Acta Acustica 2021, **5**, 55

https://doi.org/10.1051/aacus/2021048

Available online at:
https://acta-acustica.edpsciences.org

SCIENTIFIC ARTICLE OPEN ACCESS

Do near-field cues enhance the plausibility of non-individual binaural rendering in a dynamic multimodal virtual acoustic scene?

Johannes M. Arend[1,2,a,*], Melissa Ramírez[1,2,a], Heinrich R. Liesefeld[3], and Christoph Pörschmann[1]

[1] Institute of Communications Engineering, TH Köln – University of Applied Sciences, Betzdorfer Str. 2, 50679 Cologne, Germany
[2] Audio Communication Group, Technical University of Berlin, Einsteinufer 17c, 10587 Berlin, Germany
[3] Department of Psychology, University of Bremen, Hochschulring 18, 28359 Bremen, Germany

Received 27 April 2021, Accepted 8 November 2021

Abstract – It is commonly believed that near-field head-related transfer functions (HRTFs) provide perceptual benefits over far-field HRTFs that enhance the plausibility of binaural rendering of nearby sound sources. However, to the best of our knowledge, no study has systematically investigated whether using near-field HRTFs actually provides a perceptually more plausible virtual acoustic environment. To assess this question, we conducted two experiments in a six-degrees-of-freedom multimodal augmented reality experience where participants had to compare non-individual anechoic binaural renderings based on either synthesized near-field HRTFs or intensity-scaled far-field HRTFs and judge which of the two rendering methods led to a more plausible representation. Participants controlled the virtual sound source position by moving a small handheld loudspeaker along a prescribed trajectory laterally and frontally near the head, which provided visual and proprioceptive cues in addition to the auditory cues. The results of both experiments show no evidence that near-field cues enhance the plausibility of non-individual binaural rendering of nearby anechoic sound sources in a dynamic multimodal virtual acoustic scene as examined in this study. These findings suggest that, at least in terms of plausibility, the additional effort of including near-field cues in binaural rendering may not always be worthwhile for virtual or augmented reality applications.

Keywords: Binaural rendering, Nearby sound sources, Near-field head-related transfer functions, Plausibility, Multimodal environment

1 Introduction

Auditory distance perception is dominated by intensity cues [1, 2]. In reverberant environments, distance judgments are aided by changes in the direct-to-reverberant energy ratio (DRR) [1, 2], and for far-away sources (more than 15 m), high-frequency attenuation provides additional spectral cues [1, 2]. Sound sources in the proximal region[1], i.e., at distances within 1 m of the head center [3], provide further specific distance cues. In particular, interaural level differences (ILDs) exhibit significant distance-dependent changes for lateral sources. Interaural time differences (ITDs), on the other hand, are nearly independent of distance. Both effects were demonstrated by analyses of measured near-field head-related transfer functions (HRTFs) [3, 5]. Brungart [6] suggested that in the absence of the powerful intensity cue, low-frequency ILD cues ($f < 3$ kHz) dominate distance perception of nearby lateral sources in anechoic conditions. Studies by Kopčo et al. on intensity-independent distance perception of nearby sound sources in reverberant conditions yielded inconsistent results, indicating that either the DRR cue masks the ILD cue [7], or that both ILD and DRR cues support distance estimation [8]. Therefore, the relative contribution of the ILD and DRR cues to intensity-independent distance perception is currently not fully understood [9]. Furthermore, nearby sound sources show a relative emphasis of low-frequency sound pressure due to acoustic scattering by the head and torso, resulting in a low-pass filtering character that might be a spectral cue for distance estimation in the near field [1, 3]. The acoustic parallax effect may also affect perception and distance estimation of nearby sound sources [1, 2]. This effect occurs because close sources cause a significant difference between the angle of the source relative to the left or right ear, resulting in a lateral shift of some of the high-frequency features of the HRTF [1].

*Corresponding author: Johannes.Arend@th-koeln.de
[a] Johannes M. Arend and Melissa Ramírez contributed equally to this work.
[1] In the following, we also use the term *near field* to refer to the proximal region [3] or peripersonal space [2], i.e., the area within 1 m of the listener's head center, rather than to describe the frequency-dependent acoustic near field in the sense of physical acoustics [1, 4].

As briefly outlined above, previous research mainly focused on distance estimation accuracy of nearby sound sources and has obtained partly conflicting results regarding the contribution of the various near-field cues to distance perception [1, 2, 6, 10]. A recent study also investigating the influence of binaural cues on distance estimation of nearby sound sources reviews several studies on this topic and discusses the differing results [11]. Further studies in virtual acoustics that used near-field HRTFs synthesized from far-field HRTFs by applying distance variation functions (DVFs) also exclusively evaluated the influence of the synthesized near-field cues on distance estimation accuracy [12, 13]. Moreover, many of the above-mentioned studies tested distance estimation accuracy under unimodal (audio-only) conditions. In most of them, listeners had a passive role (i.e., they could not interact with the sound scene) and had to judge the distance of stationary or dynamic sound events, which, if the study was conducted in virtual acoustics, were even often reproduced with static binaural synthesis only (see, e.g., Arend et al. [11] for an overview). To better evaluate individual auditory distance cues, these methods often attempted to eliminate other cues (e.g., the intensity cue by level normalization), resulting in unnatural stimuli. Thus, whereas such experimental methods are well suited to understand the contribution of individual distance cues to distance perception and how they interact with each other, they do not ideally reflect the way humans perceive their multimodal environment and estimate, for example, the distance to a (nearby) sound source in real-life.

Rummukainen et al. [14] presented the only study we are aware of that investigated perceptual aspects of near-field HRTFs beyond distance estimation accuracy, and that was conducted in a six-degrees-of-freedom (6-DoF) multimodal virtual reality (VR) environment, thereby including visual and proprioceptive cues in addition to auditory cues. In their experiment, listeners either actively moved around a static virtual sound source or dynamically moved the virtual sound source around their head. The participants' task was to rate binaural renderings based on intensity-scaled far-field HRTFs or multi-distance near-field HRTFs, among others, according to their preference. Surprisingly, listeners liked both HRTF types equally. However, the authors pointed out that further studies are needed, especially as the closest distance examined in their study was 0.50 m, which means that the strongest near-field cues were not present.

Thus, whereas it is generally assumed that including near-field cues in binaural rendering leads to a more realistic reproduction, and especially experienced listeners often report that near-field effects are subjectively audible, studies such as Rummukainen et al. [14] raised first doubts on the perceptual importance of near-field HRTFs in multimodal environments. However, to the best of our knowledge, no study has examined yet whether using near-field HRTFs for binaural rendering in a dynamic multimodal scene enhances the *plausibility* [15] of the virtual acoustic environment (VAE) compared to using intensity-scaled far-field HRTFs, i.e., whether using near-field HRTFs results in a binaural reproduction of nearby sound sources that, based on the listener's *inner reference* and personal experience, is more in agreement with their expectation towards the corresponding real event than binaural rendering using intensity-scaled far-field HRTFs.

The plausibility of virtual environments has been discussed extensively in the literature of various research areas, and the above-mentioned definition by Lindau & Weinzierl [15] is in line with what Slater [16] referred to as *plausibility illusion* and Hofer et al. [17] recently described as *external plausibility*. Essentially, external plausibility refers to how *consistent* the virtual environment is with the users' real-world knowledge [17], and whether an event in the virtual environment could actually occur in the real world [16]. Thus, external plausibility is expressed by the user (or more precisely, in this case, the listener) judging something in the virtual environment to be factually true or accurate, or by events in the virtual environment to be highly likely or typical of the real world [17].

Assessing the plausibility of a VAE provides, therefore, a comprehensive measure for the quality of the virtual presentation that includes various perceptual factors. It is an important perceptual criterion for VR and augmented reality (AR) applications, as its assessment also examines how the acoustic representation agrees with other modalities of the virtual scene (e.g., visual, haptic, or proprioceptive) and whether there are no apparent contradictions between the modalities that would reduce or even break plausibility [18]. As such, plausibility has recently become a popular measure for the perceptual evaluation of VR and AR audio applications.

However, the various audio and acoustic studies that have assessed the plausibility of VAEs often differ in their experimental methods and procedures. Lindau & Weinzierl [15] proposed a test paradigm in which either a real (loudspeaker reproduction) or a virtual (binaural reproduction) stimulus is presented in each trial, and the participants have to decide in a yes/no task whether the stimulus comes from a real loudspeaker or a virtual representation of the loudspeaker. Some studies followed this procedure, in which the stimulus is presented either through a real loudspeaker or binaurally through headphones, for example, to evaluate the plausibility of pseudobinaural recordings [19], or 6-DoF parametric binaural rendering [20]. However, the test paradigm proposed by Lindau & Weinzierl [15] has also been adapted (by the same research group) to assess the plausibility of room acoustic simulations. In the study by Brinkmann et al. [21], there was no real source serving as an explicit reference. Instead, listeners were presented with either simulation- or measurement-based auralizations (only virtual stimuli) and had to rate whether the stimuli correspond to a real room. In line with this, several other approaches have been proposed to assess the plausibility of VAEs in cases where no real counterpart is available to use as an explicit reference. For example, Neidhardt et al. [22, 23] evaluated the plausibility of position-dynamic virtual acoustic realities in which listeners move towards a virtual sound source using either a continuous or ordinal plausibility rating scale. Amengual Garí et al. [24] evaluated

the plausibility of 3-DoF parametric binaural rendering using a two-alternative forced-choice (2AFC) procedure. Either both stimuli were virtual, or one of them was a real loudspeaker, and participants had to rate which of the two stimuli they perceived as more plausible. Most recently, Neidhardt & Zerlik [25] conducted two experiments to assess the plausibility of position-dynamic binaural rendering using a yes/no task. In one experiment, participants were presented with virtual stimuli only, whereas in another experiment, participants were presented with either real or virtual stimuli. The authors concluded that because of their different advantages and disadvantages, both methods are relevant and valid for assessing the plausibility of a VAE.

Surprisingly, even though very recent research such as VRACE [26] focuses on binaural rendering in the near field, and although several binaural renderers use near-field cues or respectively (synthesized) near-field HRTFs to reproduce nearby sound sources (e.g., the commercially available renderers from Oculus [27], MagicLeap [28], and Resonance Audio [29] as well as the open-source renderers Spat [30], Anaglyph [31], or 3DTI Toolkit [32]), it is still unknown whether binaural rendering with near-field HRTFs increases the plausibility for naive (non-expert) listeners compared to a much easier to implement rendering with intensity-scaled far-field HRTFs. However, it is crucial to know whether the additional computing effort of including near-field cues is worthwhile in terms of plausibility and overall reproduction quality, especially for complex real-time applications with limited computing resources, such as mobile AR applications with 6-DoF.

To close this gap and investigate whether near-field HRTFs provide a more plausible binaural reproduction of nearby sound sources than intensity-scaled far-field HRTFs, we performed two listening experiments in an anechoic 6-DoF VAE. In both experiments, participants controlled the position of a virtual sound source by moving a small handheld loudspeaker, which provided visual and proprioceptive cues in addition to the auditory cues and aided the application-oriented AR experience. In a 2AFC procedure, the participants had to compare non-individual anechoic binaural renderings based on either synthesized near-field HRTFs or intensity-scaled far-field HRTFs and judge which of the two rendering methods led to a more plausible representation, i.e., which one was more congruent with their expectations based on the visual- and haptic sensation as well as based on their inner reference and personal experience. We hypothesized that they would rate the renderings using near-field HRTFs as more plausible, as this reproduction method yields a more physically correct representation of nearby sound sources.

We employed a multimodal sensory-motor test paradigm where participants moved the sound source because this results in a more natural scenario that better emulates the way humans perceive their environment than the more extensively investigated unimodal passive paradigms. Besides, previous research showed that multisensory stimulation improves sound localization [33, 34]. Recent findings by Valzolgher et al. [35] also indicated that kinesthetic cues resulting from moving a sound source with one's own hand could contribute to the updating of spatial hearing and thus improve sound localization performance. In this line of thinking, providing more reliable (real) visual, motor, and proprioceptive information simultaneously, together with the (simulated) auditory information, should help listeners optimally associate auditory cues to the spatial location of a sound source [34, 35]. Thus, the multimodal virtual environment employed in our study should (1) facilitate auditory localization and (2) provide the listener with more information to assess the plausibility of a virtual sound source more reliably than is possible in a unimodal environment. Listeners were able to judge the plausibility of the binaural renderings based not only on their inner reference and listening experience but also on the simultaneous real information (visual, motor, and proprioceptive). This aided the identification of possible discrepancies between the real and virtual worlds and thus detecting breaks in plausibility.

The two experiments, each performed with a different group of subjects, differed only regarding the test signal used. In Experiment 1, we used pink noise bursts to provide extremely critical and ideally controllable stimuli that clearly reveal all near-field cues. Then, to generalize the results of Experiment 1 to a more application-oriented setup, we used female speech as a test signal in Experiment 2.

2 Experiment 1

2.1 Method

2.1.1 Participants

Sixteen participants (ages 22–55 years, $M = 33.3$ years, $Mdn = 28$ years, $SD = 11.2$) with self-reported normal hearing took part in the experiment on a voluntary basis. Four of the participants are members of our laboratory and therefore classified as expert listeners. The remaining participants were engineering students or research assistants from other laboratories at the university and classified as naive listeners. All participants were naive as to the purpose of the study.

2.1.2 Setup

The experiment took place in the sound-insulated anechoic chamber of TH Köln, which provided the appropriate acoustic environment for the anechoic binaural renderings simulating the handheld loudspeaker. The experiment was implemented, controlled, and executed by a purpose-built Python application running on a PC. For real-time dynamic binaural synthesis, we employed the open-source tool PyBinSim [36] in combination with a pair of HTC VIVE trackers (update rate of 120 Hz). One tracker was mounted on the headphones (Sennheiser HD600), and the other tracker was attached to the handheld loudspeaker (JBL Clip+), providing 6-DoF tracking data of both. Based on the tracking data, the Python application calculated the loudspeaker's azimuth, elevation, and distance relative to

the participant's head orientation and position and sent these spherical coordinates to PyBinSim by Open Sound Control (OSC) messages. The application also used OSC messages to control the renderer, e.g., to start and stop audio playback or to change between HRTF datasets. Additionally, the application logged the relative tracking data at a sampling rate of 30 Hz.

The graphical user interface of the application was presented on a screen located at a distance of about 2 m in front of the seated participant. A Numark Orbit MIDI controller served as the input device for the participants' responses. We used an RME Babyface audio interface as digital-to-analog converter and headphone amplifier at 48 kHz sampling rate and a buffer size of 64 samples. The separate buffer of PyBinSim was set to 128 samples.

2.1.3 Materials

We employed measured far-field HRTFs from a Neumann KU100 dummy head [37], a dataset widely used in both commercial applications and research. The HRTF set was transformed to the spherical harmonics (SH) domain at a sufficiently high spatial order of $N = 44$, allowing artifact-free SH interpolation to obtain HRTFs for any desired direction, which was necessary in the present case for accurate HRTF synthesis. Both the intensity-scaled far-field HRTFs as well as the near-field HRTFs were synthesized for distances from 0.12 m to 1.20 m in steps of 1 cm on a spatial sampling grid with a resolution of 1° in the horizontal direction and 5° in the vertical direction, limited to ±15° in elevation.

The near-field HRTFs were synthesized by applying distance variation functions (DVFs) to the far-field HRTFs [12]. The DVFs were generated from a spherical head model [38] with the ears positioned at azimuth $\phi = \pm 90°$ and elevation $\theta = 0°$. The optimal head radius of the spherical head model was 9.19 cm, calculated according to Algazi et al. [39] based on the dimensions of the Neumann KU100 dummy head. In general, DVFs are calculated for each distance and direction as the ratio of the pressure on the sphere emanating from a sound source at a desired distance in the near field to the pressure on the sphere emanating from a sound source in the far field, with the pressure on the sphere evaluated solely at the ear positions. Thus, a DVF approximates the changes of an HRTF as a sound source varies in distance, such as alterations in intensity and spectrum or frequency-dependent changes in ILD. Additionally, a cross-ear parallax correction was applied [40] to account for high-frequency parallax effects induced by the pinna, which the DVF is unable to take into account [12]. Appropriate far-field HRTFs for the left and right ear are first selected for the respective distance and direction (using SH interpolation) based on a geometric parallax model and then filtered with the corresponding DVFs, resulting in the desired near-field HRTFs. The described processing, which is similar to the implementation in state-of-the-art renderers such as Spat, Anaglyph, or 3DTI Toolkit, was performed using the `supdeq_dvf` function of the `SUpDEq toolbox`[2].

To synthesize the intensity-scaled far-field HRTFs, an HRTF set was first obtained by SH interpolation according to the spatial sampling grid, and then its level was matched to that of the near-field HRTF set for the highest distance of 1.20 m. This set was then adjusted in level according to the inverse-square law to generate the HRTFs for closer distances. Thus, the intensity-scaled far-field HRTFs do not contain any of the prominent near-field cues included in the synthesized near-field HRTFs, such as the significant increase in (low-frequency) ILD for lateral sources, the low-pass filtering character, and the parallax effects.

Figure 1 (left) shows the low-frequency ($f < 3$ kHz) horizontal plane ILDs (which Brungart [6] suggests are the dominant auditory distance cue in the near field) for the intensity-scaled far-field HRTF sets (FF) and synthesized near-field HRTF sets (NF) at selected distances. As expected, the ILDs of the near-field HRTFs for lateral sources increase strongly with decreasing distance, especially for close distances (less than 0.50 m). The right plot in Figure 1 shows the corresponding ITDs, which, as expected, are nearly distance-independent and therefore almost the same for all HRTF sets. Figure 2 further shows the frequency-dependent behavior of the ILDs of the synthesized near-field HRTF sets as a function of distance. Consistently, the synthesized near-field HRTFs show strong low-frequency ILDs for lateral directions at close distances and a significant increase in ILD with increasing frequency. Overall, the described characteristics of the synthesized HRTFs are very similar to those of measured near-field HRTFs [5, 11, 41], confirming that the synthesis yields correct results. In particular, the low-frequency horizontal plane ILDs and ITDs of the synthesized near-field HRTFs are nearly identical to those of measured Neumann KU100 near-field HRTFs from [5] (see Fig. S1 in the Supplementary Material [42]), further supporting the excellent performance of the synthesis.

The test signal was a 10 s long sequence of 500 ms pink noise burst (including 10 ms cosine-squared onset/offset ramps) with an interstimulus interval of 150 ms. Broadband noise bursts are well-suited test signals to examine coloration and localization, so they were ideal for the present experiment. The sequence length of 10 s provided sufficient time to move the loudspeaker along the prescribed trajectory (see procedure in Sect. 2.1.4). To minimize the influence of the Sennheiser HD600 headphones, a generic headphone compensation filter was used. The filter was based on 12 measurements in which the headphones were put on and off the Neumann KU100 dummy head (the same one used to measure the far-field HRTFs employed in the present study) to account for re-positioning variability. The final filter was designed by regularized inversion of the complex mean of the headphone transfer functions [43] using the implementation by Erbes et al. [44]. Furthermore, to enhance the virtual acoustic representation of the handheld JBL Clip+ loudspeaker, a filter describing its on-axis frequency response was designed. The magnitude

[2] Available: https://www.github.com/AudioGroupCologne/SUpDEq

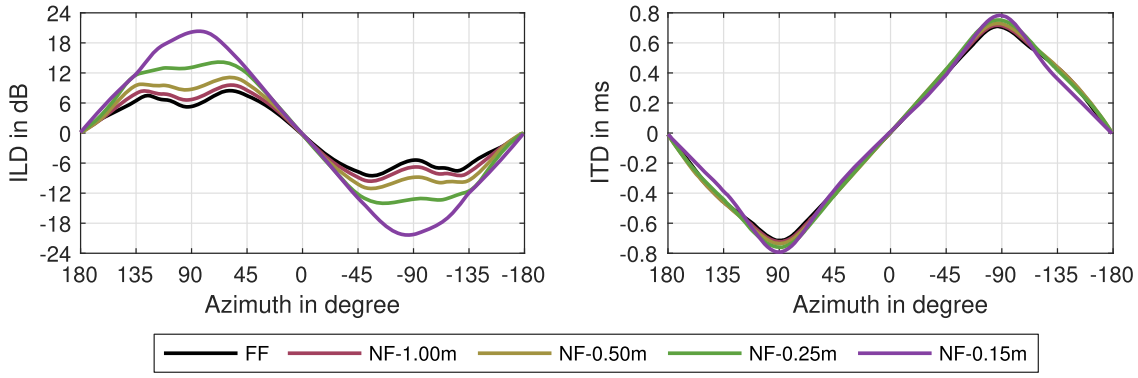

Figure 1. Low-frequency ($f < 3$ kHz) horizontal plane ILDs (left) and ITDs (right) of the intensity-scaled far-field HRTF sets (FF) and synthesized near-field HRTF sets (NF) at selected distances.

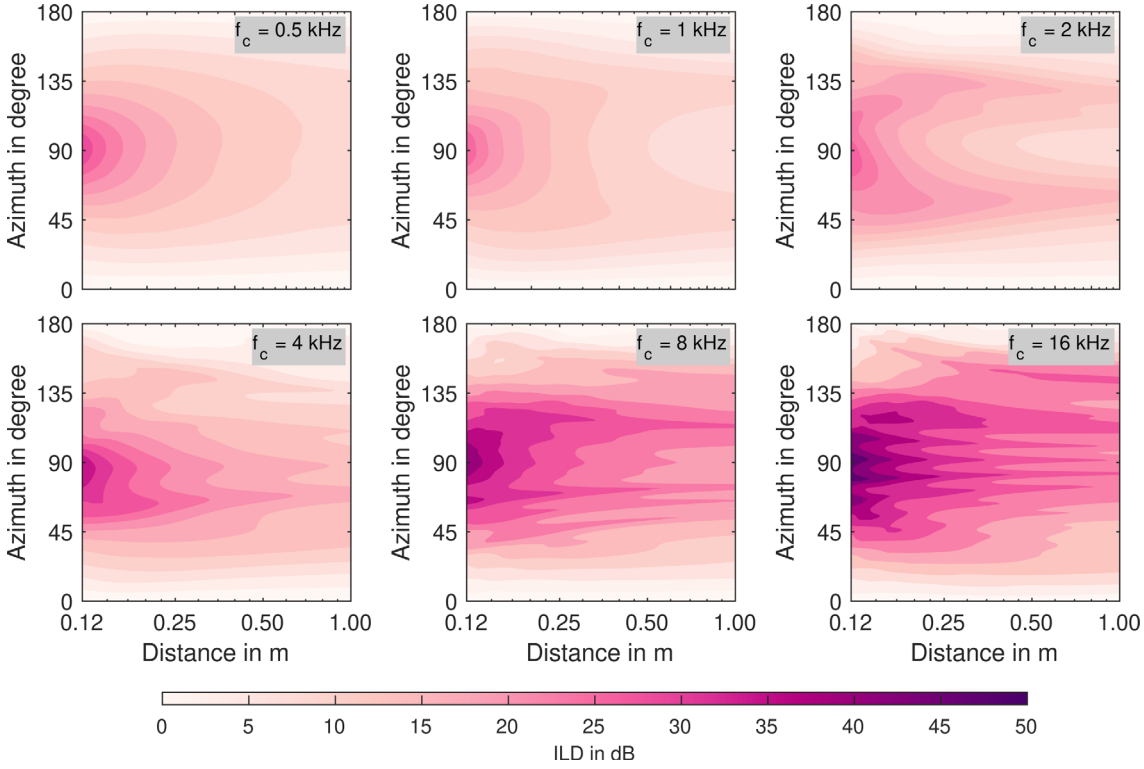

Figure 2. Horizontal plane ILDs (left hemifield) of the synthesized near-field HRTF sets as a function of distance for octave bands from 0.5 kHz to 16 kHz. Distances are shown on a logarithmic scale for a more detailed representation of the ILDs at close distances.

responses of both filters were combined to one minimum-phase finite impulse response (FIR) filter with 2048 taps, which was applied to the test signal. For more technical details, Figure S2 in the Supplementary Material [42] shows the magnitude response of the employed headphone compensation and loudspeaker filter. Informal evaluations showed that the 200 Hz low-cut of the loudspeaker filter does not affect the (binaural) near-field cues of the synthesized HRTFs. However, in pilot studies, we found that applying the filter is essential for the plausibility of the multimodal scene. Filtering out the low frequencies of the stimuli aligns the auditory impression with the visual impression of a small handheld loudspeaker. Besides, to foster reproducible research, we provide as well in the Supplementary Material [42] the Matlab script developed to synthesize the near- and far-field HRTFs, design the filters, and generate the filtered test signal.

To measure the presentation level produced over the headphones, a loudspeaker in the free field was leveled so that the playback of stimuli for frontal sound incidence produced the same electrical level at a dummy head as their playback over the headphones on the dummy head. The presentation level was then measured as the loudspeaker's equivalent free-field sound pressure level directly at the

dummy head's ear. Following this procedure, we estimated the presentation level for different conditions (without roving, meaning at a roving level of 0 dB; see roving procedure described in Sect. 2.1.4). The measured presentation level of the far-field condition for frontal sound incidence was $L_{\mathrm{Aeq}} = 49.3$ dB for a distance of 1.00 m and $L_{\mathrm{Aeq}} = 69.6$ dB for a distance of 0.12 m. The highest presentation level was $L_{\mathrm{Aeq}} = 84.5$ dB, measured for lateral sound incidence at the closest distance (0.12 m) in the near-field condition.

2.1.4 Procedure

Participants directly compared dynamic binaural renderings based on the intensity-scaled far-field HRTFs with renderings based on the synthesized near-field HRTFs in a 2AFC procedure. Each of the 100 trials in total consisted of a sequence of two 10 s intervals with an interstimulus interval of 0.5 s. The presentation order, i.e., whether the far-field or near-field rendering was presented first, was randomized. Moreover, the presentation level of each interval was randomly roved within a 10 dB range (±5 dB, steps of 1 dB, see, e.g., Kopčo & Shinn-Cunningham [7]) and participants were informed about that.

During the presentation of each interval, the participants were asked to move the handheld loudspeaker along a prescribed square-like trajectory to direct the virtual sound source through frontal and lateral areas near the head that yield strong near-field cues and thus clear differences between the rendering conditions. As a result, participants were exposed to all relevant auditory near-field cues: (1) frequent distance changes of the virtual sound source in lateral areas yielded strong variations in (low-frequency) ILD cues and distinct intensity cues, (2) movements of the virtual sound source from lateral to frontal areas very close to the head provided significant spectral, ILD, parallax, and intensity cues, and (3) frequent distance changes of the virtual sound source in frontal areas yielded strong spectral, parallax, and intensity cues.

After the presentation of both intervals, they were asked to select the interval which, as verbally instructed before the experiment, provided a more accurate representation of the expected sound field according to the sound source's positions and movements. In other words, participants had to choose the more plausible sound field representation based on their inner reference [15], life experience, and auditory, proprioceptive, and visual cues that emerged from actively moving the virtual source. The participants gave their answer by pressing a button on the MIDI controller. The answer was scored as correct when participants chose the near-field condition, following our initial hypothesis that using near-field HRTFs should be perceived as more plausible because it yields a more physically correct representation of nearby sound sources. Participants could neither repeat a trial nor continue without answering, and no feedback was provided. After an answer was registered, there was a 1 s silent pause before the next trial started. The procedure, including a presentation of the prescribed trajectory, is also illustrated in a short video, which is part of the Supplementary Material [42].

The 100 trials were split into two blocks of 50 trials with a short break in between to prevent fatigue. Before the experiment, participants were given instructions about the experimental procedure and they had to perform two training blocks to get familiar with the setup and the test procedure. In the first training block, participants were asked to practice moving the handheld loudspeaker along the prescribed trajectory. Their actual movement trajectory was displayed in real time on a computer screen so that they could visually monitor whether it conformed with the prescribed trajectory and adapt the movement trajectory based on this feedback if necessary. In the second training block, participants had to perform five trials of the experiment to practice the test procedure while still receiving the on-screen feedback. After the training, participants had no on-screen feedback on their movements to not distract them from the main task. A complete experimental session lasted about one hour, including the verbal instructions, the training blocks, and the short break.

2.2 Results and discussion

In informal post-experiment interviews, participants were asked whether the binaural reproduction was generally plausible regardless of the rendering condition. Overall, they experienced the scene as plausible, i.e., they perceived that the real loudspeaker emitted the sound, and they localized the virtual source at the position of the real loudspeaker. They reported that, in particular, moving the source and the congruence of visual, proprioceptive, and auditory cues supported the plausibility of the scene.

To verify that participants moved the loudspeaker mainly along the prescribed trajectory, we first analyzed the movement patterns based on the tracking logs. Figure 3 (left) shows the relative tracking data of Experiment 1, pooled over all participants and trials, in the form of a two-dimensional histogram. The plot shows the frequency distribution of the sound source position relative to the participants' head in the horizontal plane, defined by azimuth and distance. The prominent square-like movement pattern reflects the prescribed trajectory. As instructed, participants varied the distance of the virtual sound source to a large extent at frontal and lateral azimuth angles, which resulted in significant spectral (frontal) and ILD (lateral) changes in the near-field condition and intensity changes in both conditions. Furthermore, participants often placed the virtual source very close, both frontally and laterally, at distances between 0.20 m and 0.30 m. This also provided strong spectral (frontal) and ILD (lateral) cues in the near-field condition and thus significant differences to the far-field condition, at least from a signal-theoretic point of view. For a more detailed analysis, we provide plots of each participant's individual movement pattern in Figure S3 of the Supplementary Material [42].

Figure 4 (left) shows the results of the experiment in terms of individual p_{2AFC} values, their mean, and their 95% between-subject confidence interval (CI). The right plot of Figure 4 shows the interindividual variation in the determined p_{2AFC} values in the form of a box plot.

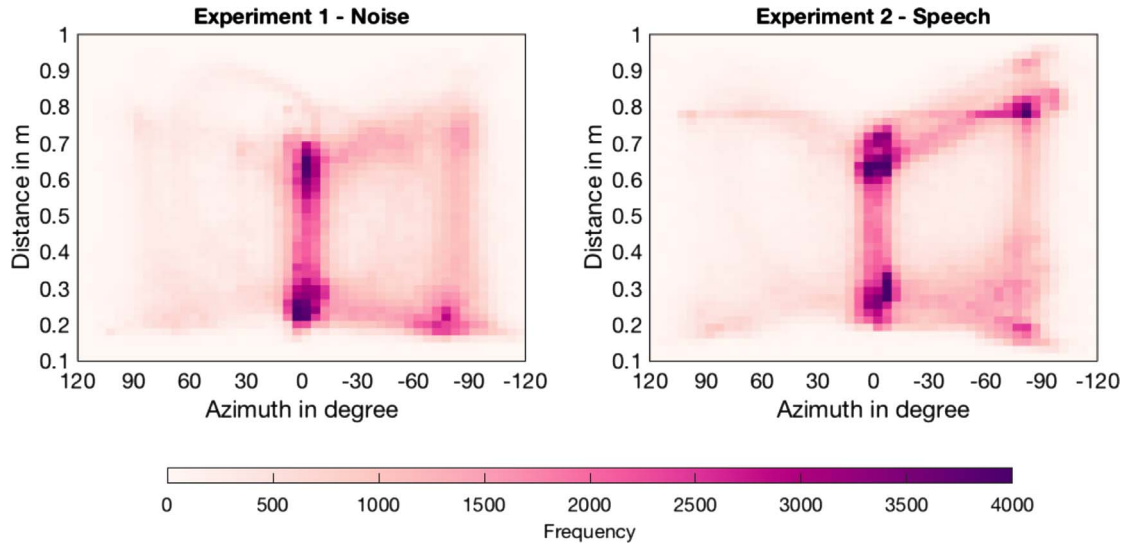

Figure 3. Two-dimensional histograms of the relative tracking data of Experiment 1 (left) and Experiment 2 (right), pooled over all participants and trials in the respective experiment.

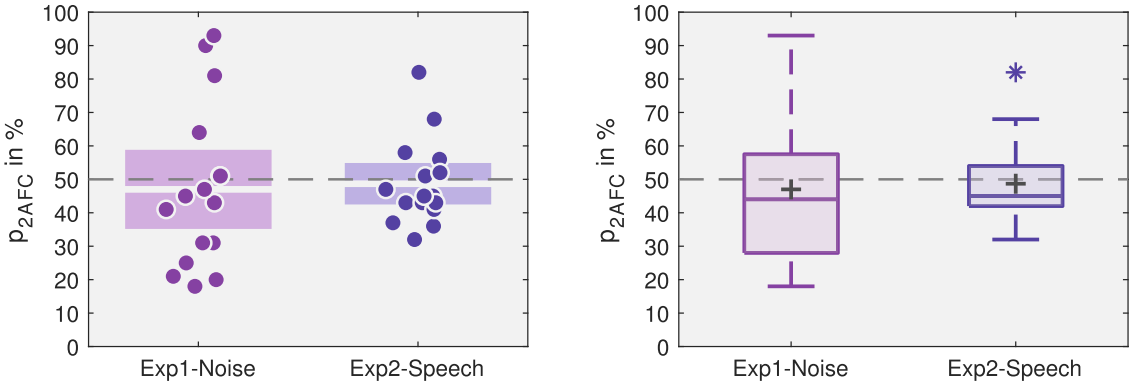

Figure 4. Results of the 2AFC test in Experiment 1 (Exp1-Noise) and Experiment 2 (Exp2-Speech). The left plot shows the determined individual percentages of correct answers p_{2AFC} as points (horizontal offset for better readability). The boxes show the mean (box notch) and the 95% between-subject CI. The gray dashed line denotes 50% chance level. The right plot shows the interindividual variation in the determined p_{2AFC} values in the form of a box plot with the median (box line), the mean (cross), and the (across participants) interquartile range (IQR); whiskers display 1.5 × IQR below the 25th or above the 75th percentile and outliers beyond that range are indicated by asterisks.

In general, the results exhibit high between-subject variance (see left plot in Fig. 4). Two participants, which both are expert listeners, performed exceptionally well (p_{2AFC} = 90% and 93%), but the majority of the participants either performed near 50% chance level or even clearly below chance. The findings suggest that the two participants strongly favored the near-field condition, whereas most other participants could not decide which condition was a more plausible reproduction (near chance performance), or even preferred the far-field condition over the near-field condition (below chance performance). Consequently, the mean and the median are slightly below chance level (see right plot in Fig. 4).

For statistical analysis of the results, we first applied a Lilliefors test for normality to the p_{2AFC} values, which showed no violations of normality (p = .151), indicating that parametric tests can be used. To analyze if the p_{2AFC} mean differs significantly from chance, we performed a one-sample t test against 50%. The test yielded no significant difference between the p_{2AFC} mean of 47% and chance level [$t(15)$ = 0.50, p = .626, d = .12]. As non-significant results of null-hypothesis significance testing cannot be interpreted as evidence for the absence of an effect, we also calculated the respective Bayes factor (BF_{01}, JZS scaling factor r = .707) for the one-sample t test. The obtained BF_{01} = 3.51 suggests that the data provide more than 3 times more evidence for the absence (rather than the presence) of an effect of near-field cues. Thus, the statistical results confirm that, on average, participants could not reliably decide which rendering method was more plausible, or in other words, on average, they found both rendering methods equally plausible.

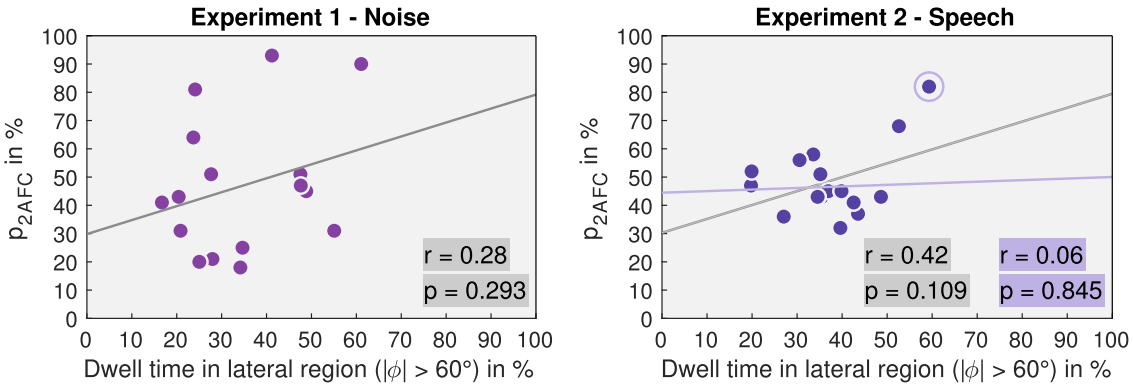

Figure 5. Participants' dwell time in lateral region ($|\phi| > 60°$) vs. their percentage of correct answers p_{2AFC} for Experiment 1 (left) and Experiment 2 (right). The solid line is the least-squares line of best fit. The right plot shows in purple the results of the correlation analysis for Experiment 2, excluding the outlier (data point circled in purple).

Next, we analyzed whether there is a correlation between participants' movement patterns and their plausibility estimates. For lateral source positions, the near-field HRTFs additionally exhibit strong ILD cues, resulting in particularly severe differences between the near- and far-field conditions. If these ILD cues affect listeners' preferences, there might be a correlation between the time subjects spend in lateral regions (*dwell time* in the following) and the percentages of correct answers. In other words, we examined whether participants who more often positioned the virtual source laterally perceived the near-field condition as more plausible. For this, we calculated the Pearson correlation between the participants' dwell time in the lateral region (proportion of relative tracking data with $|\phi| > 60°$, according to the definition of lateral positions by Brungart [6]) and the p_{2AFC} values. This yielded a non-significant positive correlation between dwell time and the p_{2AFC} values [$r(14) = .28$, $p = .293$], providing no evidence that participants who frequently positioned the virtual source laterally chose the near-field condition more often as the most plausible. Figure 5 (left) shows the corresponding scatter plot illustrating the relationship between both variables.

Finally, to determine whether plausibility ratings changed over the course of the experiment, e.g., because participants became tired or learned certain stimuli features, we analyzed the p_{2AFC} values in four epochs of 25 trials each. Figure 6 (left) shows the results of the experiment, divided among the four epochs. The plots suggest that participants remained fairly consistent in their answers over time. Thus, most participants who perceived the near-field condition as more plausible at the beginning of the experiment also did so throughout the experiment. The response behavior is similarly consistent for participants who preferred the far-field condition or perceived both conditions as equally plausible. In general, the between-subject variance seems to increase slightly over time, as participants who preferred the near- or far-field condition in particular became more stringent (more extreme) throughout the experiment, tending toward $p_{2AFC} = 100\%$ and $p_{2AFC} = 0\%$, respectively.

Statistical analysis of the data concerning the factor epoch showed no significant effect, suggesting that neither learning nor fatigue effects had a systematic impact on the participants' average responses. In particular, Greenhouse-Geisser (GG) corrected [45] one-way repeated measures ANOVA with the within-subject factor epoch revealed no significant effect of epoch [$F(3,45) = 1.38$, $p = .261$, $\eta_p^2 = .08$, $\epsilon = .62$]. In line with this, a paired t test comparing the results of the first and fourth epoch yielded no significant difference [$t(15) = 0.47$, $p = .646$, $d_z = .12$], indicating that participants answered similarly at the beginning and end of the experiment. The respective Bayes factor analysis for this pairwise comparison provided some evidence for the absence of an effect of epoch ($BF_{01} = 3.55$). Finally, a Levene's test comparing the between-subject variance of the results in the first and fourth epoch yielded no significant difference [$F(1,30) = 1.66$, $p = .208$], thus providing no statistical support for the observations that participants' answers might become more extreme towards the end of the experiment.

To quantify how consistent participants preferred one over the other rendering method across the experiment, we calculated Pearson correlations between all pairs of epochs. As shown in Table 1, these correlations were high and significant throughout, demonstrating that participants' preferences were highly consistent across epochs. By implication, the high correlations additionally show that at least those participants who strongly favored the near- or far-field condition were able to clearly discriminate the respective HRTFs.

In addition, we also examined participants' individual movement patterns across epochs (see Figs. S5–S8 in the Supplementary Material [42]). The plots show that most participants consistently performed similar movements and did not notably change their movement pattern during the experiment. These observations may indicate that, as we expected, the movement actually became automatic for participants after a short period of time (already during training or within the first few trials of the first epoch), allowing them to focus their cognitive resources on the

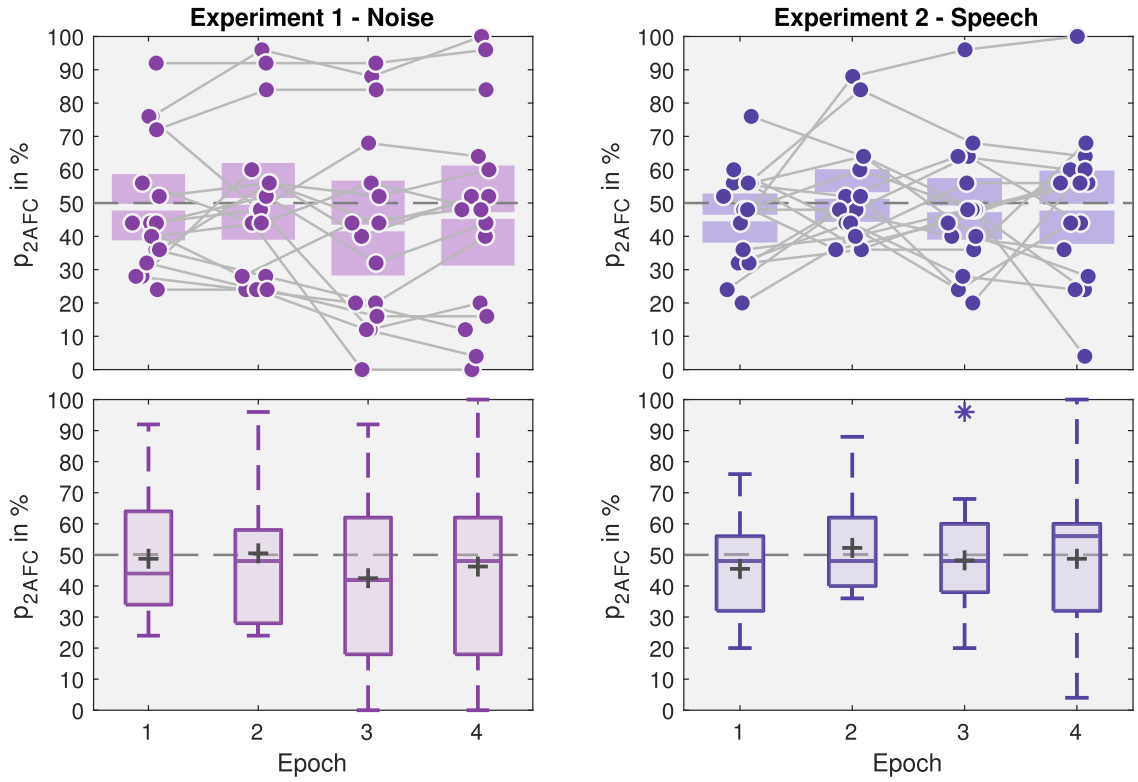

Figure 6. Results of Experiment 1 (left) and Experiment 2 (right) as a function of four epochs with 25 trials each. The top plots show the determined individual p_{2AFC} values in each epoch as points (horizontal offset for better readability). Individual data points are connected with gray lines. The boxes show the mean (box notch) and the 95% between-subject CI for each epoch. The gray dashed line denotes 50% chance level. The bottom plots show the interindividual variation in the determined p_{2AFC} values for each epoch in the form of a box plot with the median (box line), the mean (cross), and the (across participants) interquartile range (IQR); whiskers display 1.5 × IQR below the 25th or above the 75th percentile and outliers beyond that range are indicated by asterisks.

Table 1. Pearson correlation coefficients between all pairs of epoch for Experiment 1 (top) and Experiment 2 (bottom).

Epoch	1	2	3	4
	Experiment 1 – Noise			
1				
2	.78***			
3	.75***	.81***		
4	.72**	.85***	.96***	
	Experiment 2 – Speech			
1				
2	.34			
3	.12	.60*		
4	−.19	.44	.51*	

Note. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$, $N = 16$.

listening task rather than on moving the handheld loudspeaker (see, e.g., [46, 47]).

3 Experiment 2

The results of Experiment 1 provided no evidence that near-field cues enhance the plausibility of binaural rendering in a dynamic multimodal virtual acoustic scene as employed in this study. One possible explanation for these rather surprising results is that the pink noise test signal used in Experiment 1 is perceived as unnatural no matter the plausibility of the HRTFs, because pink noise rarely occurs in everyday situations. Thus, participants have no listening experience with such a stimulus and therefore might find it difficult to judge its plausibility based on their life experience and inner reference. For this reason and to provide a stimulus more commonly encountered in the near field, we used female speech as the test signal in Experiment 2, which was otherwise identical in design and procedure to Experiment 1. Furthermore, using a speech stimulus makes Experiment 2 more similar to applied scenarios, as near-field rendering of speech is important for various VR and AR applications.

3.1 Method

A new sample of 16 participants (ages 20–33 years, $M =$ 25.8 years, $Mdn = 28$ years, $SD = 3.9$) with self-reported normal hearing took part in the experiment for course credit. All participants were engineering students without experience in listening experiments and therefore classified as naive listeners. They were all naive as to the purpose of the study.

As outlined above, the only difference from the first experiment was that we used female speech as the test signal in this experiment. We chose the first, second, third, and sixth phonetically balanced sentences from the first list of Harvard sentences, spoken by a native female British English speaker [48]. The sentences were composed into a sequence of 10 s length (the same length as the noise burst sequence used in Experiment 1) with 62.5 ms silent pauses between the sentences. Similar to the first experiment, the speech test signal was filtered with the minimum phase FIR filter, combining the headphone compensation filter and the loudspeaker filter. To ensure similar presentation levels as in Experiment 1, the loudness of the speech test signal was adjusted to that of the noise test signal used in Experiment 1 according to the ITU-R BS.1770-4 recommendation [49]. The described processing can be reproduced by the Matlab script available in the Supplementary Materials [42]. In all other aspects, setup, materials, procedure, and analysis were identical to Experiment 1 (see Sect. 2).

3.2 Results and discussion

Participants in Experiment 2 also generally perceived the scene as plausible, as determined by informal post-experiment interviews. Figure 3 (right) shows the two-dimensional histogram of the relative tracking data of Experiment 2, pooled over all participants and trials. Again, it shows the square-like movement pattern reflecting the prescribed trajectory. Thus, participants in Experiment 2 also very frequently covered positions that yielded strong near-field cues in the near-field condition. Figure S4 in the Supplementary Material [42] provides individual-subject data.

Figure 4 also shows the results of Experiment 2. The majority of the participants performed near chance level (see left plot in Fig. 4), indicating that most participants could not decide which rendering method was more plausible or simply perceived both conditions as equally plausible. Only a single (outlying) participant (see right plot in Fig. 4) clearly perceived the near-field condition as more plausible than the far-field condition. Consequently, the box plot in Figure 4 (right) exhibits a rather small IQR with the mean and median slightly below chance level.

A Lilliefors test for normality showed no violations of normality ($p = .178$), so we performed a one-sample t test against chance level. In line with the plots, the test yielded no significant difference between the p_{2AFC} mean of 48.7% and chance level [$t(15) = 0.41$, $p = .684$, $d = .10$]. The respective Bayes factor analysis provided some evidence for the absence of the effect ($BF_{01} = 3.63$).

Participants' dwell time in the lateral region ($|\phi| > 60°$) did not significantly correlate with their performance [$r(14) = .42$, $p = .109$]. A close look at the corresponding scatter plot in Figure 5 (right) indicates that the (sizeable) correlation is mainly driven by the outlier, dropping to $r(13) = .06$, $p = .845$ with this outlier excluded (see results in purple in the right plot of Fig. 5). Thus, for the vast majority of participants, there is no evidence that they perceived the near-field condition as more plausible even when they frequently positioned the virtual sound source in lateral regions, producing strong binaural near-field cues and clear differences between near- and far-field conditions.

Analysis of plausibility ratings in the four epochs (each with 25 trials) showed that the majority of participants consistently performed close to chance throughout the experiment (see Fig. 6 (right)). Only one participant (the outlier) clearly tended increasingly towards the near-field condition over epochs. Thus, considering the entire data set, we did not detect a significant fatigue or learning effect. A GG-corrected one-way repeated measures ANOVA with the within-subject factor epoch showed no significant effect of epoch [$F(3,45) = 0.52$, $p = .668$, $\eta_p^2 = .03$, $\epsilon = .72$], and a paired t test comparing the results of the first and fourth epoch also showed no significant difference [$t(15) = 0.44$, $p = .664$, $d_z = .38$]. The Bayes factor analysis for the latter pairwise comparison yielded some evidence for the absence of an effect of epoch ($BF_{01} = 3.59$). Comparing the variances of the results in the first and fourth epoch with a Levene's test again yielded no significant difference [$F(1,30) = 1.48$, $p = .234$], thus providing no indication that answers would become more extreme across the course of the experiment.

For Experiment 2, we observed only few significant and relatively low correlations between plausibility ratings across epochs (see Tab. 1), indicating that participants by-and-large did not prefer one rendering method over the other with the speech stimulus used in Experiment 2. Thus, in contrast to Experiment 1, we cannot tell whether participants were even able to discriminate between the two rendering methods. Rather, it appears likely that most participants typically could not detect any clear differences between both rendering methods and, for that reason alone, could not reliably decide which rendering method was more plausible. The individual-subject movement data for each epoch indicate that participants' movements were consistent throughout the experiment (see Figs. S9–S12 in the Supplementary Material [42]), suggesting that the movement became automatic for participants already during training or within the first few trials of the experiment.

The plots in Figure 4 suggest that the results of Experiment 2 have a lower between-subject variance than those of Experiment 1. A Levene's test confirmed that the variances of the results are significantly different [$F(1,30) = 4.70$, $p = .038$]. We consider this and the absence of correlations across epochs discussed above as indication that the female speech test signal used in Experiment 2 elicited fewer perceptual differences between the near- and far-field HRTFs than the noise test signal used in Experiment 1.

4 General discussion

Previous research on near-field HRTFs and the perception of nearby sound sources mainly focused on distance estimation accuracy and the role of near-field cues (mainly the ILD cue) on distance judgments, leading to a variety of partly conflicting results on the contribution of near-field

cues to auditory distance perception (see, e.g., Arend et al. [11] for an overview). However, there is very little research investigating other perceptual aspects of near-field HRTFs, even though, especially with the emerging interest in binaural 6-DoF rendering for real-time VR and AR applications that often have limited resources, it is becoming increasingly important to determine whether simulating physically correct near-field cues is perceptually necessary. To address these questions, we conducted two 2AFC experiments in a 6-DoF multimodal AR experience investigating whether near-field HRTFs provide a more plausible binaural reproduction of nearby sound sources than intensity-scaled far-field HRTFs in dynamic multimodal virtual acoustic scenes.

The results of both experiments show no evidence that near-field cues enhance the plausibility of non-individual anechoic binaural rendering of nearby sound sources in the dynamic multimodal virtual acoustic scene designed for this study. Thus, even though in the present study the chance of perceiving a difference between the near- and far-field conditions was maximized because the multimodal AR experience provided proprioceptive and visual cues that could have conflicted with incorrect auditory cues, performance was on average (Experiment 1) or for almost each individual participant (Experiment 2) close to chance level, yielding average p_{2AFC} values slightly below but not significantly different from chance level. The equality in plausibility of the two compared HRTFs is rather surprising, given that near-field HRTFs lead to a physically more accurate representation of the nearby sound field than intensity-scaled far-field HRTFs and should therefore be perceived as more plausible on the (common) assumption that plausibility is governed by physical accuracy. Overall, the data from both experiments even show a (non-significant) trend toward p_{2AFC} values that are clearly below chance, which means numerous participants perceived the intensity-scaled far-field HRTFs as more plausible than the near-field HRTFs. On the other hand, there were participants in both experiments who favored the near-field condition. In Experiment 1, it was two expert listeners who tended toward the near-field HRTFs. However, two other expert listeners performed near or even below chance. The statistical outlier in Experiment 2, who tended to prefer near-field HRTFs, was not classified as an expert listener. Thus, there is no obvious relationship between listening experience and perceived plausibility of the near-field reproduction in the present study.

In both experiments, preference for near-field renderings did not correlate with the time participants placed the virtual sound source in the lateral region, where it would produce strong ILD cues. Moreover, both experiments did not show any learning or fatigue effects throughout the experiment, as revealed by comparing performance across four epochs.

The analysis in epochs also showed that preferences in terms of plausibility strongly correlated across epochs in Experiment 1, but lower and often non-significant correlations were observed in Experiment 2. Thus, some participants' answers were very consistent throughout the first experiment, i.e., they consistently perceived the near-field HRTFs as more plausible; others consistently preferred the far-field HRTFs. These consistent ratings also imply that these participants must have perceived differences between the two rendering methods. In Experiment 2, most participants performed near chance level in all four epochs, suggesting that they did either not have any preferences or did not even perceive any difference between the two rendering methods – even after extensive exposure to the speech stimulus and availability of clear spatial cues and rich multimodal information. This difference in consistency of ratings between experiments is also reflected in the between-subject variance: compared to Experiment 1, ratings in Experiment 2 exhibit a significantly lower between-subject variance, with individual p_{2AFC} values all closer to chance level.

One reason for this pattern of results could be that the speech signal used in Experiment 2 provided smaller perceptual differences than the noise signal used in Experiment 1, so that participants could not distinguish between the two renderings and therefore each individual participant answered more randomly in Experiment 2. As the low-frequency ILD cues are similarly excited by both test signals, we assume that the different results are because the spectral differences between the near- and far-field HRTFs (low-pass filtering character), which are strongest at higher frequencies, are much more audible for the broadband noise signal than for the speech signal, which has low energy above 8 kHz.

All these findings suggest – much to our surprise – that using near-field HRTFs or simply continuously adapting ILDs as a function of sound source distance, as done in various binaural renderers, does *not* lead (at least in dynamic multimodal environments) to a more plausible rendering of a virtual sound source in anechoic conditions for naive listeners than a simple rendering with intensity-scaled far-field HRTFs.

In a recent study conducted in a 6-DoF VR environment by Rummukainen et al. [14], listeners did not prefer measured multi-distance near-field HRTFs over intensity-scaled far-field HRTFs for non-individual anechoic dynamic binaural near-field rendering. The authors therefore concluded that including near-field HRTFs provides little benefit in a 6-DoF VR environment. However, the closest distance examined in their study was 0.50 m and the distance resolution was low, both because they used a near-field HRTF set measured with a Neumann KU100 at distances of 0.50, 0.75, 1.00, and 1.50 m [5]. As the strongest distance-dependent near-field effects occur below 0.50 m [3, 5], the authors mentioned that further studies with closer distances are necessary to be able to make a conclusion.

With the present study, we made another attempt to investigate whether near-field HRTFs provide an advantage for non-individual binaural reproduction in an anechoic 6-DoF VR or AR environment, but avoided above-mentioned drawbacks by using near-field HRTFs for very close distances down to 0.12 m at a much higher resolution of 1 cm in distance. In general, our results support the (partly inconclusive) findings of Rummukainen

et al. [14] that, from a perceptual point of view, near-field HRTFs provide little to no benefit for naive listeners in 6-DoF VR or AR multimodal applications employing binaural synthesis. In contrast, previous studies such as those by Brungart [6] or Kan et al. [12], claimed that near-field HRTFs are mandatory to generate binaural near-field rendering (based on the general assumption that a physically correct near-field representation is necessary). However, these conclusions are based on studies on distance estimation accuracy, which we did not investigate in our experiments. Thus, the importance of near-field HRTFs might differ depending on the task or application, i.e., if high-precision distance estimation accuracy in the near field is mandatory in an application, near-field HRTFs might provide advantages, whereas our results suggest that they are not necessary for an overall plausible representation of a dynamic spatial sound scene.

Our experiments, as well as the study by Rummukainen et al. [14], might indicate that correct reproduction of intensity as the primary and strongest distance cue is, in most cases, sufficient for a plausible representation of nearby sound sources in dynamic multimodal virtual environments. In line with this, experiments on distance perception revealed that, if available, the intensity cue dominates auditory distance estimation and masks the much more subtle near-field cues [11, 13]. Furthermore, a multimodal AR experience, as in the present study, provides proprioceptive and visual cues in addition to auditory cues, enhancing auditory localization and providing listeners with more information to judge the plausibility of a virtual sound source reliably. Conforming to this, previous studies on auditory space adaptation and multisensory learning effects have found evidence indicating that kinesthetic cues are additive to those evoked when the listener only pays attention to the sound source or can only see its position in space, suggesting that kinesthetic cues further support the spatial hearing updating process [35, 50, 51]. Valzolgher et al. [35], for example, considered that the sensory input achieved by multimodal stimulation, which is also supported by the human intention to act in space, could contribute to tuning the listener's sound-space correspondences. Moreover, similar to the intensity cue, these strong visual and proprioceptive cues might mask the more subtle near-field cues. To summarize, there are two possible effects of multimodal stimulation on plausibility assessment, which may even interact with each other. On the one hand, there is significant scientific evidence that multimodal stimulation combining real and simulated information improves plausibility judgments, as the different information streams can be evaluated concerning their congruency, and possible incoherences between the streams appear immediately as a break in plausibility. On the other hand, simultaneous streams containing real information congruent with the simulated auditory information might mask (in addition to the intensity cues) the less salient near-field cues, probably making the AR experience plausible even with simple distance-dependent intensity-scaling of far-field HRTFs.

The results are of particular relevance for real-time VR and AR applications with limited resources that use (mostly non-individual) binaural synthesis for 6-DoF rendering of virtual sound sources. Our results suggest that the additional (computational) effort of including near-field cues or near-field HRTF synthesis may not be necessary in terms of plausibility and reproduction quality for multimodal scenes. Furthermore, most applications reproduce reverberant environments, in which early reflections and reverberation would most probably further reduce perceptual differences between near- and far-field HRTFs. As our results suggest that even in anechoic environments using near-field HRTFs provides no perceptual benefit in terms of plausibility for naive listeners, we assume that all the more there is no benefit in using near-field HRTFs for reproducing reverberant environments.

Acknowledgments

We are grateful to the three anonymous reviewers for their constructive comments on a previous version of this manuscript. We also give special thanks to all the participants in the study. This work was supported by the German Federal Ministry of Education and Research (03FH014IX5-NarDasS and 13FH666IA6-VIWER-S).

Supplementary material

Supplementary material containing the Matlab script developed to generate the HRTFs and the filtered test signals, a video illustrating the experimental procedure, and additional results figures is available at https://doi.org/10.5281/zenodo.5656726.

Conflict of interest

Authors declared no conflict of interests.

References

1. P. Zahorik, D.S. Brungart, A.W. Bronkhorst: Auditory distance perception in humans: A summary of past and present research. Acta Acustica United with Acustica 91, 3 (2005) 409–420.
2. A.J. Kolarik, B.C.J. Moore, P. Zahorik, S. Cirstea, S. Pardhan: Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. Attention, Perception, & Psychophysics 78, 2 (2016) 373–395. https://doi.org/10.3758/s13414-015-1015-1.
3. D.S. Brungart, W.M. Rabinowitz: Auditory localization of nearby sources. Head-related transfer functions. The Journal of the Acoustical Society of America 106, 3 (1999) 1465–1479. https://doi.org/10.1121/1.427180.
4. D.S. Brungart, W.M. Rabinowitz: Auditory localization in the near-field, in Proc. of the 3rd International Conference on Auditory Display, Palo Alto, CA, USA. 1996, pp. 1–5.
5. J.M. Arend, A. Neidhardt, C. Pörschmann: Measurement and perceptual evaluation of a spherical near-field HRTF set, in Proc. of the 29th Tonmeistertagung – VDT International Convention, Cologne, Germany. 2016, pp. 356–363.

6. D.S. Brungart: Auditory localization of nearby sources. III. Stimulus effects. The Journal of the Acoustical Society of America 106, 6 (1999) 3589–3602. https://doi.org/10.1121/1.428212.
7. N. Kopčo, B.G. Shinn-Cunningham: Effect of stimulus spectrum on distance perception for nearby sources. The Journal of the Acoustical Society of America 130, 3 (2011) 1530–1541. https://doi.org/10.1121/1.3613705.
8. N. Kopčo, S. Huang, J.W. Belliveau, T. Raij, C. Tengshe, J. Ahveninen: Neuronal representations of distance in human auditory cortex. Proceedings of the National Academy of Sciences 109, 27 (2012) 11019–11024. https://doi.org/10.1073/pnas.1119496109.
9. N. Kopčo, K. Kumar Doreswamy, S. Huang, S. Rossi, J. Ahveninen: Cortical auditory distance representation based on direct-to-reverberant energy ratio. NeuroImage 208 (2020) 116436. https://doi.org/10.1016/j.neuroimage.2019.116436.
10. B.G. Shinn-Cunningham: Localizing sound in rooms, in Proc. of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments, Snowbird, Utah. 2001, pp. 17–22.
11. J.M. Arend, H.R. Liesefeld, C. Pörschmann: On the influence of non-individual binaural cues and the impact of level normalization on auditory distance estimation of nearby sound sources. Acta Acustica 5, 10 (2021) 1–21. https://doi.org/10.1051/aacus/2021001.
12. A. Kan, C. Jin, A. van Schaik: A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function. The Journal of the Acoustical Society of America 125, 4 (2009) 2233–2242. https://doi.org/10.1121/1.3081395.
13. S. Spagnol, E. Tavazzi, F. Avanzini: Distance rendering and perception of nearby virtual sound sources with a near-field filter model. Applied Acoustics 115 (2017) 61–73. https://doi.org/10.1016/j.apacoust.2016.08.015.
14. O.S. Rummukainen, S.J. Schlecht, T. Robotham, A. Plinge, E.A.P. Habets: Perceptual study of near-field binaural audio rendering in six-degrees-of-freedom virtual reality, in Proc. of IEEE VR, Osaka, Japan. 2019, pp. 1–7. https://doi.org/10.1109/VR.2019.8798177.
15. A. Lindau, S. Weinzierl: Assessing the plausibility of virtual acoustic environments. Acta Acustica United with Acustica 98, 5 (2012) 804–810. https://doi.org/10.3813/AAA.918562.
16. M. Slater: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B 364 (2009) 3549–3557. https://doi.org/10.1098/rstb.2009.0138.
17. M. Hofer, T. Hartmann, A. Eden, R. Ratan, L. Hahn: The role of plausibility in the experience of spatial presence in virtual environments. Frontiers in Virtual Reality 10, April (2020) 1–9. https://doi.org/10.3389/frvir.2020.00002.
18. U. Reiter: Perceived quality in game audio, in Grimshaw M (Ed.), Game Sound Technology and Player Interaction: Concepts and Developments, Chapter 8, IGI Global, Hershey, PA, USA. 2011, pp. 153–174. https://doi.org/10.4018/978-1-61692-828-5.ch008.
19. D. Ackermann, F. Fiedler, F. Brinkmann, M. Schneider, S. Weinzierl: On the acoustic qualities of dynamic pseudobinaural recordings. The Journal of the Audio Engineering Society 68, 6 (2020) 418–427. https://doi.org/10.17743/jaes.2020.0036.
20. J.M. Arend, S.V. Amengual Garí, C. Schissler, F. Klein, P.W. Robinson: Six-degrees-of-freedom parametric spatial audio based on one monaural room impulse response. The Journal of the Audio Engineering Society 69, 7/8 (2021) 557–575. https://doi.org/10.17743/jaes.2021.0009.
21. F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, S. Weinzierl: A round robin on room acoustical simulation and auralization. The Journal of the Acoustical Society of America 145, 4 (2019) 2746–2760. https://doi.org/10.1121/1.5096178.
22. A. Neidhardt, N. Knoop: Binaural walk-through scenarios with actual self-walking using an HTC Vive, in Proc. of the 43rd DAGA, Kiel, Germany. 2017, pp. 283–286.
23. A. Neidhardt, A.I. Tommy, A.D. Pereppadan: Plausibility of an interactive approaching motion towards a virtual sound source based on simplified BRIR sets, in Proc. of the 144th AES Convention, Milan, Italy. 2018, pp. 1–11.
24. S.V. Amengual Garí, J.M. Arend, P. Calamia, P.W. Robinson: Optimizations of the spatial decomposition method for binaural reproduction. The Journal of the Audio Engineering Society 68, 12 (2020) 959–976. https://doi.org/10.17743/jaes.2020.0063.
25. A. Neidhardt, A.M. Zerlik: The availability of a hidden real reference affects the plausibility of position-dynamic auditory AR. Frontiers in Virtual Reality 2, 678875 (2021) 1–17. https://doi.org/10.3389/frvir.2021.678875.
26. VRACE: VRACE Research Team. https://vrace-etn.eu/research-team/. Accessed: 2021-11-09.
27. Oculus: Oculus Developer. https://developer.oculus.com/blog/near-field-3d-audio-explained. Accessed: 2021-11-09.
28. Magic Leap: Magic Leap Developer. https://developer.magicleap.com/en-us/learn/guides/lumin-sdk-soundfield-audio. Accessed: 2021-11-09.
29. Resonance Audio: Resonance Audio Developer. https://resonance-audio.github.io/resonance-audio/develop/overview.html. Accessed: 2021-11-09.
30. T. Carpentier, M. Noisternig, O. Warusfel: Twenty years of Ircam Spat: Looking back, looking forward, in Proc. of 41st International Computer Music Conference (ICMC), Denton, TX, USA. 2015, pp. 270–277.
31. D. Poirier-Quinot, B.F.G. Katz: The Anaglyph binaural audio engine, in Proc. of the 144th AES Convention, Milan, Italy. 2018, pp. 1–4.
32. M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, A. Reyes-Lecuona: 3D tune-in toolkit: An open-source library for real-time binaural spatialisation. PLoS One 14, 3 (2019) 1–37. https://doi.org/10.1371/journal.pone.0211899.
33. K. Strelnikov, M. Rosito, P. Barone: Effect of audiovisual training on monaural spatial hearing in horizontal plane. PLoS One 6, 3 (2011) 1–9. https://doi.org/10.1371/journal.pone.0018344.
34. A. Isaiah, T. Vongpaisal, A.J. King, D.E.H. Hartley: Multisensory training improves auditory spatial processing following bilateral cochlear implantation. The Journal of Neuroscience 34, 33 (2014) 11119–11130. https://doi.org/10.1523/JNEUROSCI.4767-13.2014.
35. C. Valzolgher, C. Campus, G. Rabini, M. Gori, F. Pavani: Updating spatial hearing abilities through multisensory and motor cues. Cognition 204 (2020) 104409. https://doi.org/10.1016/j.cognition.2020.104409.
36. A. Neidhardt, F. Klein, N. Knoop, T. Köllmer: Flexible Python tool for dynamic binaural synthesis applications, in Proc. of the 142nd AES Convention, Berlin, Germany. 2017, pp. 1–5.
37. B. Bernschütz: A spherical far field HRIR/HRTF compilation of the Neumann KU 100, in Proc. of the 39th DAGA, Merano, Italy. 2013, pp. 592–595.
38. R.O. Duda, W.L. Martens: Range dependence of the response of a spherical head model. The Journal of the Acoustical Society of America 104, 5 (1998) 3048–3058. https://doi.org/10.1121/1.423886.
39. V. Ralph Algazi, C. Avendano, R.O. Duda: Estimation of a spherical-head model from anthropometry. The Journal of the Audio Engineering Society 49, 6 (2001) 472–479.

40. D. Romblom, B. Cook: Near-Field Compensation for HRTF Processing, in Proc. of the 125th AES Convention, San Francisco, USA. 2008, pp. 1–6.
41. J.M. Arend, C. Pörschmann: Synthesis of near-field HRTFs by directional equalization of far-field datasets, in Proc. of the 45th DAGA, Rostock, Germany. 2019, pp. 1454–1457.
42. J.M. Arend, M. Ramírez, H.R. Liesefeld, C. Pörschmann: Supplementary material for "Do near-field cues enhance the plausibility of non-individual binaural rendering in a dynamic multimodal virtual acoustic scene?". Nov. 2021. https://doi.org/10.5281/zenodo.5656726.
43. A. Lindau, F. Brinkmann: Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings. The Journal of the Audio Engineering Society 60, 1/2 (2012) 54–62.
44. V. Erbes, M. Geier, H. Wierstorf, S. Spors: Free database of low-frequency corrected head-related transfer functions and headphone compensation filters, in Proc. of the 127th AES Convention, New York, NY, USA. 2017, pp. 1–5.
45. S.W. Greenhouse, S. Geisser: On methods in the analysis of profile data. Psychometrika 24, 2 (1959) 95–112. https://doi.org/10.1007/BF02289823.
46. B. Bruya: Effortless attention: A new perspective in the cognitive science of attention and action. MIT Press, Cambridge, MA, 2010. https://doi.org/10.7551/mitpress/9780262013840.001.0001.
47. W. Schneider, R.M. Shiffrin: Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review 84, 1 (1977) 1–66. https://doi.org/10.1037/0033-295X.84.1.1.
48. P. Demonte: HARVARD speech corpus – audio recording 2019. University of Salford. Collection, 2019. URL https://doi.org/10.17866/rd.salford.c.4437578.v1.
49. ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level. International Telecommunications Union, Geneva, 2015.
50. A. Maravita, C. Spence, J. Driver: Multisensory integration and the body schema: Close to hand and within reach. Current Biology 13, 13 (2003) 531–539. https://doi.org/10.1016/S0960-9822(03)00449-4.
51. M. Gori, T. Vercillo, G. Sandini, D. Burr: Tactile feedback improves auditory spatial localization. Frontiers in Psychology 5 (2014) 1–7. https://doi.org/10.3389/fpsyg.2014.01121.

Cite this article as: Arend J. Ramírez M. Liesefeld HR. & Pörschmann C. 2021. Do near-field cues enhance the plausibility of non-individual binaural rendering in a dynamic multimodal virtual acoustic scene?. Acta Acustica, **5**, 55.