist_i} = \frac{\frac{1}{d_i}}{\frac{1}{d_i}+\frac{1}{d_j}+\frac{1}{d_k}} \quad \text{for } 3NN_{dist}$$

$$w_{area_i} = \frac{\mathcal{A}_i}{\mathcal{A}_i+\mathcal{A}_j+\mathcal{A}_k} \quad \text{for } 3NN_{area}$$

Although the 3 RIRs selected by $3NN_{dist}$ and $3NN_{area}$ are always identical for any given target position, they may lead to different interpolation weights. In $Cell_{123}$ of Figure 3, the respective weights of RIR_1, RIR_2 and RIR_3 for target position T are:

$$w_{dist_1} = 0.26 \quad w_{dist_2} = 0.42 \quad w_{dist_3} = 0.32$$
$$w_{area_1} = 0.16 \quad w_{area_2} = 0.5 \quad w_{area_3} = 0.35$$

$3NN_{dist}$ and $3NN_{area}$ lead to the same order ranking of weights between the 3 selected RIRs for their respective contribution in the spatial interpolation: $RIR_2 > RIR_3 > RIR_1$. However, $3NN_{area}$ attributes a significantly larger weight to RIR_2 and RIR_3 compared to RIR_1, whereas $3NN_{dist}$ produces a more balanced contribution scheme between the selected RIRs.

Figure 3 shows the interpolation weights of the selected RIRs along the path $[TT'']$. On average, w_{dist_i} varies more smoothly and within a range of lower extent than $w_{\mathcal{A}_i}$. Moreover, when transitioning from $Cell_{123}$ to $Cell_{234}$, w_{area_1} decreases all the way down to 0 before w_{area_4} starts to increase from 0. In contrast,

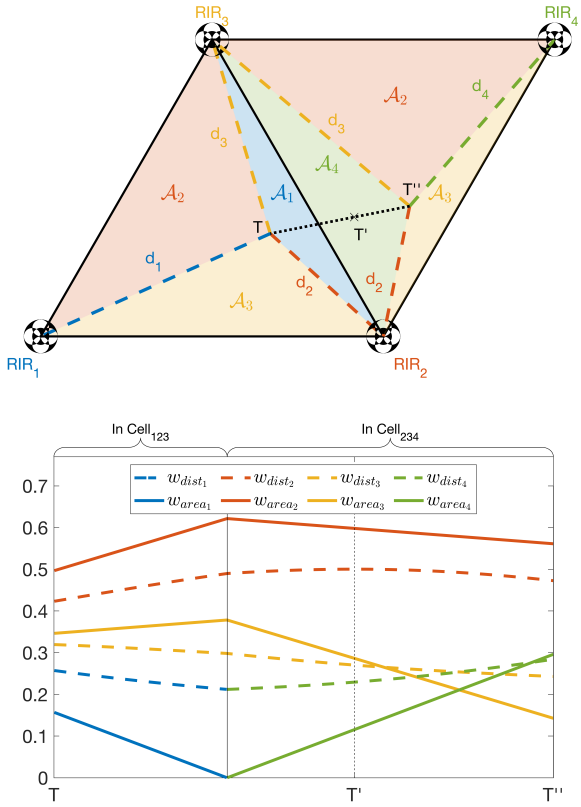

Fig. 3: (top) Visualisation of distances and surface areas of the inner triangles ($d_{i,i\in(1:3)}$ and $\mathcal{A}_{i,i\in(1:3)}$ for target position T in $Cell_{123}$ and $d_{i,i\in(2:4)}$ and $\mathcal{A}_{i,i\in(2:4)}$ for target position T'' in $Cell_{234}$). Path $[TT'']$ across $Cell_{123}$ and $Cell_{234}$ is marked by a dotted line. (bottom) RIR interpolation weights along the path $[TT'']$ as calculated by $3NN_{dist}$ (dashed lines) and $3NN_{area}$ (solid lines).

w_{dist_1} reaches 0.22 before w_{dist_4} starts to increase from the same value, *i.e.* on the edge between $Cell_{123}$ and $Cell_{234}$, RIR_1 still contributes of up to 22% in the interpolation before being substituted by RIR_4 with the same contribution. Consequently, $3NN_{area}$ may offer smoother transition (*i.e.* no abrupt RIR switches) than $3NN_{dist}$ when crossing adjacent cells.

Noteworthy, as seen in Figure 3, in every single cell crossed by $[TT'']$, $w_{\mathcal{A}_i}$ always vary monotonically whereas w_{dist_i} may change direction within the same cell. For example, as the target position moves towards T" in $Cell_{234}$, $w_{\mathcal{A}_2}$ decreases monotonically, indicat-

ing a constantly decreasing contribution of RIR_2 in the interpolation, whereas w_{dist_i} first starts to increase, reaches its maximum value at T', and starts to decrease, indicating that the contribution of RIR_2 in the interpolation first increases until T' and then decreases.

3.3 Experimental setup

The visual environment and user interface were developed in Unity and displayed in a Meta Quest 2 Head-Mounted Display (HMD). The audio scene was rendered in parallel in Max/MSP and reproduced over Sennheiser HD 600 headphones. The user's position and head orientation were tracked via the built-in cameras of the HMD and sent to the audio engine at a rate of 100 Hz via a local WiFi network using the OSC protocol. Spatial interpolation and uniformly partitioned convolution with the stimuli audio signal were performed using the RoomZ[3] plugin [11], whose crossfade time was set to 50 ms. Ambisonic rotation compensating for head orientation was done using the SceneRotator IEM plugin. Final binaural decoding used the common non-individual Neumann KU100 dummy head HRTF, distributed with the BinauralDecoder IEM plugin[4]. The I/O buffer length in Max/MSP was set to 1024 samples at 48 kHz.

3.4 Evaluation protocol

As shown in Figure 4, the virtual visual environment included a 2×2 m navigation zone in the centre of a virtual shoebox room. The room was kept empty, with realistic though rather simple textures to minimise any impact the visuals might have on the perceived auditory scene. The navigation zone was depicted by a carpet on the floor, enclosed by museum ropes at waist height attached to poles in each corner. The ropes aided the participants in understanding the extent of the navigation zone without having actually to look at the floor.

The experiment took place in the acoustically dry MotionCapture/VirtualReality room at the Institut Jean Le Rond ∂'Alembert. Before the experiment, participants were briefed on the position of the non-visual virtual auditory source, and were encouraged to explore the full extent of the navigation zone during the experiment. The actual test was preceded by a tutorial session where

[3]RoomZ website: roomz.dalembert.upmc.fr
[4]IEM plug-in suite website: plugins.iem.at

Fig. 4: (left) Screenshot of the visual environment. (right) Experimental test setup.

they could train to the task and get familiar with the user interface. Prior to any audio playback, the participants had to stand in the middle of the navigation zone in order to start the audio loop. This ensured that the perceived reference source position prior to navigation was similar across conditions. The audio source was temporarily muted if participants left the navigation zone, to prevent any unwanted auralisation artefacts. Each trial consisted of two consecutive loops of 20 s of the same stimulus and condition. The sound of a rattle, composed of a non-periodic sequence of click sounds, was chosen as the audio source, as impulsive sounds are known to be easier to localise. After the two repetitions, they rated the overall perceived instability of the source position during navigation by answering the following question: "In this scenario, how would you judge the instability of the source position when you navigate?" using a 7-point Likert scale. The odd-numbered rating marks were labelled (1) "Unnatural", (3) "Clearly noticeable", (5) "Slightly noticeable", and (7) "Unnoticeable" in ascending order. The conditions were randomised and repeated twice to gauge participants' consistency and to compensate for a potential training effect. The experimental environment can be seen in Figure 4. After the test, participants answered a questionnaire to evaluate their level of fatigue, level of self-confidence in their rating, experience with such evaluation tests, and to report other audio artefacts they might have perceived during the navigation.

3.5 Participants

A total of 27 paid subjects with an average age of 32.2 years participated (22 males, 5 females). 31% of them had already participated in at least 3 sound localisation tests, and as such are considered as expert listeners during the analysis. All participants stated having normal hearing abilities. The average duration of the experiment was 32.5 min and about 85% of

the participants reported at least a bit of fatigue after completing the test. The ratings of one participant were removed from the statistical analysis because of a low repeatability rate across repetitions of the same conditions.

3.6 Data analysis

Analyses of variances (ANOVAs) of participants' ratings were conducted to assess the effect of the different factors of panning method, room, grid size, critical listening expertise, and the first-order interaction terms between them. 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 ranksum p-values for unbalanced comparisons.

4 Results and discussion

4.1 Impact of the panning method and grid size

The panning method had a significant impact on participants' ratings ($F = 83.4$, $p < \varepsilon$). 1NN was rated overall significantly below $3NN_{dist}$ (3.2 vs 4.7, $p < \varepsilon$), itself rated below $3NN_{area}$ (4.7 vs 5.0, $p = 0.032$). Those ratings correspond to auditory source position instabilities rated on average as "clearly perceptible" for 1NN, and "slightly perceptible" for both $3NN_{dist}$ and $3NN_{area}$.

The RIR grid size also had a significant impact on participants' ratings ($F = 114.4$, $p < \varepsilon$). Those overall significantly decreased with increasing grid size: the 0.5 m grid was rated as more stable than the 1 m grid (5.1 vs 4.7, $p = 0.014$), itself rated as more stable than the 2 m grid (4.7 vs 3.0, $p < \varepsilon$). Those ratings correspond to instabilities judged as "slightly perceptible" for the 0.5 m and 1 m grids, and as "clearly perceptible" for the 2 m grid.

No significant impact of the room condition or critical listening expertise was observed on participants' ratings.

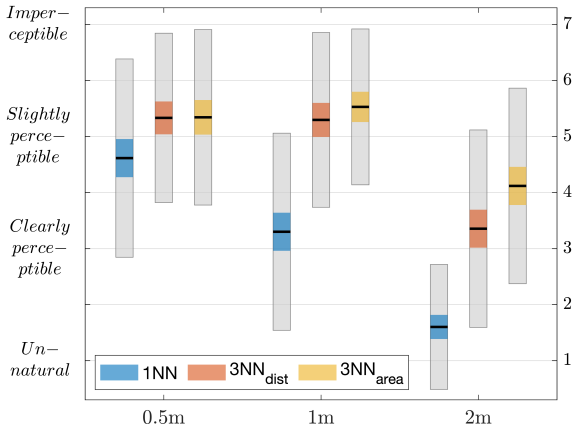

Fig. 5: Mean (—), 95% CI (coloured area), and standard deviation (grey area) of ratings of perceived source instability versus grid size, aggregated over panning method.

4.2 Further interactions

Analysis revealed a significant interaction between the grid size and the panning method regarding participants' ratings ($F = 10.3$, $p < \varepsilon$), illustrated in Figure 5. As expected, ratings overall increase across panning methods (1NN < $3NN_{dist}$ < $3NN_{area}$) and with decreasing grid size (2 m < 1 m < 0.5 m). The decomposition, however, indicates that the difference observed between the 0.5 m and 1 m grid size conditions only held for 1NN panning condition (4.6 vs. 3.3, $p < \varepsilon$), and was non-significant otherwise. It also reveals that $3NN_{area}$ panning method was actually rated higher than $3NN_{dist}$ (4.1 vs 3.4, $p = 0.002$) when using the 2 m grid size.

Interestingly, there was a significant difference between how self-reported critical listening experts and non-experts rated the different panning methods ($F = 5.1$, $p = 0.006$), as seen in Figure 6. $3NN_{area}$ was overall rated higher than $3NN_{dist}$ by the experts (5.6 vs 4.8, $p = 0.007$), while the non-experts did not perceive any difference between these panning methods. Similarly, the added value of using a 0.5 m RIR grid compared to a 1 m grid was only perceived by critical listening experts (5.6 vs 4.9, $p = 0.011$).

4.3 Discussion

The 1NN panning method (single nearest neighbour) clearly led to higher perceived source instability than

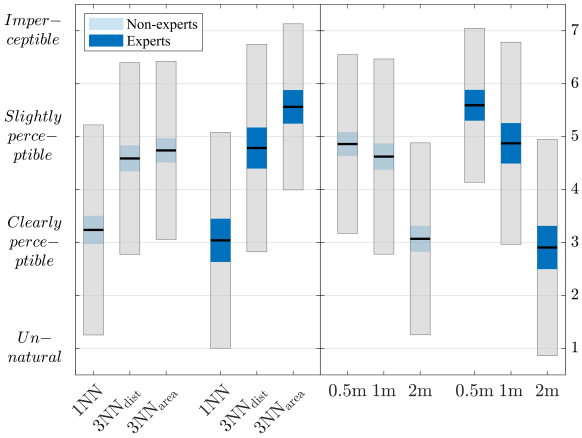

Fig. 6: Mean (—), 95% CI (coloured area), and standard deviation (grey area) of expert vs. non-expert ratings across panning method (left) and grid size (right) conditions.

the other methods, regardless of grid density. In contrast, the $3NN_{dist}$ and $3NN_{area}$ methods performed similarly for the two highest grid densities, while the latter resulted in a more stable rendering for the lowest 2 m density. This difference was more pronounced in the damped room than in the reverberant room. Those observations suggest that if the reproduction device offers sufficient CPU power to support one of the 3-RIRs panning methods, $3NN_{area}$ is the best choice overall.

$3NN_{area}$ and $3NN_{dist}$ maintained their performance for a grid density below 1 m, meaning that in the given configuration, the perceived source stability did not benefit from grid sizes below 1 m threshold when using either of the panning methods.

The 1NN method used on a 0.5 m grid density led to the same perceived stability as the other two methods used on a 2 m grid. This suggests that for a reproduction device with limited CPU power, the 1NN panning method could be used to produce a comparable level of stability as the other two methods at the cost of a higher RIR grid density requirement, at least for the tested scenarios here. This subsequently entails a higher storage requirement on the reproduction device.

Most participants reported that they preferred navigation in the front half of the navigation zone, *i.e.* closer to the auditory source, as it made source instability detection easier. In addition, most participants reported that they were mostly looking towards the source when navigating, as it eased the detection of source instability. This could be related to the fact that the minimum audible angle is smaller in the frontal listening area than on the sides [12]. These observations suggest that the instability sensitivity may be reduced for use cases comporting multiple auditory sources around the listener. In addition, one could expect reduced sensitivity as a function of source distance, as the angular span for a given grid density will be reduced the further away.

5 Conclusion

This paper reports the results of a perceptual test whose aim was to assess how, during 6DoF auralisation based on RIR convolution, auditory source stability was impacted by the RIR spatial grid density and the RIR panning method. The test examined the impact of three grid densities, three panning methods, as well as two room acoustic conditions.

Results showed that the perceived auditory source stability overall increased with increasing grid density, a result expected and in line with that reported by Neidhardt and Reif [8], Werner et al. [9]. Interestingly, perceived stability reached a plateau for grid size of 1 m and below for all but the simplest 1NN panning method. Results also indicated that that method was systematically rated below the other two, and that the $3NN_{area}$ method outperformed the $3NN_{dist}$ method for an RIR grid density of 2 m.

Self-proclaimed expert listeners proved to overall further benefit from a higher grid density. They on average preferred the $3NN_{area}$ over the $3NN_{dist}$ panning method. Finally, no significant impact of the room acoustics (reverberation time for the same geometry) on the perceived stability of the auditory source was observed.

Future work will focus on the perceptual evaluation in similar conditions of other auditory attributes impacting the authenticity of the auralisation, like the apparent source width or the sound colouration. Further tests on multi-modal interactions, such as the impact of visuals on perceived stability, are necessary to understand better how to deploy RIR based auralisations in mixed reality environments.

Results presented here can already serve as a guideline for the design of navigable auditory scenes used in

general public applications such as immersive audio-guides. Such designs usually balance auditory scene quality against CPU load and rendering device storage capacity. On the one hand, the design of high density RIR grids requires a longer time for either simulations or measurements and a higher storage capacity on the rendering device. On the other hand, using a panning method that requires 3 RIRs convolutions will be more CPU demanding, and not possible on some devices, compared to a simpler single RIR rendering method. The rating comparison between the 1NN method with a 0.5 m grid and the $3NN_{area}$ method with a 2 m grid illustrates this dilemma. The 1NN method requires three times less CPU than the $3NN_{area}$, while the 0.5 m RIR grid requires 5 times more storage than the 2 m grid (39 RIRs versus 7 RIRs) to cover the 2×2 m^2 navigation zone used in this study.

6 Acknowledgements

We would like to thank the participants who took part in the listening experiment. Funding has been provided by the European Union's Joint Programming Initiative on Cultural Heritage project PHE (The Past Has Ears, phe.pasthasears.eu), and the French project PHEND (The Past Has Ears at Notre-Dame, Grant No. ANR-20-CE38-0014, phend.pasthasears.eu).

References

[1] Tylka, J. and Choueiri, E., "Soundfield Navigation using an Array of Higher-Order Ambisonics Microphones," **Int. J. of Med. Inform.**, 49(1), pp. S33–S36, 2016, doi:10.1016/s1386-5056(98)00032-x.

[2] Patricio, E., "Toward Six Degrees of Freedom Audio Recording and Playback Using Multiple Ambisonics Sound Fields," in **Proc. of the 146th Aud. Eng. Soc. Conv.**, pp. 1–11, 2019.

[3] McKenzie, T., Meyer-Kahlen, N., Daugintis, R., McCormack, L., Schlecht, S. J., and Pulkki, V., "Perceptually informed interpolation and rendering of spatial room impulse responses for room transitions," in **Proc. of the 24th Int. Congr. on Acoust.**, 2022.

[4] Kearney, G., Masterson, C., Adams, S., and Boland, F., "Dynamic Time Warping for Acoustic Response Interpolation: Possibilities and Limitations," in **Proc. of the 17th Eur. Signal Process. Conf.**, pp. 705–709, 2009.

[5] Masterson, C., Kearney, G., and Boland, F., "Acoustic Impulse Response Interpolation for Multichannel Systems Using Dynamic Time Warping," in **Proc. of the 35th Aud. Eng. Soc. Int. Conf.**, 2009.

[6] Müller, K. and Zotter, F., "Auralization based on multi-perspective ambisonic room impulse responses," **Acta Acustica**, 4(6), p. 25, 2020, doi:10.1051/aacus/2020024.

[7] Geldert, A., Meyer-Kahlen, N., and Schlecht, S. J., "Interpolation of Spatial Room Impulse Responses Using Partial Optimal Transport," in **Proc. of the Int. Conf. on Acoust., Speech and Signal Process.**, pp. 1–5, 2023, doi:10.1109/ICASSP49357.2023.10095452.

[8] Neidhardt, A. and Reif, B., "Minimum BRIR grid resolution for interactive position changes in dynamic binaural synthesis," in **Proc. of the 148th Aud. Eng. Soc. Conv.**, 2020.

[9] Werner, S., Klein, F., and Götz, G., "Investigation on spatial auditory perception using non-uniform spatial distribution of binaural room impulse responses," in **Proc. of the Int. Conf. on Spatial Audio**, 2019, doi:10.22032/dbt.39936.

[10] Postma, B. N. and Katz, B. F., "Creation and Calibration Method of Virtual Acoustic Models for Historic Auralizations," **Virtual Reality**, 19, pp. 161–180, 2015, doi:10.1007/s10055-015-0275-3.

[11] Poirier-Quinot, D., Stitt, P., and Katz, B. F., "RoomZ: Spatial panning plugin for dynamic RIR convolution auralisations," in **Proc. of the Aud. Eng. Soc. Int. Conf. on Spatial and Immersive Audio**, submitted 2023.

[12] Mills, A. W., "On the minimum audible angle," **J. of the Acoust. Soc. of America**, 30(4), pp. 237–246, 1958.