Acta Acustica 2021, **5**, 10

https://doi.org/10.1051/aacus/2021001

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

SCIENTIFIC ARTICLE OPEN ACCESS

On the influence of non-individual binaural cues and the impact of level normalization on auditory distance estimation of nearby sound sources

Johannes M. Arend[1,2,*], Heinrich R. Liesefeld[3,4], 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
[4] Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 Munich, Germany

Received 18 June 2020, Accepted 14 January 2021

Abstract – Nearby sound sources provide distinct binaural cues, mainly in the form of interaural level differences, which vary with respect to distance and azimuth. However, there is a long-standing controversy regarding whether humans can actually utilize binaural cues for distance estimation of nearby sources. Therefore, we conducted three experiments using non-individual binaural synthesis. In Experiment 1, subjects had to estimate the relative distance of loudness-normalized and non-normalized nearby sources in static and dynamic binaural rendering in a multi-stimulus comparison task under anechoic conditions. Loudness normalization was used as a plausible method to compensate for noticeable intensity differences between stimuli. With the employed loudness normalization, nominal distance did not significantly affect distance ratings for most conditions despite the presence of non-individual binaural distance cues. In Experiment 2, subjects had to judge the relative distance between loudness-normalized sources in dynamic binaural rendering in a forced-choice task. Below chance performance in this more sensitive task revealed that the employed loudness normalization strongly affected distance estimation. As this finding indicated a general issue with loudness normalization for studies on relative distance estimation, Experiment 3 directly tested the validity of loudness normalization and a frequently used amplitude normalization. Results showed that both normalization methods lead to remaining (incorrect) intensity cues, which subjects most likely used for relative distance estimation. The experiments revealed that both examined normalization methods have consequential drawbacks. These drawbacks might in parts explain conflicting findings regarding the effectiveness of binaural cues for relative distance estimation in the literature.

1 Introduction

The primary acoustic cues for distance perception in the far field are sound intensity, direct-to-reverberant energy ratio (DRR), and spectral cues [1, 2]. Whereas the DRR cue provides absolute distance information, intensity and spectrum are relative distance cues, which means that different sounds have to be compared in order to judge distance. In reverberant environments, humans can use a combination of these cues for distance estimation, albeit spectral distance cues caused by high frequency attenuation are only available for sound sources with a distance of more than 15 m ([3], Chapter 2.3.2). In anechoic conditions, distance estimation of far field sources (below 15 m) relies mainly on intensity cues.

In the near field (sound source distance less than 1 m), interaural time differences (ITDs), interaural level differences (ILDs), and characteristic spectral cues occur in addition [2, 4]. The spectral properties of nearby sound sources change across distance. Diffraction and head-shadowing effects lead to a low-pass filtering character of nearby sound sources which might be a spectral cue for distance estimation in the near field [1, 2, 4]. While the influence of distance on the ITDs is relatively low [4, 5], ILDs change substantially across distance for nearby sound sources [4, 5].

The increase in ILDs as a sound source approaches the head is mainly caused by frequency-dependent head-shadowing effects and might be the most prominent feature of nearby sound sources. The strongest increase in ILDs can be observed for lateral sound sources at distances below 0.50 m [4]. For example, broadband ILDs obtained from HRTFs (Head-Related Transfer Functions) measured with a dummy head can increase about 10 dB to an order of 20 dB [4] or 23 dB [6] for a lateral sound source at a distance of 0.12 m or 0.25 m respectively. Because of these drastic changes in ILDs across the whole spectrum, it is assumed that ILDs play an important role for distance estimation in the near field [2].

*Corresponding author: Johannes.Arend@th-koeln.de

However, intensity has been shown to be the highest weighted distance cue [7, 8]. To evaluate the contribution of other distance cues, previous research has eliminated intensity differences between stimuli by various kinds of level equalization (see Tab. A.1 in Appendix for an overview). The *normalization method* to equalize levels has a distinct impact on the stimuli. This issue was not considered in detail in previous studies and may explain some conflicting results in the literature as summarized in the following.

Holt and Thurlow [9] conducted an experiment in anechoic conditions with level-equalized sources at distances between 1.80 m and 19 m. Binaural cues were isolated by eliminating DRR (anechoic) and intensity cues (level-equalized). The authors reported that subjects were not able to judge the distance of a frontally oriented sound source presenting a broadband noise stimulus. However, performance improved when the source was positioned laterally. In a similar experiment with level-equalized speech sources arranged in the front at distances between 0.90 m and 9.00 m, Gardner [10] observed that small head movements showed slight benefits for distance estimation. Although both studies addressed far-field sources only, the results indicate that binaural cues might provide additional information for distance estimation. Brungart et al. examined nearby sound sources in two consecutive listening experiments in anechoic conditions [7, 11]. Subjects had to judge the position of a specific sound source (approximating an acoustic point source) randomly located in their right hemifield at distances between 0.15 m and 1.00 m. To remove intensity-based cues and thus to isolate binaural cues, the amplitude of the noise stimulus was normalized dependent on distance. Additionally, to further reduce the reliability of potentially remaining intensity cues, the amplitude of the normalized stimuli was roved randomly over a 15 dB range. The results showed that distance estimation was most accurate for lateral sources and least accurate near the median plane. Moreover, accuracy of distance estimation degraded when frequency components below 3 kHz were absent. Based on these results, the authors concluded that low-frequency ILDs (below 3 kHz) are the primary and most salient cue for distance estimation of nearby lateral sound sources when no intensity cues are available, whereas listeners rely primarily on intensity cues for nearby medial sources where binaural cues are weak. In a follow-up study, Brungart and Simpson [12] conducted a similar experiment by means of a virtual auditory display based on near-field HRTFs of a KEMAR dummy head [4]. Again, subjects had to judge distance to level-equalized and level-roved (10 dB range) virtual noise sources located in their right hemifield at distances between 0.12 m and 1.00 m. In this study, however, a different type of level normalization was applied. Subjects performed worse than in the experiment with a real sound source [7, 11], which according to the authors was most likely due to the use of non-individual HRTFs. However, the authors stated that especially for lateral sound sources, subjects were still able to extract a substantial amount of distance information from the normalized nearby virtual sound sources. Kan et al. [13] also conducted an experiment in virtual acoustics applying near-field HRTFs synthesized from individual far-field HRTFs. Subjects had to judge the distance of virtual noise sources at distances between 0.10 m and 1.00 m. To remove intensity cues, the authors applied the same normalization method as Brungart et al. [7], but without level-roving. The results showed a distance discrimination for lateral sound sources within a range of 0.20 m, but the overall distance judgment performance was poor and the authors concluded that ILDs are no powerful cues. Spagnol et al. [8] conducted a similar study using synthesized and measured KEMAR near-field HRTFs. Subjects were asked to discriminate distance of virtual lateral and medial noise sources at distances between 0.20 m and 1.00 m. The researchers also applied the amplitude normalization proposed by Brungart et al. [7] without level-roving. The experiment resulted in average error rates very close to the 50% chance level, indicating that the overall performance was poor. However, similar to Kan et al. [13], performance slightly improved for lateral sources at distances below 0.20 m.

In contrast to these findings supporting the effectiveness of binaural cues, several other studies cast doubt whether binaural cues contribute to distance estimation. Simpson and Stanton [14] conducted experiments in quasi-anechoic conditions with a pulse-train source located in the front at distances between 0.30 m and 2.66 m and found that head movements had no influence on distance estimation and concluded that binaural cues are unimportant for distance perception. Rosenblum et al. [15] also noted that head movements had no influence on distance judgment accuracy. In their experiments, subjects had to judge the distance to a percussion shaker, located laterally at distances between 0.38 m and 1.10 m. However, since the sound sources were not level-equalized in both of these studies, intensity cues might have masked binaural cues so that no influence of head movements was found. Shinn-Cunningham et al. [16] addressed the question of whether binaural cues contribute to distance estimation in two experiments based on binaural synthesis using individualftable BRIRs (Binaural Room Impulse Responses) and individual HRTFs for reverberant and anechoic conditions respectively. Subjects had to judge distance of medial and lateral virtual pink noise sources at distances between 0.15 m and 1.00 m. Because the performance of untrained listeners was poor both for lateral and medial sound sources in anechoic conditions (distance perception was generally below chance), the authors concluded that binaural cues are weak or even irrelevant. Moreover, they observed that performance in anechoic conditions can improve with training, indicating that listeners can learn to use ILD cues for distance estimation of nearby sound sources. Finally, the authors stated that in reverberant conditions, DRR provides a robust distance cue in the near field, even for medial sound sources and untrained listeners. Based on these experiments, Shinn-Cunningham et al. concluded that ILDs do not contribute to distance estimation when reverberation is present, and that even in anechoic conditions, ILD cues do not lead to robust distance percepts [17, 18]. In a further study, Kopčo and Shinn-Cunningham [19] examined how changes in DRR and ILD affect distance

judgments. Again, the researchers used individual BRIRs to synthesize virtual lateral and medial sound sources at distances between 0.15 m and 1.70 m. To eliminate intensity cues, the noise stimuli were normalized in level, and to further diminish potentially remaining intensity cues, the normalized stimuli were level-roved over a 10 dB range. Similar to previous experiments, the results showed that performance was best for lateral sound sources and worse without low-frequency energy, but the authors concluded that listeners only use DRR cues to judge distance of nearby sound sources in reverberant condition. However, the authors further outlined that listeners might focus on different strategies to judge distance, depending on the listening conditions. In a later study using non-individual BRIRs, Kopčo et al. [20] qualified their statement by saying that listeners might combine the DRR and ILD cue for distance estimation, even though the DRR cue seems to be more robust and reliable than the ILD cue.

Given the conflicting results and despite the drastic variations in ILD induced by distance changes, the contribution of binaural cues to distance estimation in the near field remains an open issue. Even very similar studies strictly focusing on distance perception of nearby sound sources in anechoic conditions, like Brungart and Simpson [12] and Shinn-Cunningham et al. [16] for example, led to opposing conclusions regarding the influence of binaural cues. When comparing all studies, three factors stand out which could have had a significant influence on the respective results: the specific normalization method to eliminate intensity cues, the way head movements were considered, and the use of individual or non-individual HRTFs if binaural synthesis was applied. The following paragraphs shortly discuss those three aspects and their potential influence. Additionally, Table A.1 in Appendix gives an overview of mentioned studies including their method and the major findings concerning the contribution of binaural cues.

1.1 Normalization

One main reason for the differences between the results might be the normalization method. According to the pressure-discrimination hypothesis, just-noticeable differences (JNDs) in source distance are determined by the ability of discriminating changes in source intensity [2, 21]. Considering that the smallest detectable change in sound pressure level is about 0.4 dB for broadband noise [22], it is apparent that specific care has to be taken in the normalization to completely remove intensity cues so that intensity cannot be used instead of binaural cues. Brungart et al. [7], Kan et al. [13], and Spagnol et al. [8] applied the same distance-based normalization (see Sect. 4.1.2 for more details), which only approximately normalizes the amplitude of the stimuli (as acknowledged by Brungart et al. [7]). These three studies have in common that distance discrimination between normalized stimuli was most accurate for lateral sources really close to the head at distances below 0.20 m. In contrast, Kopčo and Shinn-Cunningham [19] normalized the stimuli so that the overall sound pressure level at the nearer ear was constant. The authors could not find any evidence that binaural cues were used for distance estimation in reverberant conditions. In previous studies, Shinn-Cunningham and Kopčo found similar results also for anechoic conditions, but the authors did not provide detailed information on their normalization method [16–18]. The comparison shows that the normalization method may significantly affect the results, for example whether intensity cues remain even after normalization. Additionally roving the stimuli in level after normalization, as for example done by Brungart et al. [7, 11] and Kopčo and Shinn-Cunningham [19], further diminishes potentially remaining intensity cues. However, roving seems not expedient for experiments on relative distance estimation, that is, experiments where at least two sound sources are presented concurrently or in quick succession, and relative distance differences have to be rated. In this case, roving would negate the normalization and introduce intensity cues that most certainly would dominate relative distance estimation (i.e., responses would likely almost exclusively be based on the level differences (re-)introduced by the roving procedure). This problem of level roving was demonstrated in a recent study from Prud'homme and Lavandier [23], where naive listeners judged distance primarily based on the roving-induced level variations, even though they were instructed to discard them. It is therefore necessary to find a reliable normalization method for studies on relative distance estimation. The performance of the various normalization methods or their possible impact on the results in studies of the type discussed above has not been systematically examined so far.

Instead of simply matching the levels of the stimuli in some way, normalizing the stimuli in loudness according to ITU-RBS.1770-4 [24] might be a better approach to remove intensity cues. In comparison to a strictly level-based analysis, loudness is a psychoacoustic measure that takes into account the frequency-dependent sensitivity of the human ear as well as the acoustic effects of the human head. In the course of development, the ITU evaluated the performance of the loudness algorithm in several listening experiments. These experiments yielded correlation coefficients of about $r = 0.98$ between perceived (subjective) loudness measurements and (objective) predicted loudness for a broad range of signals. Thus, the loudness model performs well and normalizing the stimuli in loudness should, arguably, work considerably better than previously introduced purely technical normalization methods, which do not consider the effects of human perception in the same way.

1.2 Head movements

Head movements have been shown to be another important factor influencing distance perception of nearby sound sources. Gardner [10] for example found that the small changes in binaural cues caused by head movements slightly improved distance estimation. In contrast, Simpson and Stanton [14] and Rosenblum et al. [15] reported no influence of head movements. Nevertheless, all studies outlined in this article neither captured head movements for further post-hoc analysis nor considered head movements

if binaural synthesis was applied. However, especially for nearby sound sources, head movements in the horizontal plane lead to distinct variations in ILDs. These variations might provide additional information for distance estimation in the near field and thus probably enhance human perception of auditory space, quite similar to the observation that localization performance in the horizontal and median plane improves if head movements are involved [25]. Therefore, it seems important to analyze the extent of head movements and how they influence distance estimation.

1.3 HRTFs

In comparison to non-individual HRTFs, individual HRTFs improve localization in the median plane and lead to reduced front-back confusion in static binaural synthesis because of more accurate monaural spectral cues [26–28]. In contrast, non-individual HRTFs still provide robust binaural cues for localization in the horizontal plane [26], often without relevant increase in localization error when compared to individual HRTFs [28]. However, it is not clear whether individual HRTFs lead to more accurate distance perception than non-individual HRTFs. Zahorik showed that in reverberant conditions, the performance of distance estimation of virtual sound sources is unaffected by the use of non-individualized HRTFs compared to the use of individualized HRTFs, since the strong intensity and DRR cues mask spectral deviations [1, 29, 30]. In line, Begault et al. [28] did not find any effects of individual HRTFs in anechoic and reverberant conditions on externalization, which is a perceptual attribute associated with distance perception [31]. Likewise, Yu et al. found no evidence for an influence of individual HRTFs on distance perception of nearby sound sources [32, 33]. Most recently, Prud'homme and Lavandier [23] showed that the use of non-individual BRIRs instead of individual BRIRs did not significantly affect absolute distance estimation in reverberant conditions, which confirmed Zahorik's findings. However, Hartmann and Wittenberg [34], Brimijoin et al. [35], or Baumgartner et al. [36] showed that spectral cues affect externalization, and that distorted spectral cues, such as those from non-individual HRTFs, reduce externalization and therefore perceived distance.

Thus, whereas individual HRTFs improve performance especially in median plane localization, their role in distance estimation is not entirely clear [31]. This is also evident when looking at previous studies that provide contradictory results that cannot be directly attributed to the type of HRTFs used. For example, Shinn-Cunningham et al. [16], Shinn-Cunningham [18], and Kopčo and Shinn-Cunningham [19] used individual HRTFs and could not find any evidence that binaural cues contribute to auditory distance estimation of nearby sound sources. In contrast, Brungart and Simpson [12] used generic KEMAR HRTFs and confirmed the findings from their earlier loudspeaker-based study. Brungart and Simpson [12] discussed the findings from Shinn-Cunningham et al. [16] in their article, but could not find a conclusive explanation for the conflicting results. Kopčo et al. [20] used non-individual HRTFs and found that the DRR cue masks potential ILD cues. Again in contrast, Kan et al. [13] and Spagnol et al. [8] both found a slight improvement in distance estimation performance for lateral close sources using individual or non-individual HRTFs respectively. Taken together, regardless of whether individual or non-individual HRTFs were used, the various studies led to contrary results and no direct correlation between the type of HRTFs and the respective findings can be found (see Tab. A.1 in Appendix). Rather, it seems that other factors, such as the test paradigm or the normalization method, have a greater influence than the HRTFs.

1.4 The current study

The detailed review of the studies in this field reveals a long-standing controversy and shows that the question of whether binaural cues contribute to distance perception is still an open issue. To address this, we conducted three listening experiments investigating distance perception of nearby virtual sound sources in anechoic conditions using non-individual binaural synthesis. Experiment 1 (Sect. 2) based on a multi-stimulus comparison method where subjects had to rate perceived distance to loudness-normalized (according to ITU-RBS.1770-4 [24]) and non-normalized stimuli in static or dynamic binaural rendering. In Experiment 2 (Sect. 3), we conducted a relative perceptual distance experiment between loudness-normalized nearby virtual sound sources. As Experiment 2 provided some ambiguous results, we conducted Experiment 3 (Sect. 4) as a follow-up to examine the performance of the loudness normalization and of the amplitude normalization proposed by Brungart et al. [7]. In particular, subjects of Experiment 3 rated the relative perceived loudness-difference between normalized nearby virtual sound sources.

2 Experiment 1

By asking listeners to estimate auditory distance to normalized nearby virtual sound sources in static or dynamic non-individual binaural synthesis, we tested whether distance estimation of nearby sound sources is possible without intensity cues. The particular goals were to test whether distance-related changes in binaural cues may be utilized to distinguish distance and whether head movements and the resulting variations in ILD provide additional information improving distance estimation. In another experimental condition, we maintained the distance-related level differences. The aim of this experimental condition was to examine whether head movements improve distance estimation even when intensity cues are provided, or whether intensity simply masks any additional binaural information.

2.1 Method

2.1.1 Participants

In total, 50 adults took part in the experiment for monetary remuneration (10 Euro per hour). Most of them were students in media technology or electrical engineering.

The participants were divided into two equal groups, with one group performing with head tracking (hereafter abbreviated as group *head tracking* – HT) and the other group performing without head tracking, i.e., using static binaural synthesis (hereafter abbreviated as group *static* – ST). Group HT was composed of 20 males and 5 females aged between 18 and 30 years ($M = 24.44$ years, $Mdn = 25$ years, $SD = 3.06$). Twelve participants of this group (48%) had already taken part in previous listening experiments and thus were familiar with the binaural reproduction system. Group ST was composed of 18 males and 7 females with an age between 19 and 31 years ($M = 23.76$ years, $Mdn = 22$ years, $SD = 3.28$). Here, 7 participants (28%) already had gained experience in former listening tests. However, all participants were naive as to the purpose of this experiment. Moreover, there was no previous training in distance estimation of nearby sound sources, which means that the participants had to rely on their life experience in perceiving nearby sound sources. All participants reported normal hearing.

2.1.2 Setup and stimuli

2.1.2.1 Setup

The experiment took place in the anechoic chamber of TH Köln, which has a low background noise of about 20 dB (A). The participants sat on an office swivel chair so that they could turn easily. The entire experiment was implemented, controlled, and executed with the MATLAB based software Scale [37], running on an Apple iMac. Scale handled the playback of the anechoic audio test signals as well as the internal audio routing in combination with the JACK Audio Connection Kit. For (dynamic) binaural rendering, the SoundScape Renderer [38] paired with a Fastrak head tracking system at a 120 Hz sampling rate was used. Only rotational head movements in the horizontal plane were considered, whereas vertical or translational head movements were disregarded. Via internal TCP/IP sockets, Scale controlled the renderer to switch between datasets or to change the settings according to the respective test condition. Once a second, Scale saved the head tracking data for further analysis of the head movements. The participants gave their response using an Apple iPad 2 tablet, which mirrored the graphical user interface (GUI) of Scale. The binaural audio signal was converted and amplified with an Fireface UFX audio interface and presented over AKG K601 headphones. The interface was set to a buffer size of 512 samples and a sampling rate of 48 kHz.

2.1.2.2 Test signal

As anechoic test signal, we used a pink noise burst sequence with a burst length of 1.50 s (including 10 ms cosine-squared onset/offset ramps) and an interstimulus interval of 0.50 s. Generally, a broadband signal ensures best possible localization performance ([3], Chapters 2.1 and 2.3). Concerning the special case of nearby sound sources, accurate distance judgment requires low frequency components below 3 kHz [11] or at least at around 300 Hz [19]. Hence, a pink noise signal is a good choice for this

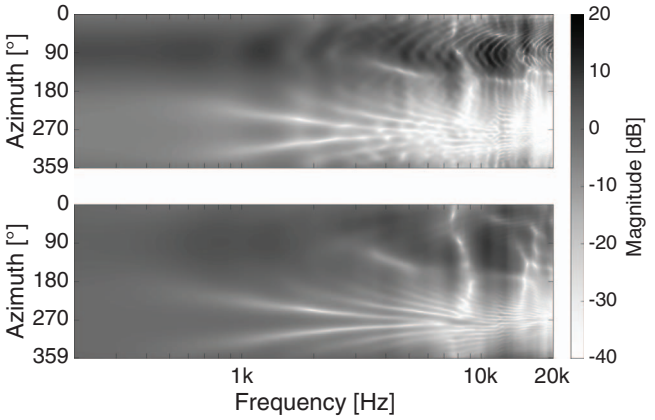

Figure 1. Left-ear magnitude spectrum of two circular grid HRTF datasets (top: $d = 0.25$ m, bottom: $d = 1.50$ m) as a function of frequency (abscissa) and source azimuth φ (ordinate). Comparing the spectrum of near- and far-field HRTFs reveals how a nearby source (top) leads to stronger high-frequency damping around the contralateral ear ($\varphi = 270°$) and increased magnitude around the ipsilateral ear ($\varphi = 90°$), also at frequencies below 3 kHz. This results in significantly higher ILDs (see Fig. 1, left).

experiment. We chose a rather long test signal to provide sufficient time for turning the head during stimulus presentation (as determined during pilot tests) to allow for head-turning related variation in binaural cues. The number of bursts played varied across experiments and is therefore described in the respective procedure paragraph.

2.1.2.3 Near-field HRTFs

To synthesize the nearby virtual sound sources, we used near-field HRTFs from a Neumann KU100 dummy head, measured at five sound source distances ($d = 0.25$ m, 0.50 m, 0.75 m, 1.00 m, 1.50 m) on a circular grid with a resolution of 1° in the horizontal plane [6, 39]. Figure 1 exemplarily shows the left-ear magnitude spectrum of two HRTF datasets (top: $d = 0.25$ m, bottom: $d = 1.50$ m) between 200 Hz and 20 kHz as a function of source azimuth (hereinafter termed direction, or simply φ). Comparing these two extremes (near field vs. far field) clearly reveals how a nearby sound source leads to stronger damping around the contralateral ear (with respect to the sound source, $\varphi = 270°$) and increased magnitude around the ipsilateral ear ($\varphi = 90°$). Moreover, decreased or increased magnitude towards the contralateral and ipsilateral ear respectively can be observed also at frequencies below 3 kHz. Consequently, the ILDs for nearby sources are distinctly higher compared to sources in the far field, as can be seen in Figure 2 (left), which shows the low-frequency ILDs ($f \leq 3$ kHz) of the HRTFs for the five distances. The polar plot shows that low-frequency ILDs, which are considered the primary cue for distance estimation of nearby lateral sources (see Sect. 1), are similar at the distances 1.50 m, 1.00 m, and 0.75 m, start to increase at a distance of 0.50 m, and rise strongly at the closest distance of 0.25 m. Same as for the ILDs, the low-frequency ITDs

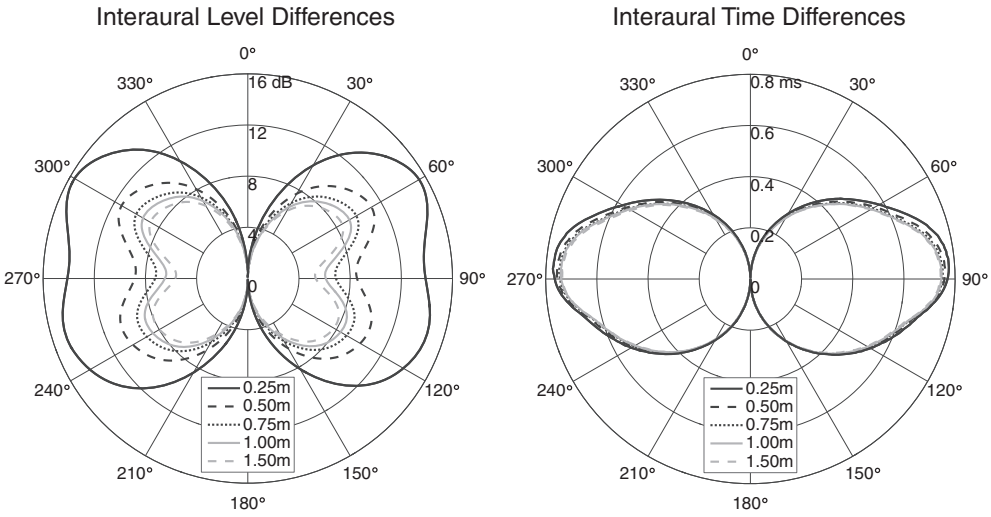

Figure 2. Low-frequency ($f \leq 3$ kHz) interaural level differences (left) and interaural time differences (right) of the used HRTF dataset. The angle represents the azimuth of the sound source (φ). The radius describes the magnitude of the level differences (in dB) or time differences (in ms). Both plots show the usual direction-dependent influence of the pinna and the head. However, typical for near-field HRTFs, the ILDs (left) increase significantly as a function of distance, whereas the ITDs (right) remain nearly constant.

presented in Figure 2 (right) show the usual direction-dependent influence of the pinna and the head. However, with only a slight increase as the source approaches the head, the ITDs are barely influenced by sound source distance. Overall, the analysis presented in this paper as well as the more detailed technical evaluation in Arend et al. [6] and Pörschmann et al. [39] confirm that the HRTFs have the intended near-field characteristics (see, e.g., [2]).

2.1.2.4 Stimuli

In the experiment, virtual noise sources in the horizontal plane (elevation $\vartheta = 0°$) at five different distances (0.25 m, 0.50 m, 0.75 m, 1.00 m, 1.50 m) and three different azimuthal positions (30°, 150°, 270°) per distance were presented, resulting in 15 nominal sound source positions. The positions were chosen to use stimuli covering the front and back hemisphere and showing rather low as well as very distinct binaural cues.

For the non-normalized stimuli, strictly distance-dependent gain values (i.e., independent of the azimuthal position) were calculated in order to maintain natural distance-related level differences. For the normalized stimuli, the loudness of each stimulus was determined according to ITU-RBS.1770-4 for frontal head orientation. For each azimuthal position, the actual determined stimulus loudness at the distance of 1.00 m was set as reference. As a result, stimuli loudness of the normalized stimuli was the same for different distances for one specific azimuthal position, but still varied with respect to the different source azimuths. This choice of a different reference loudness per source azimuth resulted in a slightly different stimuli loudness dependent on azimuth. Thus, for $\varphi = 150°$, overall stimuli loudness across distances was about 1.44 dB LKFS (Loudness, K-weighted, relative to full scale) lower than for $\varphi = 30°$, and for $\varphi = 270°$, overall stimuli loudness across distances was about 1.15 dB LKFS higher than for $\varphi = 30°$. Overall, the procedure resulted in 15 gain values for the loudness-normalized stimuli, as these values were dependent on both distance and azimuth, and 5 gain values for the non-normalized stimuli, as these values were only dependent on distance. Each gain value was assigned to the corresponding virtual sound source in the scene description file of the SoundScape Renderer, thus the actual leveling was applied to the convolution result by the renderer. A double-check with a digital audio workstation metering plugin determining loudness according to EBU R128 [40] confirmed equal loudness for all normalized stimuli.

In order to equalize the binaural chain, a headphone compensation filter according to Bernschütz ([41], Chapter 4.3.4) was applied to the pink noise test signal. The filter was a minimum phase FIR filter with 2048 filter taps. By applying this compensation filter, both the magnitude response of the Neumann KU100 – AKG K601 chain was equalized, and also the loudness normalization was maintained, because a non-flat magnitude response of the reproduction system would have affected perceived loudness at the listener's ear. The (equalized) test signal and the HRTFs (or in this case more precisely the corresponding Head Related Impulse Responses [HRIRs]) were all stored on the control computer as 16 bit/48 kHz .wav files. The playback level for the loudness-normalized stimuli was at about $L_{\mathrm{Aeq}} = 61$ dB. For the non-normalized stimuli, this playback level was assigned to a distance of 1.00 m, resulting in a maximum playback level of about $L_{\mathrm{Aeq}} = 79$ dB for the closest distance of 0.25 m ($\varphi = 270°$).

2.1.3 Procedure

We conducted the experiment with naive listeners only. Pilot tests with the normalized stimuli showed strong learning effects: First, test persons could not immediately distinguish between distances, but when they were given

detailed feedback, they learned to differentiate based on spectral changes, varying ILDs, and head movements. However, since our aim was to examine which cues influence natural distance perception in the near field, we only gave basic instructions about the procedure and refrained from a training session or a scale anchoring process.

The experiment was a 2 × 5 × 3 × 2 mixed factorial design with the between-subjects factor *head tracking* (head tracking, static) and the within-subjects factors *distance* (0.25 m, 0.50 m, 0.75 m, 1.00 m, 1.50 m), source *azimuth* (30°, 150°, 270°), and *normalization* (loudness normalization, no loudness normalization and thus distance-related level differences). We decided to use a mixed design instead of a pure within-subjects design because we observed in pilot tests that participants barely moved their head when head tracking was used as a within-subjects factor, even if we encouraged them to do so. It seemed that the random switch between dynamic and static conditions was quite confusing and distracted them from their actual task, which is maybe why they kept their head still. Further pilot tests with head tracking as a between-subjects factor worked as expected.

Each participant had to attend two separate sessions. In the first session, participants had to rate the normalized, in the second session the non-normalized stimuli. In each session, every participant had to rate the five distances for the three different source azimuths, leading to the 5 × 3 × 2 within-subjects factorial design per group.

The perceived distance had to be rated on a continuous scale with 7 anchor points ("very close", "close", "rather close", "medium", "rather distant", "distant", "very distant") in form of a multi-stimulus comparison method. The same and similar scales for ratings of relative perceived distance have already been successfully used in earlier experiments on distance perception [14, 42, 43]. The procedure was as follows. For each trial, a GUI with five value faders ranging from "very close" to "very distant" was displayed on the tablet. Each fader corresponded to one of the five actual measured distances. The source azimuth was the same for all distances (or faders) within a trial. By touching the respective fader, the participants were able to switch between the corresponding stimuli as often as required, thus also allowing for a comparison of the various stimuli (distances). Technically speaking, the HRTF filter-set switched when touching the fader while the pink noise burst sequence was played in a loop. The order of the faders per trial as well as the order of the trials itself were randomized. The procedure was repeated 10 times per azimuth, thus a full run consisted of 30 trials (with five distance ratings per trial).

Participants of both groups were given the exact same instructions. Regardless of whether they performed the experiment with or without head tracking, they were encouraged to move their head during the estimation process in the form of common localization movements, especially if they felt that movements would improve distance perception. However, they had to keep their front viewing direction because of the different source azimuths. In total, each session lasted for about one hour, including the verbal instruction and a short break.

2.1.4 Data analysis

The statistical analysis was based on the mean values per subject, thus the 10 repetitions per subject for each condition were averaged first. A Jarque-Bera test for normality failed to reject the null hypothesis for 45 out of 60 conditions at a significance level of 0.05. With Hochberg correction [44], which is a common method to correct for multiple hypothesis testing, the test failed to reject the null for all conditions. As parametric tests like the ANOVA are generally robust to slight violations of normality assumptions [45], we analyzed the data using a Greenhouse-Geisser (GG) corrected [46] four-way mixed ANOVA with the between-subjects factor head tracking and the within-subjects factors distance, azimuth, and normalization. For a more detailed analysis, several nested (GG-corrected) mixed and repeated measures ANOVAs as well as paired and independent-samples t tests (two-tailed) at a 0.05 significance level were performed on subsets of the data.

As this analysis revealed no significant effects of head tracking and nominal distance (see the results below), we further analyzed the data using Bayes factors (BF, here BF_{01}). In contrast to common null-hypothesis significance testing, BFs allow stating evidence in favor of the null hypothesis [47, 48]. In particular, the reported BFs are based on independent-samples t tests or nested repeated measures ANOVAs for all effects of additional importance. In brief, BF_{01} expresses the likelihood of the null hypothesis relative to the likelihood of the alternative hypothesis given the data. Thus, for a example, a $BF_{01} = 3$ would suggest that the data provide three times as much evidence for the null than for the alternative. The Bayesian t tests were conducted according to Rouder et al. [48], using the Jeffrey-Zellner-Siow (JZS) prior with a scaling factor of $r = .707$. The BF for the repeated measures ANOVA was calculated according to Rouder et al. [49] using the same prior assumptions.

2.2 Results

Figure 3 shows the results of Experiment 1. The data are separated with respect to the between-subjects factor head tracking and the within-subjects factor normalization, resulting in four subsets: Head Tracking – Loudness Normalization (HTNorm), Head Tracking (HT), Static – Loudness Normalization (STNorm), and Static (ST).

The mean plots in Figure 3 (left) show three notable patterns, statistically confirmed by the analysis further below: (a) There is no apparent effect of head tracking on estimated distance, which can be seen by comparing dynamic (HTNorm, HT) and static (STNorm, ST) conditions. This indicates that head movements had no significant influence on distance estimation of nearby virtual sound sources. (b) Participants did not accurately rate distance of the normalized stimuli (HTNorm and STNorm), suggesting that they could not exploit the non-individual binaural distance cues with the applied loudness normalization method. (c) As expected, participants rated according to the nominal (i.e., actually measured) distance if the

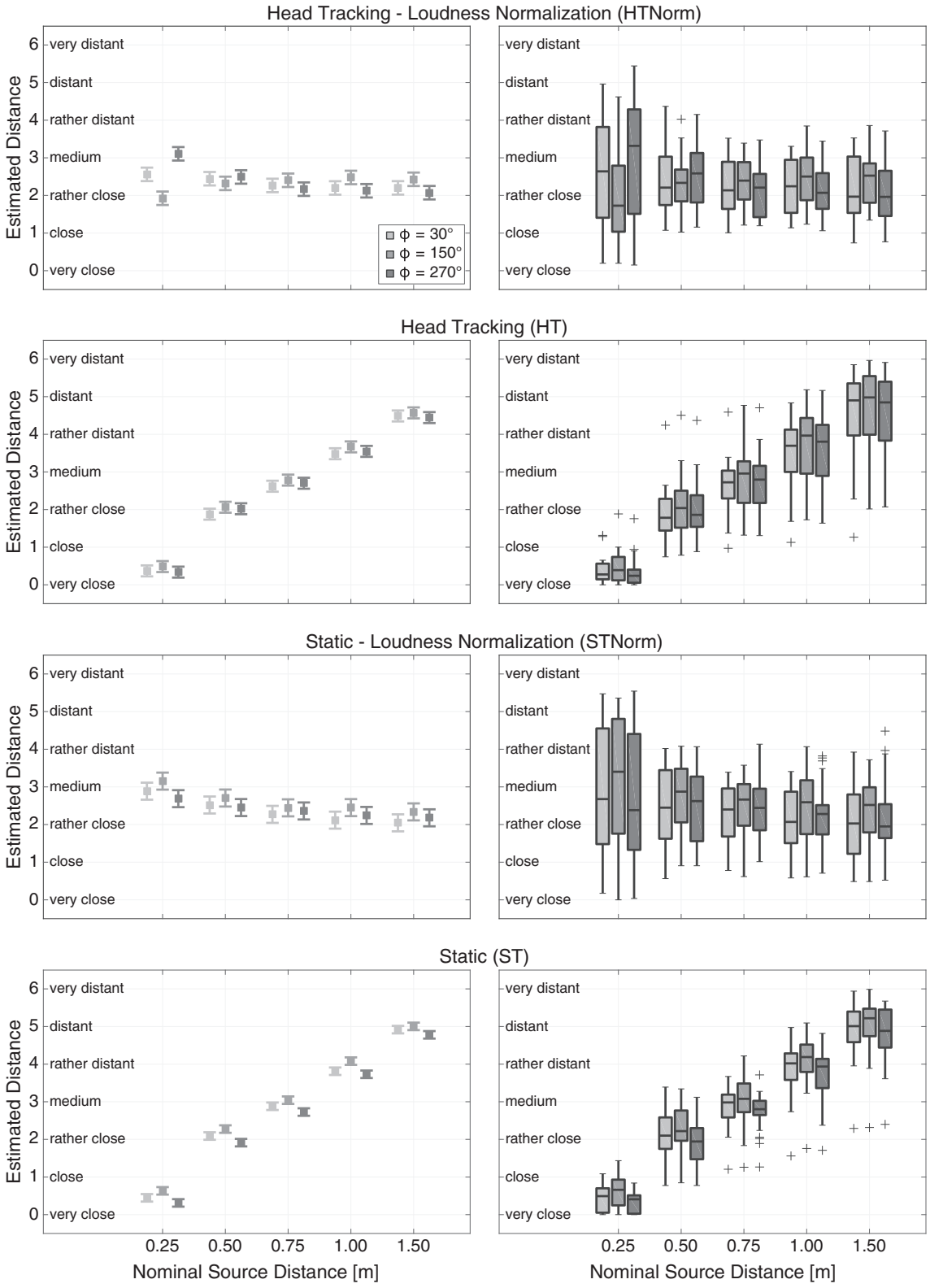

Figure 3. Mean estimated distances (left) and interindividual variation in the estimated distances (right) as a function of nominal source distance (abscissa) and nominal source azimuth (shades of gray) for the subsets: Head Tracking – Loudness Normalization (HTNorm), Head Tracking (HT), Static – Loudness Normalization (STNorm), and Static (ST). The error bars in the mean plots (left) display 95% within-subjects confidence intervals [50, 51], based on the error term of the respective distance main effect. The box plots (right) show the median and the (across participants) interquartile range (IQR) per condition; whiskers display 1.5 × IQR below the 25th or above the 75th percentile and outliers are indicated by plus signs.

Table 1. Results of the four-way mixed ANOVA with the between-subjects factor head tracking (HT) and the within-subjects factors distance (Dist), azimuth (Az), and normalization (Norm).

Source	*df*	*F*	*MSE*	ε	η_p^2	*p*
Between-subjects						
HT	1, 48	.88	11.06	–	.02	.35
Within-subjects						
Dist	4, 192	149.16	1.03	.32	.76	<.001*
Dist × HT	4, 192	.12	1.03	.32	0	.794
Az	2, 96	9.82	.33	.98	.17	<.001*
Az × HT	2, 96	6.56	.33	.98	.12	.002*
Norm	1, 48	15.54	2.71	1	.25	<.001*
Norm × HT	1, 48	.36	2.71	1	.01	.551
Dist × Az	8, 384	5.59	.08	.43	.10	.001*
Dist × Az × HT	8, 384	12.13	.08	.43	.20	<.001*
Dist × Norm	4, 192	219.68	1.15	.31	.82	<.001*
Dist × Norm × HT	4, 192	1.43	1.15	.31	.03	.241
Az × Norm	2, 96	2.37	.24	.97	.05	.101
Az × Norm × HT	2, 96	1.92	.24	.97	.04	.153
Dist × Az × Norm	8, 384	7.55	.07	.38	.14	<.001*
Dist × Az × Norm × HT	8, 384	14.31	.07	.38	.23	<.001*

Note. ε = Greenhouse-Geisser (GG) epsilon, *p* = GG-corrected *p*-values. Note that GG correction is appropriate only for within-subject tests, with more than one degree of freedom in the numerator.
*$p < .05$.

stimuli were not normalized in loudness (HT and ST) and thus natural distance-related intensity cues were provided. Also worth noting are the ratings for HTNorm at $d = 0.25$ m, since participants rated especially the stimuli with $\varphi = 270°$ (erroneously) further away than all other sources. Moreover, the box plots in Figure 3 (right) reveal that the between-subjects variance of the normalized stimuli (especially at $d = 0.25$ m) is considerably higher than for the non-normalized stimuli. This suggests that the normalized conditions might have provided conflicting and ambiguous distance cues that were interpreted or weighted differently by different individuals. In particular, participants might have been confused by binaural cues indicating a nearby sound source in the absence of matching intensity cues.

Table 1 shows the results of the GG-corrected four-way mixed ANOVA. In line with observation (a) made on the basis of the plots, no significant between-subjects effect of head tracking was found, which is also reflected by an independent-samples *t* test on data averaged across all within-subject conditions [$t_{\text{Groups}}(48) = 0.94$, $p = .352$, $d = .27$, $BF_{01} = 2.46$]. This was further confirmed by two independent-samples *t* tests separately testing subsets with and without normalization for an effect of head tracking (HTNorm vs. STNorm, HT vs. ST, averaged across all remaining factors). Both tests yielded no significant difference between the subsets [$t(48) = 0.53$, $p = .60$, $d = .15$, $BF_{01} = 3.15$]; [$t(48) = 1.23$, $p = .227$, $d = .35$ $BF_{01} = 1.92$]. Based on the BFs, the data provided about 2–3 times more evidence for the null than for the alternative, indicating that head tracking did not influence distance estimation performance whether loudness was normalized or not.

The mixed ANOVA revealed several significant within-subjects main and interaction effects though, like for example a rather complex significant four-way interaction between distance, azimuth, normalization, and head tracking. For a better interpretation of the results, we therefore analyzed the data using several nested ANOVAs. In particular, we conducted a GG-corrected two-way repeated measures ANOVA with the within-subjects factors distance and azimuth for each subset. Table 2 summarizes the results of these ANOVAs.

In line with observation (b) made on the basis of the plots, the standard repeated measures ANOVA showed no significant distance effect in the subset HTNorm. For this effect, the respective Bayesian ANOVA revealed $BF_{01} = 5.10$, suggesting that the data of subset HTNorm provide about five times more evidence for the absence (rather than presence) of an effect of distance. For subset STNorm, the main effect of distance from the standard ANOVA was not significant. The Bayesian ANOVA however provided 29,783 more evidence for the presence (rather than absence) of an effect of distance ($BF_{01} = 3.36 \times 10^{-5}$). As evident in Figure 3, this main effect reflected a negative trend of distance, that is, sources that were nominally further away were perceived as closer.

Furthermore, the results of the repeated measures ANOVAs yielded a strong distance × azimuth interaction in subset HTNorm, with an effect of azimuth present mainly for $d = 0.25$ m and to a minor degree for $d = 0.50$ m, and largely absent for the other distances, as revealed by means of further nested ANOVAs (for the sake of conciseness, these one-way repeated measures ANOVAs are not reported here). A paired *t* test yielded a significant difference between the ratings for $d = 0.25$ m and $d = 0.50$ m with $\varphi = 270°$ [$t(24) = 2.67$, $p = .013$, $d_z = .53$], confirming that participants rated the closest source with $\varphi = 270°$ further away than all the other sources. A comparison between the ratings for $d = 0.25$ m and $d = 0.50$ m with $\varphi = 150°$ or $\varphi = 30°$ using paired *t* tests yielded no significant

Table 2. Results of the two-way repeated measures ANOVAs for the subsets HTNorm, HT, STNorm, ST, each with the within-subjects factors distance (Dist) and azimuth (Az).

Source	*df*	*F*	*MSE*	ε	η_p^2	*p*
HTNorm						
Dist	4, 96	.96	1.22	.29	.04	.347
Az	2, 48	.45	.49	.96	.02	.633
Dist × Az	8, 192	18.21	.15	.27	.43	<.001*
HT						
Dist	4, 96	226.87	.81	.34	.90	<.001*
Az	2, 48	6.86	.10	.86	.22	<.001*
Dist × Az	8, 192	1.75	.02	.63	.07	.126
STNorm						
Dist	4, 96	3.21	1.94	.27	.12	.083
Az	2, 48	5.79	.42	.89	.19	.008*
Dist × Az	8, 192	1.87	.10	.38	.07	.142
ST						
Dist	4, 96	570.60	.38	.37	.96	<.001*
Az	2, 48	24.15	.13	.99	.50	<.001*
Dist × Az	8, 192	1.34	.03	.54	.05	.259

Note. ε = Greenhouse-Geisser (GG) epsilon, p = GG-corrected p-values. Note that GG correction is appropriate only for within-subject tests, with more than one degree of freedom in the numerator.
*$p < .05$.

differences [$t(24) = 2.02$, $p = .055$, $d_z = .40$], [$t(24) = 0.49$, $p = .626$, $d_z = .10$], that is, in subset HTNorm, the source at $d = 0.25$ m and $\varphi = 270°$ was the only source rated significantly different in distance compared to all other sources.

Moreover, there was a main effect of azimuth in subset STNorm without a significant distance × azimuth interaction, but further nested ANOVAs for this subset showed significant azimuth effects only for conditions with $d = 0.25$ m and $d = 1.00$ m, indicating that the influence of source azimuth on estimated distance is relatively small for the remaining levels of distance in this subset.

For subset HT and ST, the ANOVAs revealed highly significant main effects of distance, confirming observation (c) made on the basis of the plots. As expected, participants were able to distinguish distance for the non-normalized stimuli, thus validating the employed procedure. Moreover, the results yielded a rather strong main effect of azimuth for both subsets. These effects can also be seen in the plots, as the means vary in some kind of triangular pattern with respect to azimuth.

To analyze the three-way interaction between distance, azimuth, and head tracking, as well as the two-way interaction between azimuth and head tracking, we first compared the conditions with normalized stimuli (subsets HTNorm and STNorm) as a function of head tracking. Two-way mixed ANOVAs for each level of distance with the between-subjects factor head tracking and the within-subjects factor azimuth showed a highly significant azimuth × head tracking interaction for conditions with $d = 0.25$ m, and a small significant interaction for conditions with $d = 0.50$ m. The plots in Figure 3 clearly illustrate this interaction effect, as the values at $d = 0.25$ m and $d = 0.50$ m vary significantly dependent on head tracking and show opposing patterns (compare HTNorm and STNorm). Similar ANOVAs for the subsets HT and ST revealed a significant azimuth × head tracking interaction for all levels of distance except for $d = 1.50$ m. A look at the plots clarifies this interaction effect, since especially the means at $\varphi = 270°$ are notably higher for conditions with head tracking than for conditions without head tracking.

To further unpack the four-way interaction between distance, azimuth, normalization, and head tracking, we conducted two-way nested mixed ANOVAs with the between-subjects factor head tracking and the within-subjects factor normalization for each combination of levels of the factors distance and azimuth. Here, the results showed a slightly significant interaction between normalization and head tracking for only two conditions ($d = 1.50$ m, $\varphi = 30°$ and $d = 0.25$ m, $\varphi = 150°$), suggesting a rather small influence of the normalization × head tracking interaction in the context of the entire dataset.

To validate that the option to move the head was actually used more often when it had an effect on the stimulation, we also compared the head movements between groups with and without head tracking (groups HT and ST). For each participant, the standard deviation of the horizontal viewing directions (azimuth) for all conditions was calculated, leading to 25 values per group representing the amount of variation around the viewing direction of 0° (front viewing direction). The averaged standard deviation among the group with head tracking was 23.98° and 13.56° among the group without head tracking. An independent-samples t test revealed a significant difference between the two groups [$t(48) = 3.39$, $p = .001$, $d = .96$], indicating that the group with head tracking moved their head to a significantly higher degree.

2.3 Discussion

The most interesting results of Experiment 1 are that nominal distance did not significantly affect distance ratings for the normalized stimuli (except for the condition $d = 0.25$ m, $\varphi = 270°$ in the subset HTNorm) and that adapting stimuli to the current head position (head tracking) had no significant influence on estimated distance even though it influenced the degree to which participants moved their heads. The findings indicate that in most conditions, the naive listeners did not use the variations in binaural cues, whether induced by a change in nominal distance of the virtual sound source, or by a change in head orientation. The non-significant distance effect as well as the estimated Bayes factor $BF_{01} = 5.10$ for this effect for conditions with normalized stimuli and dynamic binaural rendering (subset HTNorm) generally support this assumption. Only the significant distance × azimuth interaction in the same subset, mainly driven by an effect of azimuth for $d = 0.25$ m, could be attributed to an effect of binaural cues. Surprisingly, however, the most nearby source with $\varphi = 270°$ for this distance in subset HTNorm was rated as being the furthest away. Furthermore, for subset STNorm, perceived distance decreased with nominal distance. These two, at first sight counterintuitive, findings can be better explained by the workings of intensity cues rather than binaural cues, as revealed by Experiment 3 (see Sect. 4).

The findings regarding the importance of binaural cues conflict with the well know results from Brungart et al. [7, 11, 12], who concluded that especially low-frequency ILDs are an important binaural distance cue for nearby sound sources in real or virtual acoustics. Even though the experiments differ in many ways, like for example the general setup and procedure (e.g., Brungart et al. [17] roved the level of the stimuli in their experiment on absolute distance judgments, in order to diminish potentially remaining intensity cues), the employed HRTFs, or the applied binaural synthesis (Brungart and Simpson [12] employed KEMAR HRTFs and only used static binaural synthesis), it is not directly obvious why the findings are so different. It appears possible that the test procedure and the method to eliminate intensity cues in the stimuli have a major impact on the results (see Sect. 5 for a detailed discussion).

The observed main effect of azimuth can, to some extent, be explained by the slight differences in stimuli loudness dependent on the azimuthal position, as described in Section 2.1.2. The data show a triangular pattern as a function of azimuth (see Fig. 3), except for the above-discussed conditions $d = 0.25$ m and $d = 0.50$ m in subset HTNorm, where this pattern appears inverse, leading to the distance × azimuth interaction effect in subset HTNorm. Thus, in most cases, participants rated conditions with $\varphi = 150°$ a little bit further away than the stimuli with $\varphi = 30°$, most likely because the stimuli with $\varphi = 150°$ were about 1.44 dB LKFS lower in loudness level than the stimuli with $\varphi = 30°$. On the opposite, participants mostly rated the conditions with $\varphi = 270°$ a little bit closer then the stimuli with $\varphi = 30°$, as the stimuli with $\varphi = 270°$ were about 1.15 dB LKFS higher in loudness level than the stimuli with $\varphi = 30°$. These effects of azimuth occurred especially for conditions without normalization.

3 Experiment 2

In Experiment 1, nominal distance had no significant effect on distance ratings for the normalized stimuli except for a single condition, suggesting that participants mostly did not use binaural cues, which vary strongly with nominal distance, for distance estimation. To further verify this somewhat surprising outcome, we conducted a psychophysical two-alternative forced choice (2AFC) test, providing a more sensitive test than the previously used method. In the forced-choice procedure, participants had to judge the relative perceptual distance between two virtual (nearby) sound sources. If at all relevant for distance estimation, barely perceptible differences possibly included in the near-field HRTFs would be easier to detect in such a direct comparison than in the multiple-stimulus test used in Experiment 1. Since the focus of Experiment 2 was on binaural cues, we only examined conditions with loudness normalization and head tracking.

3.1 Method

3.1.1 Participants

Seventy-three participants with an age between 19 and 47 years took part in Experiment 2 (58 males, 15 females, $M = 23.37$ years, $Mdn = 23$ years, $SD = 4.21$). All of them were students in media technology and participated for course credit. None of them had participated in Experiment 1 and they were all naive as to the purpose of this study. Thus, as in the previous experiment, participants had to rely on their life experience in distance perception of nearby sources. Here, only five participants (7%) already had experience with the binaural reproduction system. This small number resulted from the fact that most of the subjects from our commonly used subject pool had already participated in Experiment 1 and thus were not allowed to take part in the second experiment. All participants had self-reported normal hearing.

3.1.2 Setup and stimuli

Head-tracking and loudness normalization were used throughout Experiment 2. In all other aspects, technical setup and stimuli were identical to Experiment 1 (see Sect. 2.1.2).

3.1.3 Procedure

The experiment was a 4 × 3 × 2 within-subjects factorial design with the factors *distance pair* (0.25 m vs. 0.50 m, 0.25 m vs. 0.75 m, 0.25 m vs. 1.00 m, 0.25 m vs. 1.50 m), source *azimuth* (30°, 150°, 270°), and *presentation order* (close – far, far – close). Distances were always compared with reference to the closest distance ($d = 0.25$ m), because the ILD and the spectrum differ most strongly from the

other distances. The factor azimuth describes the azimuthal position of the two virtual sources to be compared, and the factor presentation order describes whether the closest sound source at $d = 0.25$ m was presented first (close – far) or last (far – close).

The procedure of the experiment was as follows. On each trial, a sequence composed of four stimuli was presented. In this sequence, the first and the last two stimuli were always the same, resulting in two stimulus pairs which had to be compared. Thus, the two to-be-compared distances (A and B) were presented twice (A–A–B–B). Each stimulus pair had a total length of 3.50 s (2 × stimulus of 1.50 s + 0.50 s interstimulus interval). Between both stimulus pairs, there was an interstimulus interval of 1.00 s, resulting in a playback time of 8.00 s for each trial. Similar to Experiment 1, we decided to use a rather long test signal as well as stimulus repetitions to provide enough time to move the head during playback.

After playback, participants had to report whether they perceived the second stimulus pair closer or further away than the first one, by pressing the corresponding button on the GUI presented on the tablet. The two buttons were arranged on a vertical line, with the upper one labeled "further away" and the lower one labeled "closer". Participants could neither repeat a trial nor continue without giving an answer. After a response was registered, the next trial followed immediately. A full run consisted of 12 trials per condition, leading to a total of 24 (conditions) × 12 (trials) = 288 trials. The order of conditions was randomized for each participant.

Before starting the test, participants were given instructions about the general procedure. Since most of the participants were new to the field of virtual acoustics, the instruction also included a brief introduction on dynamic binaural synthesis. Furthermore, as in Experiment 1, they were encouraged to perform localization movements with their head if they felt that distance perception improved when doing so. At the same time, because of the different source positions, they were instructed to keep their main line of vision straight ahead and they were not allowed to turn their body, e.g., sideways or to orientate themselves to the sound source. After the instructions, participants conducted a short training session composed of six trials to get familiar with dynamic binaural synthesis and with the test procedure. Altogether, the experiment took about 1 h, including the verbal instruction, the training session, and a short break after half of the trials.

3.1.4 Data analysis

For each subject, the 12 repetitions per condition were averaged first, leading to a quasi-metric variable with a value between 0 and 1 describing the proportion of correct answers. These proportion data follow a binomial distribution, where generally the variance is a function of the mean and variances tend to be small at both ends of the range but large in the middle. As a consequence, it is questionable to use parametric tests with raw proportions, since the assumption of normality and homogeneity of variance might be violated to a certain extent, even though the usual parametric tests like t test or ANOVA are robust to these violations. To lessen this issue, we applied an arcsine square root transformation to the raw data, which is a typical procedure for proportions. The transformation removes the correlation between means and variances and stretches out both ends of the distribution of proportions while compressing the middle, resulting in homogenized variance and improved normality ([52], Chapter 10.2). The statistical analysis, which was quite similar to the one performed in Experiment 1, was conducted using the transformed data.

We analyzed the transformed data using a three-way repeated measures ANOVA with the within-subjects factors distance pair, azimuth, and presentation order. A Jarque-Bera test for normality failed to reject the null hypothesis for 19 out of 24 conditions. With Hochberg [44] correction, the test failed to reject the null for all conditions. We nevertheless corrected for slight violations of ANOVA assumptions using the GG correction [45]. To analyze the data in greater detail, we conduced several nested (GG-corrected) repeated measures ANOVAs as well as one-sample t tests (two-tailed) at a 0.05 significance level. All t tests were corrected for multiple hypothesis testing using the Hochberg [44] method.

3.2 Results

Figure 4 shows the results of the experiment. We separated the data with respect to the factor presentation order, resulting in two subsets, further labeled Close – Far (CF) and Far – Close (FC). For better interpretability, the plots show the raw instead of the transformed data. Most strikingly, Figure 4 (left) shows that all means were below chance level. This suggests that participants did hear a difference between the stimuli, but, surprisingly, perceived the closer source as farther away and vice versa. The corresponding box plots in Figure 4 (right) reveal a rather high variance of the results with whiskers often covering the entire range of proportions, which again shows that different participants interpreted or weighted the available distance cues differently.

The ANOVA summarzied in Table 3 yielded a significant main effect of distance pair and significant interaction effects between distance pair and azimuth as well as between azimuth and presentation order. We further analyzed if the condition means differ significantly from chance by conducting 24 one-sample t tests against arcsine-square root transformed chance level (0.7854), revealing significant deviations from chance for all conditions (all $ps < .001$).

To unpack the observed main and interaction effects (see Tab. 3), we conducted several nested ANOVAs. Overall, the main effect of distance pair was present in both subsets, but appears to be much stronger in subset CF. As regards the main effect of distance, performance decreased with increased distance between the virtual sound sources and consequently with intensified inter-stimulus differences. Paradoxically, this decreased performance can be explained by participants perceiving clearer differences between the stimuli with increased distance between the virtual sound

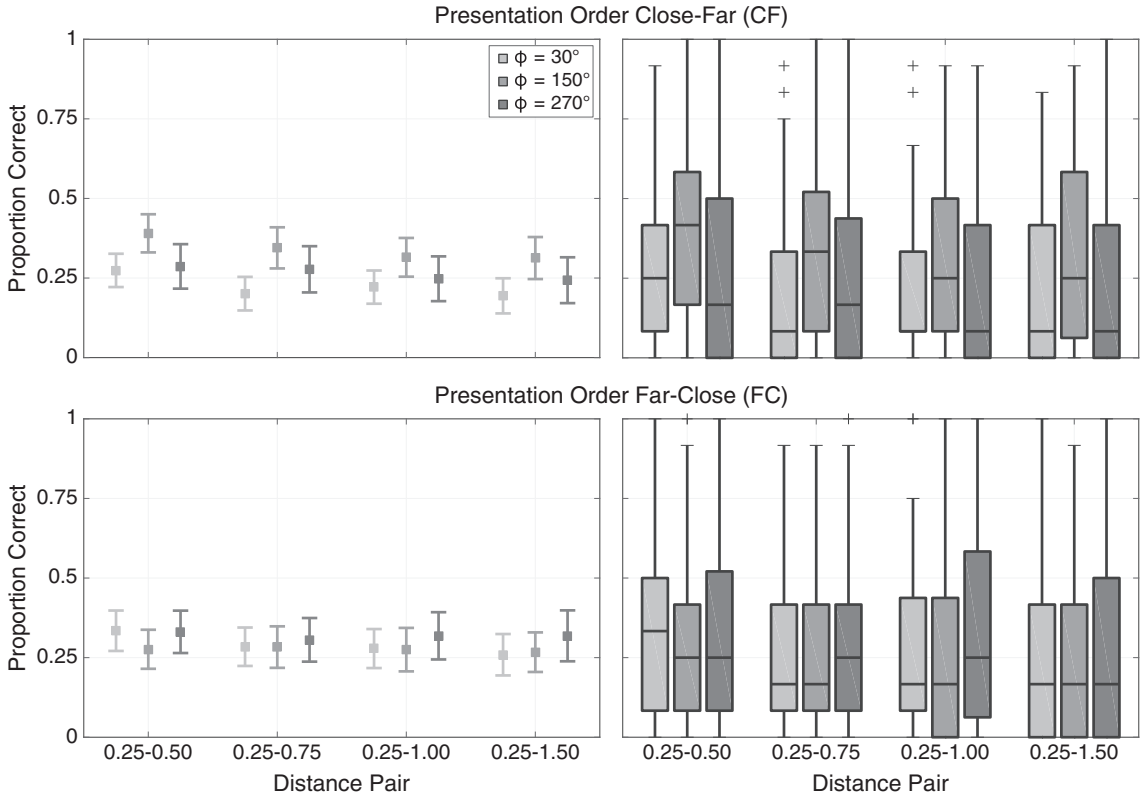

Figure 4. Mean proportions of correct answers (left) and interindividual variation in the proportions (right) as a function of distance pair (abscissa) and nominal source azimuth (shades of gray), separated with respect to presentation order: Close – Far (CF) and Far – Close (FC). For better interpretability, the plots show the raw data instead of the transformed data that was submitted to statistical testing. The error bars in the mean plots (left) display 95% confidence intervals based on the respective one-sample t tests comparing the means of the raw data against non-transformed chance level of 0.5. The box plots (right) show the median and the (across participants) interquartile range (IQR) per condition; whiskers display 1.5 × IQR below the 25th or above the 75th percentile and outliers are indicated by plus signs.

Table 3. Results of the three-way repeated measures ANOVA with the within-subjects factor distance pair (DP), azimuth (Az), and presentation order (PO).

Source	*df*	*F*	*MSE*	ε	η_p^2	*p*
DP	3, 216	21.28	.03	.97	.23	<.001*
Az	2, 144	1.52	.41	.92	.02	.223
PO	1, 72	2.35	.18	1	.03	.130
DP × Az	6, 432	2.31	.03	.94	.03	.036*
DP × PO	3, 216	2.46	.03	.97	.03	.066
Az × PO	2, 144	23.71	.07	.93	.25	<.001*
DP × Az × PO	6, 432	1.53	.03	.95	.02	.169

Note. ε = Greenhouse-Geisser (GG) epsilon, p = GG-corrected p-values. Note that GG correction is appropriate only for within-subject tests, with more than one degree of freedom in the numerator.
*$p < .05$.

sources. The error rate increased because they misinterpreted the provided cues or because wrong cues were given.

Regarding the rather weak distance pair × azimuth interaction effect, two nested ANOVAs indicated a small-but-significant interaction effect between those two factors and a strong main effect of azimuth in subset CF, whereas there was no main effect of azimuth or interaction in subset FC. This can easily be seen in the mean plots, as the values vary clearly as a function of azimuth in Figure 4 (CF – left), but seem to be almost independent of azimuth in Figure 4 (FC – left). These observations also explain the highly significant interaction effect between azimuth and presentation order (Tab. 3). In particular, when the closer source was presented first, the source azimuth had a significant influence on the number of correct answers. Apparently, in this case, the performance for virtual sound sources with an azimuthal position of 30° was much more below chance level than for sources with an azimuthal position of 150°. As opposed to this, the source azimuth had no significant influence on proportions of correct answers if the farther source was presented first.

3.3 Discussion

In line with Experiment 1, participants of Experiment 2 did not correctly employ the strong changes in non-individual binaural cues induced by a variation in nominal sound source distance. Rather, participants predominantly made false responses, which resulted in mean values

significantly below chance level for each tested condition. This indicates that participants did perceive a difference between the stimuli, but made the wrong conclusion with regard to their relative distance. These results might tentatively be explained in two different ways: (a) incorrect use of spectral cues or (b) overcorrection by the normalization method.

The first tentative explanation is based on the signal properties of nearby sound sources: As briefly outlined in Sections 1 and 2.1.3, low frequencies increase relative to high frequencies as a sound source approaches the head, leading to a low-pass filtering character of nearby sound sources. Consequently, stimuli with a distance of 0.25 m were always more dull than stimuli at any other distance. Hence, participants might have most often classified the duller stimulus as the farther source, and the brighter stimulus as the closer source. This assumption is in line with studies from Butler et al. [53] or Little et al. [54], who showed that sounds with decreased high-frequency components relative to low-frequency components are perceived to be further away. Thus, spectrum has a dual role in estimation of distance, as already revealed by Coleman [55], since the relative decrease of high frequency components can be a cue for a source nearby or far away. A dominance of the spectral cue (and the misinterpretation of the provided spectral differences) would also explain the strong effect of distance pair found in the statistical analysis. As outlined above, the proportions decreased as a function of distance pair, which suggests that participants perceived stronger differences with increased distance between the sources. In fact, the spectral differences increase as the distance between the sources becomes larger.

Alternatively, remaining intensity differences between the stimuli may be responsible for the counterintuitive results: As the participants mostly rated the sources at $d = 0.25$ m as further away, it could simply be that the stimuli with a distance of 0.25 m were perceived as slightly quieter than all other stimuli. The perceived loudness differences between the stimuli might have increased as a function of distance pair, which would explain the strong effect of distance pair. Of course, it is also possible that intensity and spectral cues both contributed to the effect. However, if sufficiently strong, the wrong intensity cues most probably masked the spectral cues. As none of the effects was clearly evident in Experiment 1, we assume that differences in spectrum or loudness are more salient in a direct comparison, as provided in Experiment 2, and kind of blur in a multi-stimulus comparison, as in Experiment 1.

Concerning the observed interaction effect between azimuth and presentation order, we could not find any plausible explanation. Thus, it is uncertain why especially the proportions for stimuli with $\varphi = 150°$ were closer to chance level if the closer source was presented first (presentation order close – far). Based on these observations, we can only generally conclude that there were less perceptible differences between these specific stimuli. However, the findings cannot be explained by any signal properties of the stimuli and a detailed exploration of the described effect is beyond the scope of the present study. Indeed, the influence of source azimuth and presentation order on perceived distance of nearby sound sources remains an interesting research question for further studies.

4 Experiment 3

To clarify whether a difference in perceived loudness might explain the surprising below-chance performance in Experiment 2, Experiment 3 directly tested for differences in perceived loudness after normalization. We examined both loudness normalization as employed in this study and amplitude normalization according to Brungart et al. [7] as employed in many previous experiments. To test for perceptible loudness differences between the stimuli, we performed comparison tests according to the SAQI test paradigm [43]. In particular, participants had to judge the relative perceptual loudness difference between two virtual (nearby) sound sources on a bipolar seven-point scale. Similar to Experiment 2, we only examined conditions with (the respective) normalization and head tracking. Regarding the outcome of this follow-up study, we expected to observe that the loudness normalization reduces perceptible loudness differences better than the amplitude normalization.

4.1 Method

4.1.1 Participants

Seventeen male students in media technology or electrical engineering with an age between 19 and 42 years ($M = 24.12$ years, $Mdn = 22$ years, $SD = 5.66$) participated in the experiment on a voluntary basis. Five of them had already participated in Experiment 1, but none of them had taken part in Experiment 2, which was the more recent experiment with largely similar stimuli and procedure. Eight participants (47%) already had experience with the binaural reproduction system and the test environment. All of them reported normal hearing and were naive as to the purpose of this study.

4.1.2 Setup and stimuli

As this was a follow-up study, the experimental setup, the test signal, and the HRTFs were exactly the same as in Experiments 1 and 2 (refer to Sect. 2.1.2). All conditions were with head tracking and the respective normalization method. To get the additional gain values for the amplitude-normalized stimuli, the scaling factor S according to Brungart et al. [7] was calculated for each of the 15 positions (5 distances and 3 directions). This factor is based on the distance of the source from the left and right ears of the listener, such as $S = 1/((50/d_l) + (50/d_r))$, where d_l and d_r is equal to the distance in cm to the left and right ear respectively. The distance between both ears was defined as 0.20 m. The calculated scaling factors were then added to the gain values for the non-normalized stimuli, resulting in amplitude-normalized stimuli when being rendered.

To ensure good comparability between the normalization methods, the gain values were set to the same values at the reference distance of 1.00 m. For stimuli closer or farther away, the gain values obviously differed between the two normalization methods. At the closest distance of 0.25 m, the differences in gain values were greatest. Depending on the direction, the gain values for the amplitude normalization were about 2–4 dB higher than for the loudness normalization in this case.

4.1.3 Procedure

In addition to the independent variables considered in the previous experiment, Experiment 3 involved the two different normalization methods. This resulted in a $4 \times 3 \times 2 \times 2$ within-subjects factorial design (48 conditions) with the factors *distance pair* (0.25 m vs. 0.50 m, 0.25 m vs. 0.75 m, 0.25 m vs. 1.00 m, 0.25 m vs. 1.50 m), source *azimuth* (30°, 150°, 270°), *presentation order* (close – far, far – close), and *normalization method* (loudness, amplitude).

The procedure according to the SAQI assessment for loudness was as follows. On each trial, two different stimuli were presented successively. Corresponding to the stimulus length of 1.50 s and an interstimulus interval of 0.50 s, the total playback time of each trial was 3.50 s. In contrast to Experiment 2, each stimulus was only presented once. However, the total length of the test signal was exactly the same, again to provide enough time for potential head movements and to assure comparability with the other experiments.

After each trial, participants had to rate if they perceived the second stimulus louder or quieter than the first one. The size of the perceived loudness difference had to be given on a bipolar seven-point scale with the comparative scale ends named quieter and louder. The scale was numbered from 0 to 3, with 0 being in the middle and 3 being at both scale ends. It was displayed on the GUI in form of a vertically aligned continuous fader, thus selecting interim values between the given numbers was possible. To avoid a bias towards a specific scale range, the fader knob was reset to 0 (center position) at the beginning of each trial. Thus, if no loudness difference between both stimuli could be perceived, the fader knob could simply remain untouched. By pressing a button displayed on the GUI, participants could continue to the next trial. Each trial was only presented once and participants could not repeat the playback. A full run consisted of 6 trials per condition, resulting in a total of 48 (conditions) $\times$ 6 (trials) $=$ 288 trials. The order of conditions was randomized for each participant.

Before starting the test, participants were given instructions about the general procedure. Participants new to the field of virtual acoustics were also briefly introduced into binaural reproduction technology. Similar to the previous experiments, participants were allowed to turn their head, but they were instructed to keep their front viewing direction and not to turn their body. At the beginning of the test, participants had to conduct a short training session composed of six trials. This way, they could get familiar with the procedure, binaural rendering, and the loudness range of the stimuli. In total, each test session took about one hour, including the verbal instruction, the training trials, and a short break after half of the test.

4.1.4 Data analysis

The statistical analysis was based on normalized mean values per subject. Thus, the 6 repetitions per subject for each condition were averaged first and then normalized to the range from -1 to 1. A Jarque-Bera test for normality failed to reject the null hypothesis for 40 out of 48 conditions. With Hochberg [44] correction, the test failed to reject the null for all conditions. As the ANOVA is very robust to small violations of its assumptions [45], we conducted a GG-corrected four-way repeated measures ANOVA with the within-subjects factors distance pair, azimuth, presentation order, and normalization method. For further analysis, we performed several nested (GG-corrected) repeated measures ANOVAs as well as one-sample t tests (two-tailed) at a 0.05 significance level. To compensate for multiple hypothesis testing, all t tests were corrected using the Hochberg [44] method.

4.2 Results

Figure 5 shows the results pooled over presentation order and separated with respect to normalization method. As can be seen, the participants rated the more distant sources louder than the closer ones if loudness normalization was applied. In contrast, they perceived the more distant sources quieter than the closer ones for the conditions with amplitude normalization. Thus, the plots indicate that both tested normalization methods did not work properly and furthermore led to conflicting results.

The ANOVA summarized in Table 4 yielded a significant main effect of distance pair and normalization method, but no significant main effect of azimuth or presentation order. Furthermore, the analysis revealed a significant interaction effect between azimuth and normalization method, which is of particular interest here, as well as several other two- and three-way interaction effects, which we refrain from discussing in detail in the following in order to focus on the main outcome of the experiment.

To more directly test whether the loudness differences remaining after normalization are significant, we performed 24 one-sample t tests comparing the respective results of the pooled conditions against zero. For 20 conditions, the t tests yielded a significant difference between the respective condition mean and zero ($p < .001$ for all). Only the conditions distance pair 0.25 m vs. 0.75 m, amplitude normalization at all three levels of azimuth as well as the condition distance pair 0.25 m vs. 1.00 m, amplitude normalization, $\varphi = 150°$ were not significantly different from zero.

Furthermore, the pattern of the effect of distance pair differs between the two normalization methods. Whereas the results for loudness normalization have an almost constant offset from zero with only a slight slope as the

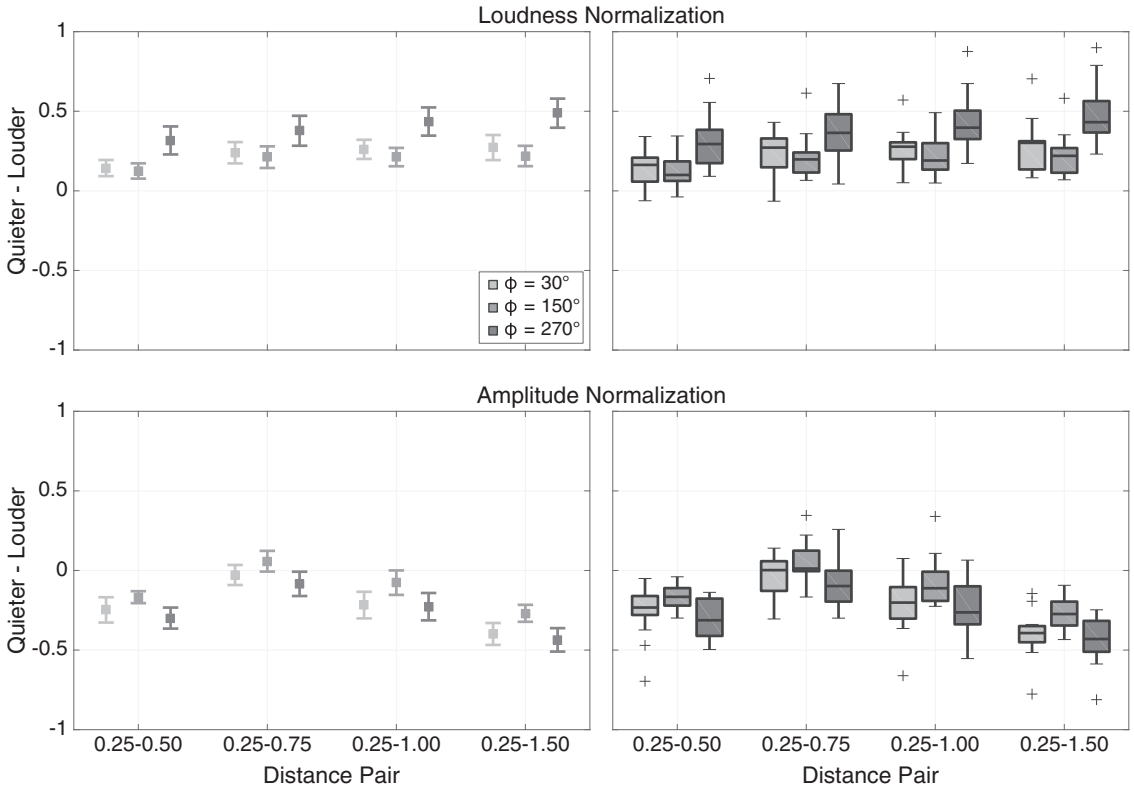

Figure 5. Mean ratings of the SAQI assessment for loudness (left) and interindividual variation in the ratings (right) as a function of distance pair (abscissa) and nominal source azimuth (shades of gray), pooled over presentation order and separated with respect to normalization method. The error bars in the mean plot (left) display 95% confidence intervals based on the respective one-sample t tests comparing the condition means against 0. The box plots (right) show the median and the (across participants) interquartile range (IQR) per condition; whiskers display 1.5 × IQR below the 25th or above the 75th percentile and outliers are indicated by plus signs.

Table 4. Results of the four-way repeated measures ANOVA with the within-subjects factor distance pair (DP), azimuth (Az), presentation order (PO), and normalization method (NM).

Source	*df*	*F*	*MSE*	ε	η_p^2	*p*
DP	3, 48	61.72	.02	.52	.79	<.001*
Az	2, 32	3.03	.11	.99	.16	.063
PO	1, 16	.01	.12	1	0	.919
NM	1, 16	223.47	.21	1	.93	<.001*
DP × Az	6, 96	1.07	.01	.65	.06	.377
DP × PO	3, 48	8.63	.01	.83	.35	<.001*
Az × PO	2, 32	16.27	.05	.62	.50	<.001*
DP × NM	3, 48	80.54	.02	.69	.83	<.001*
Az × NM	2, 32	92.81	.02	.98	.85	<.001*
PO × NM	1, 16	53.45	.04	1	.77	<.001*
DP × Az × PO	6, 96	3.21	.01	.81	.17	.012
DP × Az × NM	6, 96	1.93	.01	.74	.11	.109
DP × PO × NM	3, 48	7.54	.01	.79	.32	.001*
Az × PO × NM	2, 32	6.35	.02	.78	.28	.009*
DP × Az × PO × NM	6, 96	.49	.01	.69	.03	.753

Note. ε = Greenhouse-Geisser (GG) epsilon, p = GG-corrected p-values. Note that GG correction is appropriate only for within-subject tests, with more than one degree of freedom in the numerator.
*$p < .05$.

distance between the sound sources increases, the results for amplitude normalization vary considerably more depending on distance pair and do not to follow a linear trend. Two nested ANOVAs yielded a significant main effect of distance pair with loudness normalization [$F(3, 48) = 42.75$, $p < .001$, $\eta_p^2 = .73$, $\varepsilon = .80$] and with amplitude

normalization [$F(3, 48) = 79.19$, $p < .001$, $\eta_p^2 = .83$, $\varepsilon = .56$]. Thus, intensity cues remain with both normalization methods, but it seems that the offset of the loudness normalization method could be compensated much more easily with a linearly decreasing gain function.

Finally, we examined the influence of azimuth in more detail. The same two nested ANOVAs as described in the previous paragraph showed a significant main effect of azimuth with loudness normalization [$F(2, 32) = 34.23$, $p < .001$, $\eta_p^2 = .68$, $\varepsilon = .81$] and with amplitude normalization [$F(2, 32) = 10.20$, $p < .001$, $\eta_p^2 = .39$, $\varepsilon = .85$]. Eight paired t tests comparing the loudness-normalized conditions with $\varphi = 270°$ against conditions with $\varphi = 30°$ and $\varphi = 150°$ showed that the stimuli with $\varphi = 270°$ led to significantly higher ratings in perceived loudness differences ($p < .01$ for all), i.e., participants perceived the source at $d = 0.25$ m and $\varphi = 270°$ as the least loud.

The results reveal that the loudness normalization attenuates very close sources too much. Thus, sources at $d = 0.25$ m (especially for $\varphi = 270°$) were always quieter than sources farther away, which was particularly perceptible in the direct comparison task. In contrast, the amplitude normalization amplifies very close sources too much and at the same time attenuates the more distant sources. As a result, sources at $d = 0.25$ m were mostly distinctly louder than sources farther away, especially when compared to sources at $d = 1.50$ m. Thus, the amplitude normalization actually results in intensity cues associated with a natural distance shift, and therefore allows for correct distance discrimination based on the intensity cues that it is supposed to remove. Overall, it seems that very close (lateral) sound sources are most critical and that both normalization methods have strong drawbacks.

4.3 Discussion

Experiment 3 revealed that neither the commonly employed amplitude normalization nor the supposedly more suitable loudness normalization correctly removed perceptible loudness differences. Especially for the (lateral) closest source at $d = 0.25$ m, both methods achieved the worst results. Based on these surprising results, we can now explain the counterintuitive results of Experiment 2 as well as some observations of Experiment 1, and can also give a possible explanation for some contradictory findings in the literature.

In Experiment 2, it appears that the inaccurate loudness normalization resulted in intensity cues that participants exploited for distance discrimination. As the reference source at $d = 0.25$ m was always perceived as quieter than any of the other sources, it is not surprising that the participants mostly rated the second (louder) source as closer. Moreover, as shown in Experiment 3, the perceived loudness differences increased as a function of distance pair. This is in line with the findings from Experiment 2 where participants perceived clearer differences between the stimuli with increased nominal distance between the virtual sound sources (main effect of distance pair). Spectral cues, however, most probably played only a minor role or were simply masked by intensity cues. These intensity cues strongly affected the results of the sensitive direct-comparison task employed in Experiments 2 and 3, but apparently played less of a role in the multiple-stimulus comparison procedure in Experiment 1. However, in subset HTNorm of Experiment 1, the normalized source at $d = 0.25$ m and $\varphi = 270°$ was rated further away than the other sources at $d = 0.25$ m. This is in line with the results of Experiment 3, which revealed that this stimulus was perceived as the least loud among all stimuli. Furthermore, the ratings for the loudness normalized stimuli in Experiment 3 follow the same trend (v-shaped pattern) as the ratings for subset HTNorm and $d = 0.25$ m and $d = 0.50$ m in Experiment 1, indicating that participants estimated relative distance in these conditions of Experiment 1 according to the perceived azimuth-dependent loudness differences between the sources. Relatedly, in subset STNorm, stimuli that were nominally farther away were consistently rated as closer, likely because they were perceived as louder. Thus, remaining intensity cues caused by the (erroneous) loudness normalization, as revealed in Experiment 3, are a reasonable explanation for these on first sight surprising findings of Experiment 1, but due to the azimuth effects, a binaural influence on distance estimation cannot be generally ruled out.

5 General discussion

Previous studies have yielded conflicting results regarding the question whether (individual or non-individual) binaural cues contribute to distance perception in the near field. To address this open research question, we conducted three listening experiments using non-individual binaural synthesis. Experiment 1 was designed as a broader study to get a better insight into various potential influences on auditory distance perception. In a multi-stimulus comparison task, subjects had to estimate distance of loudness-normalized and non-normalized nearby sources in static and dynamic binaural synthesis. To isolate binaural cues in the (supposedly) best possible way, we normalized the stimuli in loudness according to ITU-RBS.1770-4. Experiment 2 strictly focused on binaural cues of nearby sound sources and their potential influence on auditory distance perception. Here, subjects had to judge the relative perceived distance between loudness-normalized sources in dynamic binaural rendering. Finally, Experiment 3 assessed the performance of the employed loudness normalization and of the frequently used amplitude normalization proposed by Brungart et al. [7].

The results of Experiment 1 suggest that in most examined conditions, naive listeners did not make use of non-individual binaural cues for distance estimation of nearby loudness-normalized sound sources in anechoic conditions, despite the drastic physical changes in binaural cues (especially in ILDs) due to changes in nominal sound source distance or head movements. In Experiment 2, participants even performed significantly below chance, that is, they mostly interpreted the closer source as the

source farther away. This surprising result was explained by Experiment 3, which revealed that the employed loudness normalization overcorrected so that closer sources were perceived as less loud than farther sources. As a result, the participants in Experiment 2 always compared a slightly quieter source to a somewhat louder source and therefore most probably discriminated distance based on intensity cues, which provided clearly perceptible differences in the sensitive direct-comparison test of Experiment 2. In the multiple-stimulus comparison task of Experiment 1, the effects of remaining intensity cues in conditions with normalized stimuli were weaker, but in line with those observed in Experiments 2 and 3.

Experiment 3 also revealed that previous studies on the effect of binaural cues on distance estimation were likely compromised by the opposite drawback of amplitude normalization. In particular, with amplitude normalization, close sources are still perceived louder than far sources. In other words, the here considered amplitude normalization did not fulfill its intended function. As a consequence, in previous studies employing amplitude normalization, participants might have been able to correctly perceive distance changes based on intensity cues instead of binaural cues. Thus, the present test series clearly demonstrates the problem of normalization as a means to remove intensity cues: with an imperfect normalization, intensity cues remain, which then dominate distance estimation and mask all other cues. Regarding non-individual binaural cues, our results show no clear evidence that, despite their strength in the near field, they contribute to distance estimation of nearby sound sources in anechoic conditions when weak residual intensity cues are still present. However, given the demonstrated drawbacks of the normalization methods causing these residual intensity cues, further studies with other test and normalization methods are needed to clarify the role of binaural cues for distance estimation of nearby sound sources.

Results of our Experiment 1 are in line with Shinn-Cunningham [18] and Kopčo and Shinn-Cunningham [19], who concluded that individual binaural cues are irrelevant for distance perception of nearby sound sources in anechoic conditions [18] and furthermore could not find direct evidence that binaural cues affect distance judgments in reverberant conditions [19]. In a follow-up study using non-individual BRIRs, Kopčo et al. [20] qualified the latter statement by suggesting that the DRR cue is more robust and reliable than the ILD cue, but that the brain actually combines both cues to process distance estimation and does not simply rely on a DRR-to-distance mapping.

In contrast, Brungart et al. [7] as well as Kan et al. [13] and Spagnol et al. [8] for example found that individual as well as non-individual binaural cues affect distance estimation especially for lateral sound sources. In fact, in all these studies, the amplitude normalization proposed by Brungart et al. [7] was applied, which leads to very close sources being presented too loudly according to the results of Experiment 3. Thus, amplitude-normalized stimuli are similar to a natural presentation of sound sources at different distances where closer sources are always (a little) louder. Furthermore, our experiment showed that the perceived loudness differences were particular strong for the lateral sound source ($\varphi = 270°$). Given that intensity is considered as the most dominant cue for distance perception and that even small intensity differences can lead to a change in perceived distance [1, 21], it might be that participants of the above mentioned studies exploited subtle intensity differences between the stimuli for correct distance estimation instead of binaural cues. However, since Brungart et al. [7, 11, 12] roved the level of the normalized stimuli, it is unlikely that participants were able to exploit intensity cues in their studies. Nevertheless, already small intensity differences between the stimuli might be important in localization experiments (without level-roving) as conducted by Kan et al. [13], and especially in a direct-comparison test (relative distance estimation) as for example performed by Spagnol et al. [8], these differences most certainly affect distance estimation. Thus, based on Experiment 3, some results of the above mentioned studies might also be explained by residual intensity cues. However, as previous studies used other stimuli, had other conditions, and applied other individual or generic HRTFs or even used loudspeakers instead of virtual acoustics, the residual-intensity-cue explanation of their results must await further dedicated studies.

Theoretically, roving the level of the stimuli and thus diminishing remaining intensity cues might be a way to compensate for the drawbacks of normalization, especially in studies on absolute distance perception. However, in practice, roving the level seems not expedient for experiments on relative distance estimation applying direct-comparison tasks and normalization. In particular, level-roving would reintroduce intensity cues and thus negate the attempt of the normalization method to eliminate intensity differences between the stimuli. As a result, intensity would most certainly mask any other cue, and listeners would therefore estimate distance purely based on variation in stimulus intensity induced by the roving, i.e., they would most probably perceive the louder source as closer than the quieter source (as confirmed by results of Experiments 2 and 3). Especially naive listeners, such as all participants in the presented listening experiments, seem to be affected by this since they mostly cannot ignore the strong intensity distance cues induced by roving, even if they are explicitly instructed to do so (for direct evidence from a recent pertinent study, see [23]).

Our findings apply to situation where non-individual HRTFs are used. It appears possible that the use of individual HRTFs would affect the results. As discussed in the Introduction (see Sect. 1), it is not clear how exactly and under what circumstances individual HRTFs improve distance perception as compared to non-individual HRTFs. In the special case of distance estimation without DRR cues (anechoic) and without or only weak intensity cues, as tested in the present study, listeners might weight spectral cues more strongly than in natural acoustic conditions. Thus, whereas non-individual spectral cues do not seem to affect distance perception in more realistic listening situations [31], distance perception in our experiments under

anechoic conditions was maybe affected by impaired monaural spectral cues caused by the non-individual HRTFs, similar as for example shown by Baumgartner et al. [36]. Therefore, listeners may have perceived the sources closer (or to some extent less externalized) than would have been the case when using individual HRTFs. However, according to subject reports, listeners perceived sound sources as sufficiently externalized in our experiments. Ultimately, whether individual binaural and spectral cues would have an effect on distance estimation of normalized stimuli is another empirical question which remains to be investigated in future studies.

6 Conclusion

In the present study, we examined how non-individual binaural cues contribute to auditory distance estimation of nearby sound sources. We conducted three experiments in virtual acoustics with naive and untrained listeners. In a multiple-stimulus comparison task (Experiment 1), non-individual binaural cues did not evidently influence distance estimation of nearby sound sources. In a more sensitive direct-comparison task (Experiment 2), listeners might have judged distance based on remaining intensity differences, even with loudness normalization applied. The final experiment (Experiment 3) showed that the loudness normalization applied here, as well as the amplitude normalization introduced by Brungart et al. [7], leave intensity cues that might mask any subtle binaural distance cues. In sum, the present set of experiments has revealed that eliminating all intensity cues without overcorrecting is not a trivial task and that the normalization method should be carefully considered and evaluated when designing a distance perception experiment.

This also means that the influence of binaural cues cannot be correctly investigated with test methods based on stimulus normalization that do not employ additional approaches to effectively suppress remaining intensity cues. Therefore, it remains an open question whether binaural cues contribute to distance estimation of nearby sound sources, and several previous studies cannot be taken as conclusive evidence for or against the role of binaural cues. In fact, our results indicate that several conflicting findings in the literature regarding the role of binaural cues in distance estimation can be explained by differences in normalization methods across the various studies, i.e., maybe subjects only evaluated distance based on remaining salient intensity cues, which masked subtle binaural cues.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF 03FH014IX5-NarDasS). We would like to thank the participants of the experiments for their patience and commitment. We also thank Tim Lübeck for his assistance in conducting the experiments. The research data for this article are available at https://doi.org/10.5281/zenodo.4445283.

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, Shahina 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.
3. J. Blauert: Spatial Hearing – The Psychophysics of Human Sound Localization. MIT Press, Cambridge, MA, 1996.
4. D.S. Brungart, W.M. Rabinowitz: Auditory localization of nearby sources. Head-related transfer functions. Journal of the Acoustical Society of America 106, 3 (1999) 1465–1479.
5. R.O. Duda, W.L. Martens: Range dependence of the response of a spherical head model. Journal of the Acoustical Society of America 104, 5 (1998) 3048–3058.
6. J.M. Arend, A. Neidhardt, C. Pörschmann: Measurement and perceptual evaluation of a spherical near-field HRTF Set, in Proceedings of the 29th Tonmeistertagung – VDT International Convention, 2016, pp. 356–363.
7. D.S. Brungart, N.I. Durlach, W.M. Rabinowitz: Auditory localization of nearby sources. II. Localization of a broadband source. Journal of the Acoustical Society of America 106, 4 (1999) 1956–1968.
8. 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.
9. R.E. Holt, W.R. Thurlow: Subject orientation and judgment of distance of a sound source. Journal of the Acoustical Society of America 46, 6B (1969) 1584.
10. M.B. Gardner: Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. Journal of the Acoustical Society of America 45, 1 (1969) 47–53.
11. D.S. Brungart: Auditory localization of nearby sources. III. Stimulus effects. Journal of the Acoustical Society of America 106, 6 (1999) 3589–3602.
12. D.S. Brungart, B.D. Simpson: Auditory localization of nearby sources in a virtual audio display, in: Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, 2001, pp. 107–110.
13. A. Kan, C. Jin, A. Schaik: A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function. Journal of the Acoustical Society of America 125, 4 (2009) 2233–2242.
14. W.E. Simpson, L.D. Stanton: Head movement does not facilitate perception of the distance of a source of sound. Journal of the Acoustical Society of America 86, 1 (1973,) 151–159.
15. L.D. Rosenblum, A. Paige Wuestefeld, K.L. Anderson: Auditory reachability: an affordance approach to the perception of sound source distance. Ecological Psychology 8, 1 (1996) 1–24.
16. B.G. Shinn-Cunningham, S. Santarelli, N. Kopčo: Distance perception of nearby sources in reverberant and anechoic listening conditions: Binaural vs. Monaural Cues, in Poster presented at the 23rd MidWinter meeting of the Association for Research in Otolaryngology, St. Petersburg, Florida, 2000.

17. B.G. Shinn-Cunningham: Distance cues for virtual auditory space, in Proceedings of the First IEEE Pacific-Rim Conference on Multimedia, Sydney, Australia, 2000, pp. 227–230.
18. B.G. Shinn-Cunningham: Localizing sound in rooms, in Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments, Snowbird, Utah, 2001, pp. 17–22.
19. N. Kopčo, B.G. Shinn-Cunningham: Effect of stimulus spectrum on distance perception for nearby sources. Journal of the Acoustical Society of America 130, 3 (2011) 1530–1541.
20. 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.
21. D.H. Ashmead, D. Leroy, R.D. Odom: Perception of the relative distances of nearby sound sources. Perception & Psychophysics 47, 4 (1990) 326–331.
22. G.A. Miller: Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. Journal of the Acoustical Society of America 19, 4 (1947) 609–619.
23. L. Prud'homme, M. Lavandier: Do we need two ears to perceive the distance of a virtual frontal sound source? Journal of the Acoustical Society of America 148, 3 (2020) 1614–1623.
24. ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level. International Telecommunications Union, Geneva, 2015.
25. T. Djelani, C. Pörschmann, J. Sahrhage, J. Blauert: An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization. Acta Acustica united with Acustica 86, 6 (2000) 1046–1053.
26. E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman: Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America 94, 1 (1993) 111–123.
27. H. Møller, M.F. Sørensen, C.B. Jensen, D. Hammershøi: Binaural technique: do we need individual recordings? Journal of the Audio Engineering Society 44, 6 (1996) 451–469.
28. D.R. Begault, E.M. Wenzel, M.R. Anderson: Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. Journal of the Audio Engineering Society 49, 10 (2001) 904–916.
29. P. Zahorik: Distance localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America 108 (2000) 2597.
30. P. Zahorik: Auditory display of sound source distance, in Proceedings of the International Conference on Auditory Displays, 2002, pp. 1–7.
31. V. Best, R. Baumgartner, M. Lavandier, P. Majdak, N. Kop: Sound externalization: A review of recent research. Trends in Hearing 24 (2020) 1–14.
32. G. Yu, L. Wang: Effect of individualized head-related transfer functions on distance perception in virtual reproduction for a nearby source, in Proceedings of the AES International Conference on Spatial Reproduction – Aesthetics and Science, 2018, pp. 1–5.
33. G. Yu, L. Wang: Effect of individualized head-related transfer functions on distance perception in virtual reproduction for a nearby sound source. Archives of Acoustics 44, 2 (2019) 251–258.
34. W.M. Hartmann, A. Wittenberg: On the externalization of sound images. Journal of the Acoustical Society of America 99, 6 (1996) 3678–3688.
35. W. Owenbrimijoin, A.W. Boyd, M.A. Akeroyd: The contribution of head movement to the externalization and internalization of sounds. PLoS One 8, 12 (2013) 1–12.
36. R. Baumgartner, D.K. Reed, B. Tóth, V. Best, P. Majdak, H. Steven Colburn, B. Shinn-Cunningham: Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. Proceedings of the National Academy of Sciences of the United States of America 114, 36 (2017) 9743–9748.
37. A.V. Giner: Scale – conducting psychoacoustic experiments with dynamic binaural synthesis, in Proceedings of the 41st DAGA, 2015, pp. 1128–1130.
38. M. Geier, J. Ahrens, S. Spors: The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods, in Proceedings of the 124th AES Convention, Amsterdam, The Netherlands, 2008, pp. 1–6.
39. C. Pörschmann, J.M. Arend, A. Neidhardt, A spherical near-field HRTF Set for auralization and psychoacoustic research, Proceedings of the 142nd AES Convention, Berlin, Germany, 2017, pp. 1–5.
40. EBU R128: Loudness normalisation and permitted maximum level of audio signals. EBU – European Broadcasting Union, Geneva, 2014.
41. B. Bernschütz: Microphone arrays and sound field decomposition for dynamic binaural recording, Doctoral dissertation, TU Berlin, 2016.
42. C. Pörschmann, C. Störig: Investigations into the velocity and distance perception of moving sound sources. Acta Acustica united with Acustica 95. 4 (2009,) 696–706.
43. A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, S. Weinzierl, A spatial audio quality inventory (SAQI), Acta Acustica united with Acustica 100, 5 (2014) 984–994.
44. Y. Hochberg: A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 4 (1988) 800–802.
45. G.V. Glass, P.D. Peckham, J.R. Sanders: Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research 42, 3 (1972) 237–288.
46. S.W. Greenhouse, S. Geisser: On methods in the analysis of profile data. Psychometrika 24, 2 (1959) 885–891.
47. E.-J. Wagenmakers: A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review 5 (2007) 779–804.
48. J.N. Rouder, P.L. Speckman, D. Sun, R.D. Morey, G. Iverson: Bayesian t tests for accepting and rejecting the null hypothesis. Psychonomic Bulletin & Review 16, 2 (2009) 225–237.
49. J.N. Rouder, R.D. Morey, P.L. Speckman, J.M. Province, Default Bayes factors for ANOVA designs, Journal of Mathematical Psychology 56, 5 (2012) 356–374.
50. G.R. Loftus, M.E.J. Masson: Using confidence intervals in within-subject designs, Psychonomic Bulletin & Review 1, 4 (1994) 476–490.
51. J. Jarmasz, J.G. Hollands: Confidence intervals in Repeated-Measures Designs: The number of observations principle. Canadian Journal of Experimental Psychology 63, 2 (2009) 124–138.
52. R.R. Sokal, F. James Rohlf: Introduction to Biostatistics, 2nd ed. Dover Publications Inc, Mineola, NY, 2009.
53. R.A. Butler, E.T. Levy, W.D. Neff: Apparent distance of sounds recorded in echoic and anechoic chambers. Journal of Experimental Psychology: Human Perception and Performance 6, 4 (1980) 745–750.
54. A.D. Little, D.H. Mershon, P.H. Cox: Spectral content as a cue to perceived auditory distance. Perception 21, 3 (1992) 405–416.
55. P.D. Coleman: Dual role of frequency spectrum in determination of auditory distance. Journal of the Acoustical Society of America 44, 2 (1968) 631–632.

Appendix

A.1 Overview of discussed studies

Table A.1. Overview of studies investigating the contribution of binaural cues to distance estimation of (nearby) sound sources. (+) Binaural cues contribute. (°) Unclear or mixed findings. (−) Binaural cues do not contribute.

Study	Method	Normalization	Findings and conclusion
Holt and Thurlow [9]	Anechoic conditions. Far-field sources between 1.80 m and 19 m. Participants judged distance in feet.	Level [dB(A)]	(+) Performance improved for lateral sources. Binaural cues are important for distance perception.
Brungart et al. [7]	Anechoic conditions. Near-field sources at distances between 0.15 m and 1.00 m. Participants pointed to the perceived location.	Distance-related amplitude normalization and level-roving	(+) Most accurate distance estimation for lateral sources. ILDs are salient cues for distance estimation.
Brungart and Simpson [12]	Static binaural synthesis with near-field KEMAR HRTFs. Near-field sources at distances between 0.12 m and 1.00 m. Participants pointed to the perceived location.	Signal power and level-roving	(+) Performance worse than in Brungart et al. [7], maybe due to non-individual HRTFs. Still proper distance estimation for lateral sources. ILDs are salient cues for distance estimation.
Gardner [10]	Anechoic conditions. Far-field sources at distances between 0.90 m and 9.00 m. Participants judged distance by choosing a loudspeaker.	Level [dB(B)]	(°) Bad performance for frontal sources. Small head movements led to better performance. Changes in binaural cues might be beneficial.
Kan et al. [13]	Static binaural synthesis with synthesized near-field HRTFs based on individual far-field HRTFs. Near-field sources at distances between 0.10 m and 1.00 m. Participants pointed to the perceived location.	Same as Brungart et al. [7], but without level-roving	(°) Poor performance. Minor distance discrimination for lateral sources at distances < 0.20 m. ILDs are no powerful cues.
Kopčo et al. [20]	Static binaural synthesis with non-individualized near-field BRIRs. Near-field sources at distances between 0.15 m and 1.00 m. 2AFC test – Participants indicated whether the second source was closer or farther than the first one.	Near-ear level [dB (SPL)] and level-roving	(°) Distance estimation based on DRR and ILD cue combination, but DRR cues are more dominant and reliable.
Spagnol et al. [8]	Static binaural synthesis with synthesized near-field HRTFs based on KEMAR far-field HRTFs. Near-field sources at distances between 0.20 m and 1.00 m. 2AFC test – Participants indicated whether the second source was closer or farther than the first one.	Same as Brungart et al. [7], but without level-roving	(°) Poor performance. Similar to Kan et al. [13], slightly improved performance for lateral sources at distances < 0.20 m. ILDs are no powerful cues.
Simpson and Stanton [14]	Quasi-anechoic conditions. Near- and far-field sources at distances between 0.30 m and 2.66 m. Participants rated perceived distance on a scale.	None	(−) No influence of head movements on distance estimation. Binaural cues are not important for distance perception.
Rosenblum et al. [15]	Acoustically normal room. Near-field sources at distances between 0.38 m and 1.10 m. Participants judged the source reachability.	None	(−) No influence of head movements on distance judgment accuracy. Binaural cues are not important for distance judgment.
Shinn-Cunningham et al. [16]	Static binaural synthesis with individual near-field HRTFs/BRIRs. Near-field sources at distances between 0.15 m and 1.00 m. Participants judged distance with a GUI.	No sufficient information	(−) Poor performance. ILD cues do not contribute to distance perception in reverberant conditions and do not provide robust distance percepts even in anechoic conditions.
Kopčo and Shinn-Cunningham [19]	Static binaural synthesis with individual near-field BRIRs. Far- and near-field sources at distances between 0.15 m and 1.70 m. Participants judged distance with a GUI.	Near-ear level [dB (SPL)] and level-roving	(−) Performance was better for lateral sources than for frontal sources and worse without low-frequency energy. In reverberant conditions, only DRR cues are used to judge distance, and not the ILD cues.

Cite this article as:Arend JM, Liesefeld HR & Pörschmann C. 2021. 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.