DOI: 10.61782/fa.2025.0055

FORUM ACUSTICUM
EURONOISE 2025

ACCOMMODATION TO VIRTUAL SOUND RENDERING IN A REVERBERANT ROOM WITH 3DOF AND 6DOF INTERACTION

Julie Meyer[1,2*] **Lorenzo Picinali**[2]

[1] Department of Engineering, King's College London, United Kingdom
[2] Dyson School of Design Engineering, Imperial College London, United Kingdom

ABSTRACT

This study investigates the perception of real (loudspeaker) and virtual (binaural rendering through headphones) sound sources in a reverberant room with 3DoF and 6DoF interaction. More specifically, the study looks at the effect of accommodation to the virtual audio rendering while carrying out a listening experiment aimed at evaluating the presence of real and virtual sources. The experiment contains three blocks of triplets organised as follows: the 10 first triplets always contained one real sound source, the next 25 triplets had only virtual sources, and the 10 last triplets had again one real sound source. The results show that most of the participants correctly identified the presence of the real sound source during the first block, and started to identify virtual sound sources as real ones in the second block, possibly indicating an accommodation to the virtual rendering. In the third block, participants correctly identified the presence of the real sound source again.

Keywords: *binaural rendering, 6DoF movements, accommodation, audio augmented reality*

1. INTRODUCTION

In acoustic or audio augmented reality (AAR) scenarios, where the aim is to seamlessly blend virtual auditory events with the physical world, several factors affect the perceived end results. The individual nature of human localization cues, the room acoustics, and the use of transparent devices (e.g., headphones) to reproduce both the real and virtual sound sources are among such factors, together with vision and multisensory integration [1]. This means that the rendering requirements in order to properly blend the virtual and real domain signals will vary depending on the context (e.g., acoustic environment) and on the level of interaction (e.g., 3DOF or 6DOF). For example, creating fully static (i.e., with neither head rotations nor translation movements) binaural renderings that would seamlessly blend into anechoic environments presents less challenges than doing so for fully dynamic (i.e., including movements in 6DoF) ones in echoic environments. As it is of interest to create AAR scenarios resembling realistic life conditions, this study looks at 3DoF and 6DoF movements within a relatively reverberant room.

Another relevant aspect to consider is the choice of testing paradigm employed, and its associated experimental design to evaluate the suitability of the virtual audio rendering in augmented reality (AR) contexts. To this date, a few different paradigms have been proposed, although no agreement on which is the most useful for this purpose exists [2]. Authenticity, which refers to the perceptual identity with an explicitly presented real event (agreement with an external reference) [3], has proven to be hard to achieve in AAR scenarios with dynamic rendering. On the other hand, plausibility [4], which is typically evaluated with a yes/no experimental test design, does not require direct comparisons between real and virtual sound sources. As such, it is less strict than the first, but presents some limitations in AR contexts since it is likely that real sound sources overlap with virtual ones. To overcome such limitations, the notion of transfer-plausibility

**Corresponding author*: julie.meyer@kcl.ac.uk.

was proposed in [5, 6], and more precisely defined in [2] as "*A virtual sound source is transfer-plausible if it is believed to be real in the presence of real sound sources*".

The study from Neidhardt and Zerlik [7] evaluated the plausibility of AAR realised with position-dynamic binaural rendering. Their study comprised two experiments; in the first one only virtual sound sources were included, while the second one exploited a mix of real (i.e., loudspeaker) and virtual ones. The results revealed that there was an influence of the presence/absence of real sound sources in the plausibility assessment of the AAR rendered in fully-dynamic binaural conditions. Indeed, the inexperienced listeners tended to accept the virtual sound sources as real in the first experiment, while experts distributed their answers approximately equally. In the second experiment, expert listeners were able to identify the virtual sound sources quite reliably. The authors reported that for inexperienced listeners, the responses varied largely in the second experiment. The results from another study, which evaluated two binaural rendering approaches in AAR scenarios with position-dynamic binaural rendering, also showed that predominantly presenting virtual stimuli to the participants resulted in lower percentage of correct answers compared to presenting equal amounts of real and virtual sound sources, or even more real sound sources than virtual ones [8].

Building upon these observations, the present study examined closely the effect of presenting unbalanced numbers of real and virtual sound sources to the participants to verify if there was an accommodation effect to listening to the virtual audio renderings. This was achieved by employing two experimental blocks, one at the beginning and one at the end of the experiment, in which triplets of speech stimuli consisted of both real and virtual sound sources, and another longer block, in the middle, where only virtual sound sources were presented. Because none of the experimental designs associated with the existing paradigms allowed to examine this effect, a variation of the typical binary (yes/no) test was employed.

2. EXPERIMENT

2.1 Binaural rendering

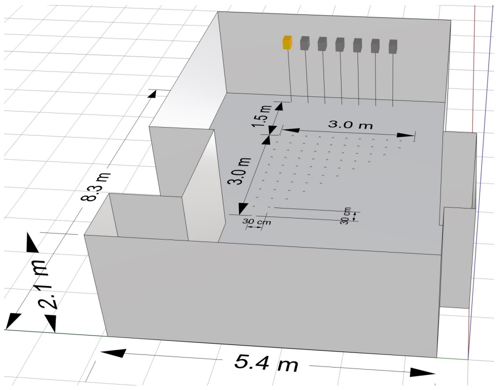
2.1.1 Spatial room impulse responses measurement

To capture the room acoustics and reproduce it binaurally, spatial room impulse responses (SRIRs) were first measured in the room where the experiment took place. The room, located in the Dyson School of Design Engineering building at Imperial College London, has approximate dimensions 5.4 m × 8.3 m × 2.1 m and is relatively reverberant (the reverberation time is shown in Fig. 1b). The SRIRs were measured at 66 positions using a spherical microphone array (em32 Eigenmike, MH Acoustics, USA) for seven sound source positions. Each sound source emanating from an active loudspeaker (Genelec 8010A, Finland) played back a logarithmic sine sweep of 2.730 s spanning frequencies from 0 Hz to 24 kHz (a fade-in and fade-out of 50 samples was applied to avoid distortions), although a single loudspeaker position was employed for this specific experiment (the orange coloured rectangular prism in Fig. 1a). The loudspeakers were connected to an audio interface (Antelope Audio, Bulgaria) which was plugged to a MacBook Pro laptop (Apple, USA). The microphone array was connected to a USB audio interface (Madiface Pro, RME, Germany), which was also connected to the laptop. The sine sweeps playback/recording/deconvolution was controlled via a custom-made Max/MSP (Cycling '74, USA) patch. During the measurements, the centre of the microphone array and the loudspeakers were located at a height of 1.55 m from the floor level. For each of the 66 measured positions, the SRIR was first denoised (detailed information on the denoising method can be found in [9]) and then encoded to Ambisonics to result in a 4th order Ambisonic signal.

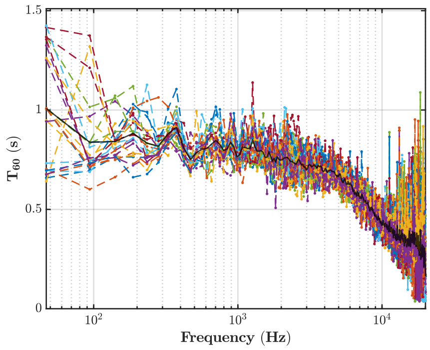
2.1.2 Hardware and software

The participants were wearing a pair of extra-aural headphones (AKG K1000, Austria) onto which a Vive Tracker (HTC Corp., Taiwan) was mounted to allow tracking of the head position and rotations (in the x-, y-, z-, pitch-, yaw-, roll-axis). The tracking data was sent via open sound control (OSC) messages from Unity to a custom-made Max/MSP patch (different from the one used during the SRIRs measurement) running on a 64-bit Windows 11 laptop (Lenovo Legion 5, Beijing, China).

The binaural renderings were produced in real time by using a combination of the *6DoFconv* and the *ambiBIN* plug-ins from SPARTA [10] in the Max/MSP patch. The *6DoFconv* plug-in executed the convolution between the test stimuli (details in Sec. 2.2.2) and the SRIR depending on the participant's position on the 66-position grid in the room (more detailed information on this plug-in can be found in [11]). The *ambiBIN* plug-in decoded the 4th order Ambisonic signal to binaural with the magnitude least-squares decoding approach [12]. The employed HRTFs for the binaural renderings were those of the KEMAR



(a)



(b)

Figure 1: (a) 3D model of the room where the SRIRs were measured. (b) Reverberation time, T_{60}, calculated from the 4th order Ambisonics signal of the SRIR measured at a single position (corresponding to the more distant source-receiver combination).

(GRAS Sound & Vibration, Denmark) dummy head. Before starting the experiment, the height of the loudspeakers was adjusted to match with this of the participant's ear canal entrances'. The audio stream (binaural or loudspeaker) was sent to an USB audio interface (Ultralite mk3, MOTU, USA) onto which was plugged a radio frequency transmitter (MEI 1000 G2, LD Systems, Germany). The radio frequency receiver was connected to a portable amplifier (EM2, Donner, China) into which the headphones were plugged. The participants held a mini keypad to both trigger the playback of the stimuli and to indicate their answer. The keypad activated a custom-made Python script which sent OSC messages to the Max/MSP patch to control which audio stream to render. Figure 2 shows a diagram describing the hardware and software used in the experimental setup.

2.2 Experimental design

2.2.1 Test design

As mentioned in Sec. 1, previous studies showed that a larger number of auditory illusions (i.e., believing that a virtual sound source is coming from the surrounding environment rather than from the headphones) [13] are created when evaluating the plausibility of fully dynamic AAR scenarios with either only virtual sound sources or more virtual sound sources than their real counterparts. Based on these results, this study aims at assessing how much and how fast such auditory illusions can be created by the mere fact of presenting virtual audio stimuli only in AR contexts. In other words, this study aims exploring the extent to which the participants can accommodate to virtual audio renderings.

Although a few paradigms exist (and as such, a few associated experimental designs) to evaluate the suitability of virtual audio renderings, none of the proposed ones allowed to test for this accommodation effect in an AR context where real and virtual sound sources overlap. Indeed, assessing authenticity is based on direct comparisons between real and virtual sound sources, and thus would not allow to solely present virtual sound sources to the participants. Plausibility is limited for AR contexts where both real and virtual sound sources can overlap since it is typically assessed by presenting one stimulus at a time to the participants. Finally, transfer-plausibility which was proposed to be assessed using a three-alternative forced-choice test design would also be incompatible with presenting virtual sound sources only. This is because such a test design requires the participants to identify which of the stimuli is virtual [2]. As such, if only virtual sound sources are presented, this question becomes inappropriate. For these reasons, the present experimental design employed a variation of the yes/no test in which a sequence of three stimuli (i.e., a stimulus triplet) was presented instead of one stimulus presented at a time.

The participants were presented with a series of 45 triplets of stimuli. The series was decomposed in three blocks which differed in terms of number of real sound

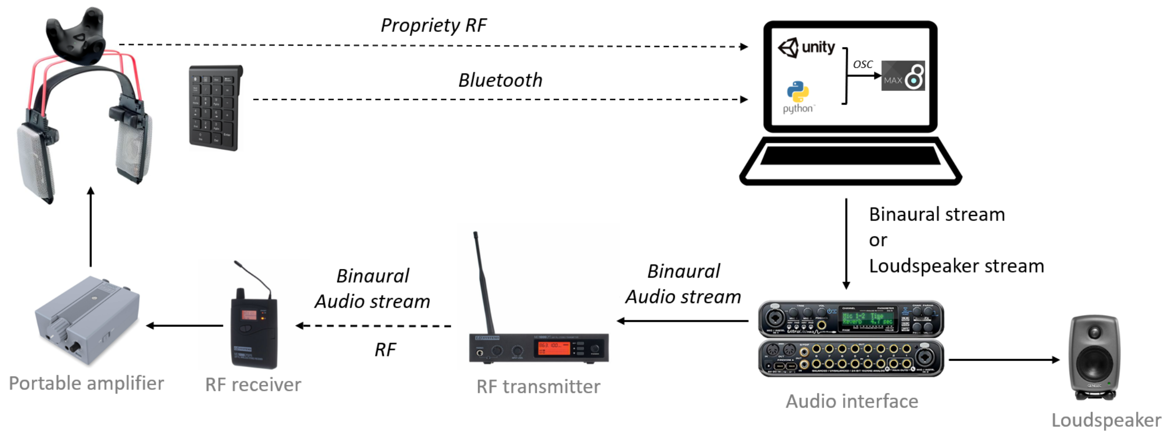


Figure 2: Diagram showing the hardware and software used in the experimental setup.

source present in the triplet, as indicated in Table 1. The task was to respond to the question "Is there a real sound source?" with a Yes or a No answer for each presented triplet. As mentioned in Sec. 2.1.2, they held a mini keypad to both trigger the playback of the sound sources and to indicate their answer using keys clearly labelled "Yes" and "No" on the keypad. Prior to providing their answer, they were required to listen to each stimulus forming the triplet at least once and the playback was limited to two times per stimulus. To rule out response bias or shifts in decision criteria, none of the participants was made aware of the organization of the three blocks of triplets, which implied, e.g., that they did not know real sound sources were absent in block 2. They were only informed that either one or no real sound source could be present within each triplet.

2.2.2 *Stimuli*

The stimuli were 25 randomly selected speech sentences (both female and male talkers) from the Harvard Sentences speech corpus [14]. The three stimuli constituting a triplet were randomly chosen among the 25 ones. Each time a triplet was formed, a new random selection of a stimulus triplet was done. The duration of each stimulus was on average 4 seconds. The series of 45 stimulus triplets was completed for each of the 3DoF and 6DoF conditions, totalling 90 answers per participants. The stimuli were first filtered to compensate for differences in the frequency response of the audio processing chain of the loudspeaker and headphones renderings. All the stimuli were then normalized in amplitude and their reproduction level adjusted so that the sound pressure level measured at 1 meter away from (and facing) the loudspeaker was approximately 65 dB(A). The reproduction level of the binaural renderings was calibrated at the same level using a binaural microphone (EARS, miniDSP, Hong Kong).

Table 1: Block structure of the experimental design.

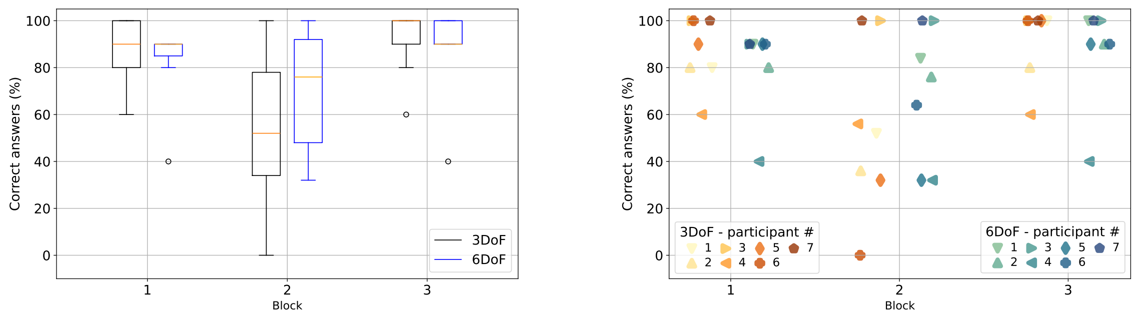
	Block 1	Block 2	Block 3
Number of triplet	10	25	10
Number of real sound source	1	0	1
Number of virtual sound source	2	3	2

2.2.3 *Participants*

Seven participants (1 female, 6 males) having self-reported normal hearing took part in this study. The participants were between 22 to 32 years of age (mean = 27 years, standard deviation = 3 years) and had little to some previous experience in participating to listening experiments. Participation was voluntary and not compensated.

3. RESULTS AND DISCUSSION

The results are shown in Fig. 3. The participants took on average 11 and 10 minutes to complete the experiment in the 3DoF and 6DoF conditions,respectively. The Shapiro-Wilk tests run on the data shown in Fig. 3 indicate that some of the rates for correct answers are not normally distributed (results not shown), requiring therefore the use of non-parametric statistical methods.



(a) (b)

Figure 3: (a) Boxplot showing the percentage of correct answers per block of the experiment for each DoF condition, pooled across participants. (b) Scatter plot showing the percentage of correct answers per block of the experiment for each DoF condition, for each participant.

The Friedman test was employed (at the 5% significance level) on the percentage of correct answers obtained separately for each DoF condition. The tests' results indicate that the rates of correct answers are different in the three blocks of the experiment for the 3DoF ($\chi^2 = 9.294$, $p = 0.010$) and the 6DoF ($\chi^2 = 6.348$, $p = 0.042$) conditions. *Post-hoc* multiple comparison tests using the Tukey's honest significant difference procedure further indicate that the difference is between the blocks 2 and 3 for both DoF conditions.

Unlike the study from Neidhardt and Zerlik [7] in which the participants were presented with real sound sources in part II of the experiment (whereas only virtual sound sources were in part I), here the real sound sources were presented already in block 1. Despite the medians being lower in block 2 compared to block 1 for both DoF conditions (see Fig. 3a), there is no statistical difference between the rates of correct answers in the two first blocks. As such, it could be argued that the accommodation effect was mitigated by the availability of the real sound source in block 1. As for comparing block 2 and 3 together, the present results are comparable with those from [7] in that the availability of a real sound source significantly increased the correct answer rates compared to presenting virtual stimuli only.

It is possible that the accommodation effect that this study attempted to examine was partly due to expectation effects of having equal numbers of trials containing 0 or 1 real sound source. Despite the relatively short times the participants took to complete the experiment, fatigue could have also played a role in the judgements. Nonetheless, this type of experiment could be useful to evaluate and compare the suitability of virtual audio rendering approaches for ecologically-valid audio augmented applications. Indeed, it is herein hypothesised that a non-suitable virtual audio rendering approach would lead to a high percentage of correct answers, whereas a suitable one would produce a low percentage of correct answers in block 2. The question of how to determine what is high or how remains however open, and could be specifically determined by looking at the performance of each participant.

Future work will include a detailed analysis of the block which contained virtual sound sources only, as it is hypothesised that during that block the percentage of correct identification of the real source is gradually becoming lower trial after trial. This is because of the adaptation process which, rather than helping in the correct identification of the real source, is tricking the brain in believing that a virtual source is real. Future work should also focus on introducing different sound source spatial locations, as done in the experimental design evaluating the concept of transfer-plausibility [2]. Additional work should also be done to refine the current test design such that more appropriate statistical analysis can be applied. In fact, for typical yes/no test designs, signal detection theory (e.g., [15]) can be used to evaluate the sensitivity index and the decision criterion, which could be better indicators of the discriminability between real and virtual reproduction than the percentage of correct responses. Furthermore, and as was measured in [16], it is expected that the results obtained in this study were limited by the acoustic alterations due to wearing the headphones, since these were not accounted for in the binaural rendering. Finally, including sound source directivity could be examined to potentially create more convincing blend between the real and virtual sources.

4. CONCLUSION

This study investigated the perception of real and virtual sound sources in a reverberant environment with two different degrees of movements allowed: 3DoF with head movements only and 6DoF with translation in addition to head movements. Based on the premises that humans with normal hearing are able to adapt to certain cues remapping (such as adapting to altered localization cues), the experiment examined the accommodation effect to listening to virtual audio rendering. Results showed that the availability of a real sound source in block 1 mitigated the accommodation effect in that the percentage of correct answers in detecting the presence of the real sound source was not different across the two first blocks. However, after listening to only virtual sound sources, the participants were able to correctly detect the real sound source with higher rates.

5. ACKNOWLEDGMENTS

The authors would like to thank Rapolas Daugintis from Imperial College London for his help in re-compiling the *6DoFconv* plug-in. This project has received funding from the SONICOM project (www.sonicom.eu), a European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017743.

6. REFERENCES

[1] S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg, "A summary on acoustic room divergence and its effect on externalization of auditory events," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, IEEE, 2016.

[2] N. Meyer-Kahlen, S. Schlecht, S. V. A. Garí, and T. Lokki, "Testing Auditory Illusions in Augmented Reality: Plausibility, Transfer-Plausibility, and Authenticity," *Journal of the Audio Engineering Society*, vol. 72, no. 11, pp. 797–812, 2024.

[3] F. Brinkmann, A. Lindau, and S. Weinzierl, "On the authenticity of individual dynamic binaural synthesis," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 1784–1795, 2017.

[4] A. Lindau and S. Weinzierl, "Assessing the plausibility of virtual acoustic environments," in *Forum Acusticum*, European Acoustics Association, 2011.

[5] S. A. Wirler, N. Meyer-Kahlen, and S. J. Schlecht, "Towards transfer-plausibility for evaluating mixed reality audio in complex scenes," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2020.

[6] N. Meyer-Kahlen, S. A. Garí, T. McKenzie, S. Schlecht, and T. Lokki, "Transfer-plausibility of binaural rendering with different real-world references," in *Jahrestagung für Akustik*, 2022.

[7] A. Neidhardt and A. M. Zerlik, "The availability of a hidden real reference affects the plausibility of position-dynamic auditory AR," *Frontiers in Virtual Reality*, vol. 2, p. 678875, 2021.

[8] V. Martin, J. Meyer, and L. Picinali, "Exploring dynamic binaural rendering approaches in 6DoF acoustic augmented reality," in *DAGA 2025. Jahrestagung für Akustik*, Deutsche Gesellschaft für Akustik eV (DEGA), 2025.

[9] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Denoising directional room impulse responses with spatially anisotropic late reverberation tails," *Applied Sciences*, vol. 10, no. 3, p. 1033, 2020.

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

[11] T. McKenzie, N. Meyer-Kahlen, R. Daugintis, L. McCormack, S. Schlecht, and V. Pulkki, "Perceptually informed interpolation and rendering of spatial room impulse responses for room transitions," in *International Congress on Acoustics*, pp. 1–11, Acoustical Society of Korea (ASK), 2022.

[12] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proceedings of the DAGA*, vol. 44, pp. 339–342, 2018.

[13] K. Brandenburg, F. Klein, A. Neidhardt, U. Sloma, and S. Werner, "Creating auditory illusions with binaural technology," *The Technology of Binaural Understanding*, pp. 623–663, 2020.

[14] "IEEE Recommended Practice for Speech Quality Measurements," *IEEE No 297-1969*, pp. 1–24, 1969.

[15] D. M. Green, J. A. Swets, *et al.*, *Signal detection theory and psychophysics*, vol. 1. Wiley New York, 1966.

[16] C. Schneiderwind, A. Neidhardt, and D. Meyer, "Comparing the effect of different open headphone models on the perception of a real sound source," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2020.