JASA ARTICLE

Localization accuracy of phantom sound sources on the horizontal plane by bilateral hearing aid users in aided free-field and non–free-field conditions

Janani Fernandez,[1,2,3,a)] Petteri Hyvärinen,[1,4] and Abigail Anne Kressner[2,3]
[1]*Department of Information and Communications Engineering, Aalto University, 02150 Espoo, Finland*
[2]*Department of Health Technology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark*
[3]*Copenhagen Hearing and Balance Centre, Rigshospitalet, 2100 Copenhagen, Denmark*
[4]*Department of Otorhinolaryngology, Kuopio University Hospital, 70210 Kuopio, Finland*

ABSTRACT:
This study investigates the use of amplitude panning in a localization accuracy test and the influence of a non-ideal environment on its feasibility as a clinical tool. The horizontal localization accuracy of 16 normal-hearing participants and ten bilateral hearing aid users was assessed for real and amplitude panned sound sources produced over loudspeakers. Localization accuracy was measured with speech-shaped noise in both an anechoic chamber (free-field) and an acoustically treated listening room (non–free-field). The root mean square error between the response angle and the target angle was calculated for each participant. Thus, the root mean square error for the two sound source types for each test environment could be calculated and compared, and also contrasted against existing literature. Statistical analysis of the control group results revealed an effect of the target angle, method used (real vs amplitude panning) and environment (free-field vs non–free-field). An interaction between target angle and environment was also found. For the hearing aid user group, however, only an effect of target angle was found, which may lend support to simpler setups with fewer loudspeakers in non–free-field environments. However, the effect of the room varied between individuals within this group, thereby warranting further exploration.
 https://doi.org/10.1121/10.0035828

(Received 18 July 2024; revised 16 January 2025; accepted 27 January 2025; published online 13 February 2025)

[Editor: Pavel Zahorik] Pages: 1151–1161

I. INTRODUCTION

Research and development for assistive hearing devices have historically focused on restoring language comprehension to the device users. Sound source localization accuracy for hearing device users, on the other hand, remains relatively less explored but may also be considered as an important facet of hearing. This is because the ability to localize sound sources may allow a person to react more appropriately to potential dangers, such as nearby traffic, and may also improve speech intelligibility through the spatial release from masking phenomenon (Arbogast *et al.*, 2005; Noble *et al.*, 1997). Therefore, being able to evaluate the source localization abilities of hearing device users within clinical settings, during their device fittings, may allow the devices to be better optimized for this source localization task, thus ultimately improving the wearer's quality of life.

The ideal way to assess localization abilities is to place a listener in a free-field environment and to play an audio signal out of a set of loudspeakers located in different directions (Blauert, 1997). The localization error is then typically defined as the angle between the true direction and the direction reported by the listener. Naturally, the greater the number of loudspeakers present in the setup, the higher the spatial resolution is for the test. Such free-field conditions that afford high spatial resolution via a high density of loudspeakers are, however, not always available. Thus, many previous localization studies, particularly those involving hearing device users, have been conducted in sound-treated booths and listening rooms (Brungart *et al.*, 2017; Dorman *et al.*, 2016; Drennan *et al.*, 2005; Ellis and Souza, 2020; Johnson *et al.*, 2017; Van den Bogaert *et al.*, 2006). In such environments, however, sound source reflections may lead to localization shifts due to the summing localization phenomenon (Blauert, 1997). This effect has mostly been explored within the context of normal-hearing (NH) listeners, whereby if a coherent sound arrives within approximately 2 ms of the initial sound, the two sound events are merged into one phantom image, and localized somewhere in-between the two directions (Blauert, 1997). Summing localization has also been shown to affect localization in hearing-impaired persons, as evidenced by the successful application of sound reproduction methods that rely on the summing localization phenomenon (Fernandez *et al.*, 2022). On the other hand, hearing device users may experience a reduced precedence effect (Akeroyd and Guy, 2011;

[a)]Email: janani.fernandez@aalto.fi

Goverts *et al.*, 2002), which is a phenomenon that helps to mitigate the influence of sounds arriving after 2 ms on sound localization and therefore influences sound source localization in non–free-field, reverberant conditions.

An important practical consideration for listening rooms in clinics is that they are typically relatively small in size; thus, they usually feature fewer loudspeakers in their setups. While the use of fewer loudspeakers lowers costs, it also reduces the possible spatial resolution of localization tests conducted in the space. This resolution may be improved, i.e., made finer, by using movable mountings, but such setups are usually quite expensive and may also be too large for the given space. Therefore, to address the problem of low spatial resolution, there has been growing interest in using spatialization algorithms to synthesise phantom sound sources over headphones or loudspeakers, in order to evaluate and potentially improve the localization abilities of hearing-impaired listeners (Besing and Koehnke, 1995; Ellis and Souza, 2020; Fernandez *et al.*, 2022; Mueller *et al.*, 2012; Nisha *et al.*, 2023; Simon *et al.*, 2017; Smith-Olinde *et al.*, 1998; Syeda *et al.*, 2023; Zenke, 2021). These sound source spatialization algorithms are principally tasked with producing signals that lead to the appropriate perceptual cues being delivered to the listener when they are played over the reproduction setup, such that a sound source is perceived as emanating from a specific target direction. These perceptual cues include interaural time differences, which are incurred due to the source's angle of incidence, interaural level differences, which largely arise due to head-shadowing, and monaural cues caused by the acoustical interactions with the pinnae (Blauert, 1997). These perceptual cues may also be collectively described via the head-related transfer function (HRTF) (Blauert, 1997).

To produce phantom sound sources over headphones, one may directly convolve a monophonic input signal with an HRTF corresponding to the desired direction. Generic HRTFs, often measured using dummy heads (Armstrong *et al.*, 2018; Engel *et al.*, 2023), may be utilized for this task and have been deployed in previous studies involving hearing device users (Mueller *et al.*, 2012; Smith-Olinde *et al.*, 1998; Syeda *et al.*, 2023). However, since the listener's head geometry and anthropomorphic features are highly individual, headphone-based spatialization can benefit from the use of personalized HRTFs, which tailor the spatialization to each listener (Brungart *et al.*, 2017; Møller *et al.*, 1996; Wenzel *et al.*, 1993; Wightman and Kistler, 1989). There is also evidence to support the use of personalized HRTFs over generalized HRTFs for hearing device users (Brungart *et al.*, 2017). However, such personalized HRTFs are typically costly and time-consuming to acquire for a dense grid of directions, since this requires the listener to be placed in a free-field environment with microphones in their ear canals and then remain motionless as the HRTF measurements are made for different directions. Furthermore, if one wishes to also relay dynamic spatial cues to the listener by incorporating head-tracking into the system to follow listener head-movements, then a very dense HRTF measurement grid may be required. Therefore, personalized binaural listening using measured HRTFs is generally only conducted on a small scale within academic contexts. While recent efforts have explored simulating HRTFs (Brinkmann *et al.*, 2019; Ziegelwanger *et al.*, 2015), this requires a detailed head and pinna scan, and the required computations for the full audible frequency range may be prohibitive to conduct on a large scale.

To produce phantom sources over loudspeakers instead, one may employ amplitude panning (Pulkki, 1997). Unlike HRTF convolution-based spatialization, the listener can instead experience the appropriate localization cues due to the real physical pathway between a loudspeaker and the listener's two ears, which removes the need for an individualized HRTF measurement setup, and also allows the assessment of the user's localization ability when wearing their actual devices, i.e., the hardware and software they are acclimatized to using in everyday life. Such an approach to generate phantom sources, along with other similar approaches, have been previously explored with several hearing-impaired and hearing device user populations through perceptual evaluations (Best *et al.*, 2015; Ellis and Souza, 2020; Fernandez *et al.*, 2024; Fernandez et al., 2022; Koski *et al.*, 2013; Mansour *et al.*, 2021). There have also been investigations into the use of phantom sound sources for testing the hearing devices themselves via objective metrics (Cubick and Dau, 2016; Grimm *et al.*, 2015; Koski *et al.*, 2014; Simon *et al.*, 2017). Some of these studies have been conducted in free-field spaces (Best *et al.*, 2015; Cubick and Dau, 2016; Koski *et al.*, 2014; Koski *et al.*, 2013; Mansour *et al.*, 2021) and some in non–free-field listening rooms (Ellis and Souza, 2020; Fernandez *et al.*, 2024; Fernandez *et al.*, 2022; Koski *et al.*, 2013; Simon *et al.*, 2017). The majority of these studies though have focused on examining speech intelligibility in sound field environments produced via the use of virtual audio rendering techniques, primarily concluding that the speech reception thresholds obtained are reliable (Best *et al.*, 2015) and result in thresholds similar to those obtained in the corresponding real sound field environment that the methods intend to mimic (Koski *et al.*, 2013). However, it is unclear whether these conclusions apply to localization thresholds, i.e., if virtual audio rendering techniques are also reliable when used in a localization test, as distortions to spatial cues may be more apparent in a localization task than in a speech intelligibility task. Of the aforementioned studies, only two have specifically investigated the use of these methods for localization tasks within listening rooms over loudspeakers. While both studies concluded that they are indeed viable methods for the task at hand, the studies were conducted in only non-ideal conditions, meaning no direct reference comparison was made to free-field conditions (Ellis and Souza, 2020; Fernandez *et al.*, 2022), which makes it difficult to independently assess the influence of the specific room on the outcome.

Therefore, the main contribution of this study is the direct comparison of bilateral hearing aid (BiHA) users' localization performance for phantom and real sound sources presented over loudspeakers in both a free-field environment and a sound-treated listening room. Localization errors were measured for ten BiHA users in both free-field and non–free-field environments with real and phantom sources and directly compared to assess the effect of the listening environment and the type of sound source (i.e., real or phantom source). For reference, the same comparison is made with 16 NH listeners. This study design attempts to clarify whether there is an effect of the source type on localization accuracy for both NH listeners and BiHA users; to what extent the test environment (i.e., free-field or non–free-field) affects localization accuracy for both of these participant groups; and whether there is an interaction between these two aspects.

II. METHODOLOGY

A. Test environments

The effects of two test environments were explored in this study: a free-field and a non–free-field (but acoustically treated), listening room, environment. The first room, an anechoic chamber, was the Audio Visual Immersion Lab at the Technical University of Denmark (Lyngby, Denmark). The chamber is fitted with a spherical loudspeaker arrangement consisting of 64 loudspeakers, of which 24 were placed at 15° intervals on the horizontal plane, i.e., 0° elevation. From this ring, 15 loudspeakers were used in the listening testing, and they were positioned in a semicircle from −105° to 105°. Time and magnitude corrections were applied to individual loudspeakers based on impulse response measurements conducted from the listening position.

The non–free-field environment, on the other hand, is an acoustically treated listening room located at the Copenhagen Hearing and Balance Centre, Rigshospitalet (Copenhagen, Denmark). The walls of the room were acoustically treated to achieve a broadband reverberation time (RT_{60}) of 0.13 s, calculated according to the ISO 3382-1:2009 (2009) standard. The room contained a loudspeaker arrangement set at similar positions to the anechoic chamber, with the exception of the setup consisting of loudspeakers embedded within the acoustic paneling along the square walls of the room. As in the free-field environment, the 15 loudspeakers located between –105° and 105° were used in the listening test. In a similar manner to the adjustments made in the free-field environment, impulse response measurements taken at the listening position were used to apply time, magnitude, and level corrections to the individual loudspeakers.

B. Stimuli

The stimuli consisted of speech-shaped noise samples from the HARVARD speech corpus, which was generated using 52nd order linear predictive coding (Demonte, 2019). The sampling rate of the stimuli was 48 kHz. A total of four samples, each of length 2 s duration were isolated from the original sound file. Onsets and offsets of 10 ms each were applied to all stimuli.

The stimuli were created using the Reaper software (Cockos, San Francisco, CA) in conjunction with Virtual Studio Technology audio plugins. The SPARTA Panner (version 1.6.2), which implements the vector base amplitude panning (Pulkki, 1997) method, was used to create sound files, which on playback created phantom sound sources at the specified panning angle (McCormack and Politis, 2019). Additionally, the Institute of Electronic Music and Acoustics (Graz, Austria) plugin suite's distance compensator (version 1.12.0)[1] was utilized in the rendering chain for stimuli in the listening room environment to compensate for its rectangular, i.e., non-spherical, loudspeaker arrangement.

Real sound sources were created at loudspeaker locations every 15° between −90° and 90°. Phantom sound sources were also created at every 15° to correspond with the same positions as the real sound sources by utilizing the nearest loudspeakers on either side of the specified target angle, i.e., not utilizing the loudspeaker at that position if the specified panning angle coincided with a loudspeaker position. In addition to these locations, there were an additional six phantom sound sources positioned between loudspeakers so that the phantom sources were placed every 7.5° between –45° and 45°. Once again, the pair of loudspeakers nearest to the intended target angle were selected. Thus, a total of 32 source directions were considered: 13 for the real sound sources and 19 for the panned sources. Taking into account the four different sound files, a total of 128 stimuli were created for the tests. All stimulus files were normalized such that playback level was approximately 75 dB(A) at the listening position.

C. Participants

The BiHA user group consisted of ten bilateral hearing aid users. Their age ranged from 24 to 76 years, with a mean age of 54 years and a standard deviation of 16.6 years. Five members of the group were female. All participants in the group had symmetric hearing loss; i.e., the difference in the threshold values between the two ears did not exceed 15 dB at any frequency on their pure-tone audiograms. Table I contains anonymized information on the devices utilized by the test participants, which included models from Interon, Oticon, Rexton, Siemens, and Widex.

To form a control group, 16 participants with NH were recruited, i.e., persons with auditory threshold values of 25 dB HL or lower in the frequency range 125 Hz to 8 kHz. The age range of the group was 20 to 30 years, with a mean age of 25 years and standard deviation of 2.4 yrs. It should be noted that the two groups were not age-matched.

The audiograms for all test participants are plotted in Fig. 1 with gray lines marked with diamonds for members of the control group and squares for members of the BiHA user group. The average audiogram for each group is plotted with bold black lines.

TABLE I. Hearing device brand and model for the BiHA user group.

Participant ID	Brand	Model
1	A	1
2	A	2
3	A	3
4	A	4
5	A	5
6	B	6
7	B	6
8	C	7
9	D	8
10	E	9

D. Test design and procedure

In each test environment, the test participant was seated in the center of the loudspeaker setup facing the center loudspeaker, i.e., the loudspeaker positioned at azimuth 0°. The participants were tasked with determining the angle from which they heard the stimuli to be emanating. The test participants were asked not to move their heads and to face the center loudspeaker during playback. They recorded their answers via the use of a touchscreen graphical user interface (GUI) hosted on an iPad. The GUI consisted of a picture, in which the participant's position was marked with a gray circle and the positions of the loudspeakers were marked with red rectangles. The participant could then mark the direction they perceived the sound to be coming from by placing a circle in that direction relative to their position. A screenshot of the interface in which a participant's response is marked with a blue circle is presented in Fig. 2. This circle also denotes which loudspeaker the participant was facing during the test.

As can be seen in the screenshot, there were four buttons on the interface. Stimuli would only play after they selected the "Play Sound" button and were not repeated. Once the stimuli finished playing, the participant was able to mark their answer by selecting a position on the picture. If they made an error, it was possible for the participant to delete the circle and place a new one on the diagram or for the participant to select and drag the incorrectly placed circle to the intended position. Once the participant was satisfied with their answer, they were instructed to select the "Done" button, then select "Play Sound" once they were ready for the next trial. Each stimulus was repeated four times. A total of 128 trials were conducted.

Half the participants performed the listening tests in the free-field environment first, while the other half performed the test in the listening room environment first. The tests were completed in two sessions, one in each environment. The average duration of each session was 35 min. Note that the same test interface was utilized in both environments. Participants were informed that some stimuli will be coming from positions in between loudspeakers. Additionally, before starting the test, participants in the BiHA user group were asked to put their devices into their default "everyday" setting for listening. This was done in order to ensure that their localization performance would be based on the cues they are most acclimatized to in their day-to-day lives.

E. Data analysis

The root mean square error (RMSE) was calculated for each group, for each test environment and sound source type. This calculation, as well as the generation of plots, was performed using MATLAB (The MathWorks, Natick, MA). In addition, RMSE values were also calculated for each participant to be used for further statistical analysis, with one value computed for each source position across each test environment and sound source type, resulting in 52 such RMSE values for each participant. It should be noted this statistical analysis did not include the extra six phantom positions between –45° and 45°; i.e., statistical analysis was performed using only phantom positions at which a corresponding real sound source was positioned during the listening test. A mixed effect model RMSE $\sim$ angle $+$ environment $+$ source_type $+$ environment $\times$ source_type $+$ angle $\times$ environment $+$ angle $\times$ source_type $+ (1|\text{participant})$ was then constructed using the R (version 4.3.3) (R Core Team, 2024) package lme4 (Bates, 2014).

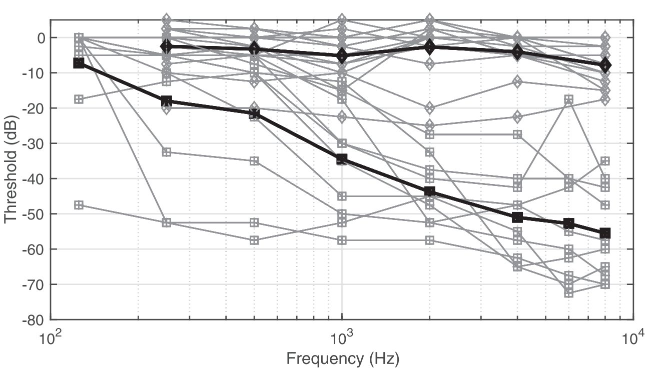

FIG. 1. The audiograms of the test participants. Each gray line represents the audiogram of a single participant, averaged over both ears. Square represents the audiograms of the BiHA user group. Diamonds represent the NH group. Black lines indicate the average audiogram of each group.

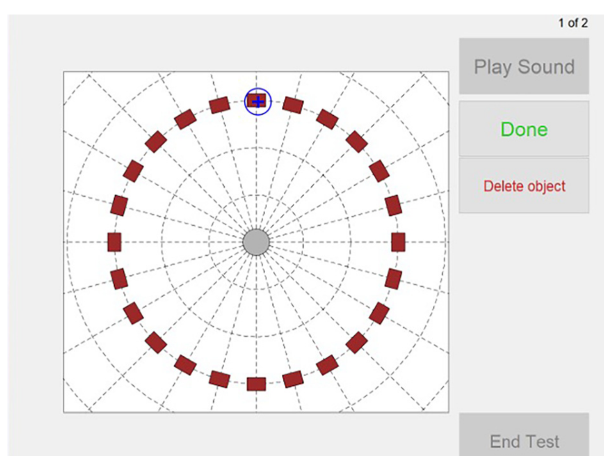

FIG. 2. Screenshot of the test interface.

JASA
https://doi.org/10.1121/10.0035828

Here, the fixed effects are modeled by three variables: environment is a categorical variable with two values corresponding to the two different test environments, source_type is an indicator variable with values corresponding to conditions with panning or no panning applied, and angle is a predictor variable indicating the (absolute) target angle. Finally, participant is a random intercept term, modeling the interindividual variability in baseline localization accuracy. For tests of model significance, Satterthwaite's approximation for the degrees of freedom was used, as implemented in the function anova in the package lmerTest (Kuznetsova *et al.*, 2017).

III. RESULTS

The response angles estimated by the participants were calculated based upon their placement of circles in the GUI. Figures 3 and 4 show the responses, whereby each subplot contains an individual's responses in a specific listening environment. The responses of a single individual are, therefore, plotted in two subplots, side by side: one for each listening environment. There are a total of 32 subplots for NH group and 20 for the BiHA user group. The target angle is shown on the *x* axis, while the *y* axis denotes the response angle. The trials for real sound sources are represented by circles colored light blue and light green, and the phantom

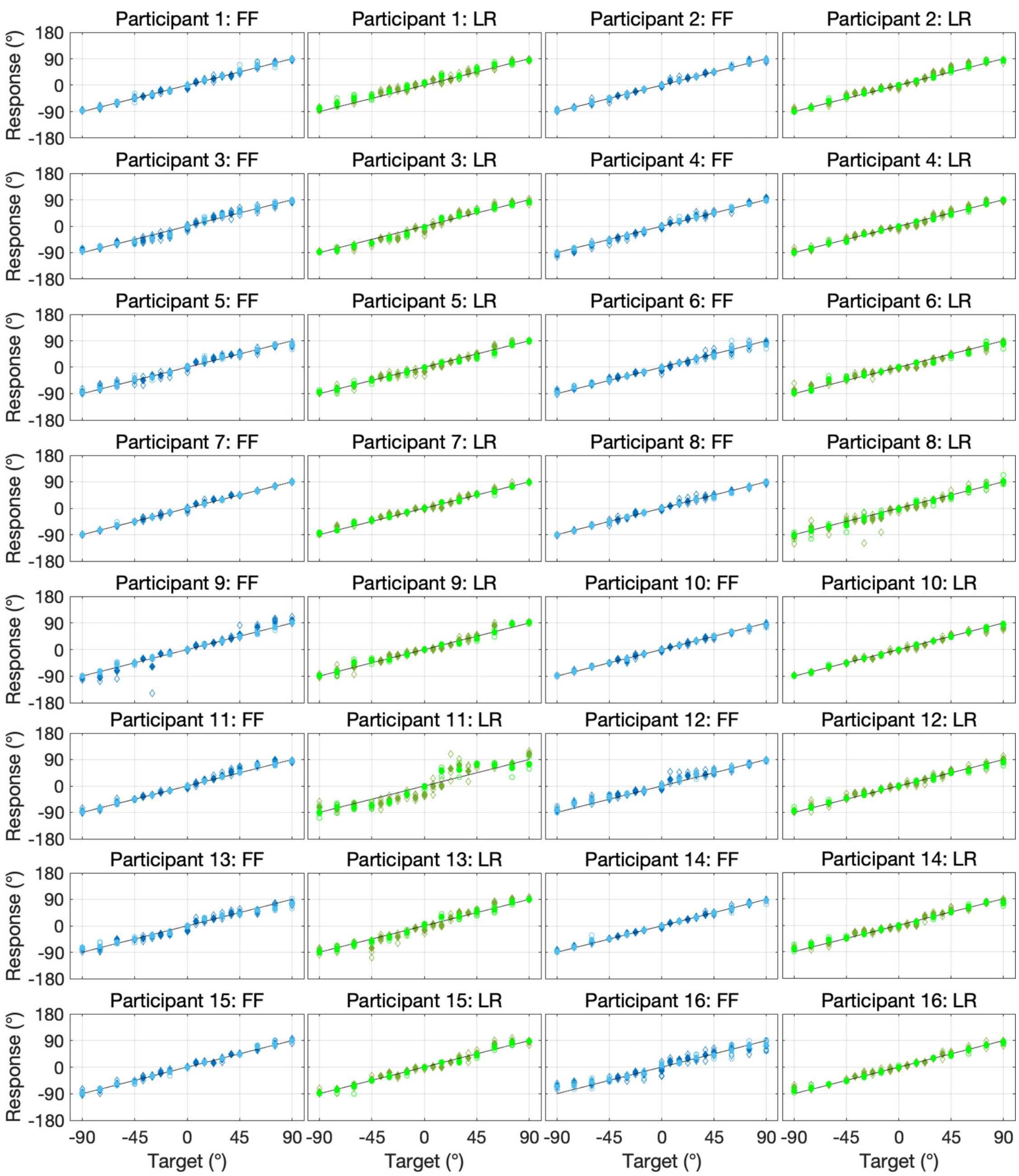

FIG. 3. NH results, separated by listening environment. Unfilled light blue and light green circles represent the real sound sources. Dark blue and dark green diamonds represent phantom sources. The mean response for each angle is indicated with filled circles and diamonds for real and phantom sources, respectively. Solid gray line indicates the perfect target angle and response angle relation. FF, free-field; LR, listening room.

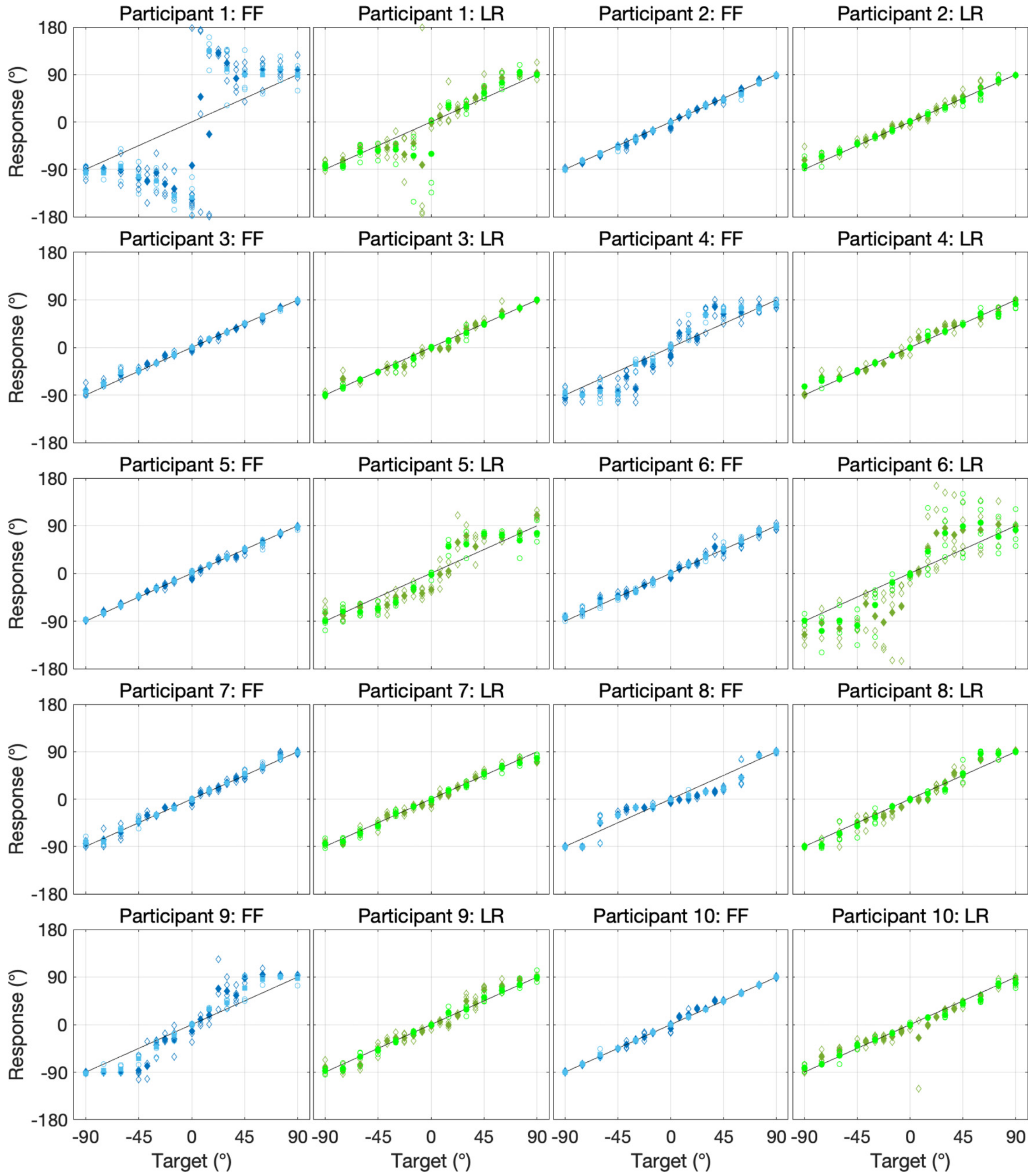

FIG. 4. BiHA user group results, separated by listening environment. Light blue and light green circles represent the real sound sources. Dark blue and dark green diamonds represent phantom sources. The mean response for each angle is indicated with filled circles and diamonds for real and phantom sources, respectively. Solid gray line indicates the perfect target angle and response angle relation. FF, free-field; LR, listening room.

source trials are represented by dark blue and dark green colored non-filled diamonds. The average response angle for each of the four trials is represented with a solid circle and a solid diamond for the real sound and phantom sound sources, respectively. Likewise, the individual trials are represented by unfilled circles or diamonds, depending upon the type of sound source.

Localization accuracy is high for all NH listeners, as shown by the fact that the responses lie along the diagonal in Fig. 3. This trend in accuracy is true for both real and phantom sources. Of note is that the response patterns are largely similar and consistent regardless of the environment, with very few exceptions, i.e., participant 11. In contrast, however, participant 4 shows that there is substantial variation among the HI listeners. A small number of participants, i.e., participants 1, 4, 8, and 9 show an obvious decrease in error when in the listening room environment, while others show the opposite trend, i.e., participants 5, 6, and 10. Additionally, participants 1 and 6 display front-back confusions.

Figures 5(a) and 5(b) depict scatter plots of the NH and BiHA user groups response error as a function of the target angle, where the error is calculated by subtracting the response angle from the actual angle. The responses for all trials from all individuals are used to generate these scatter

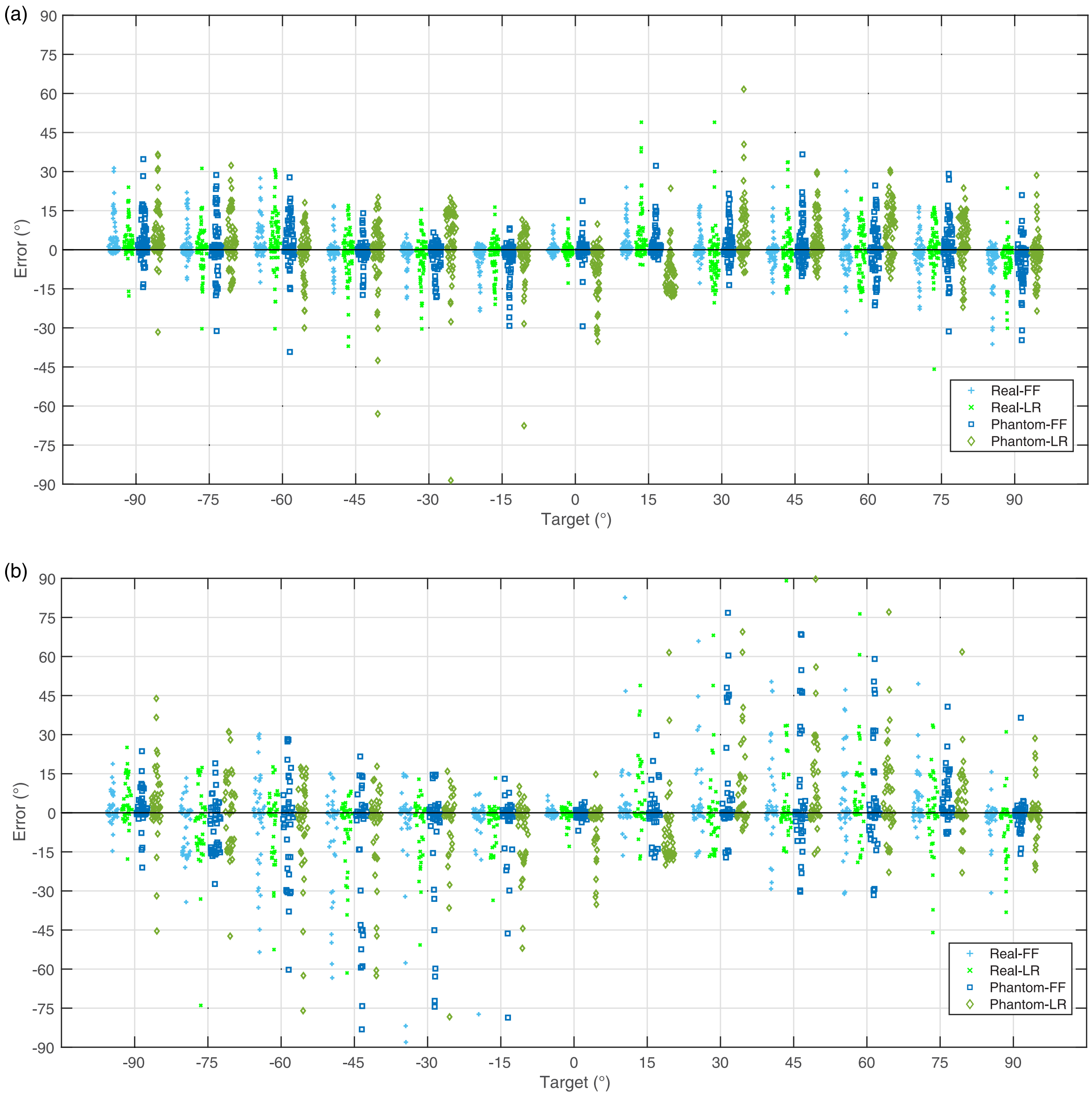

FIG. 5. Error calculated over all trials, plotted over the target angle. Excluded from these plots are outlier data points whose values were outside the range $-90 < \text{error} < 90$. Real, real sound source; Phantom, phantom sound source; FF, free-field; LR, listening room.

plots meaning there are 64 data points (i.e., 4 data points from 16 participants) per target angle for each condition in Fig. 5(a), while there are, respectively, only 40 data points in each scatter cluster in Fig. 5(b). Excluded from these plots are errors that are greater than 90°. This was a total of one data point for the NH group results (one for the free-field, phantom results) and 41 data points for BiHA group (15 for the free-field with real sound sources; 14 for the free-field with phantom sound sources; four for the listening room with real sound sources; and eight in the listening room with phantom sound sources).

For the NH group, the RMSE for real sources in the free-field environment was 5.3°, while the RMSE calculated over the responses for phantom sources whose positions corresponded with these real sources was 6.8°. The RMSE was 6.7° for all phantom source positions present in the test for free-field conditions. In the listening room setup, the RMSE was 7.5° for the real sources, and 9.8° when calculated over the corresponding phantom sources. When all phantom source positions were considered, the RMSE reduced to 9.4°.

For the BiHA user group, the RMSE was 13.0° for real sources in the free-field environment, 14.8° when calculated over the corresponding phantom source positions, and 16.2° when calculated over all the phantom source positions. In the listening room, the RMSE was 12.9°, 14.0°, and 15.6°, respectively. Tables II and III list the confidence intervals, estimates, standard errors (Std. Error), degrees of freedom (df), t values and p values ($Pr(>|t|)$) of the mixed effect models. The statistical analysis revealed that, for the NH group, all three factors, i.e., target angle, environment, and source type, had a significant effect on the localization test results. In addition, there was an interaction between the target angle and the environment. However, no interactions

TABLE II. Model coefficients for the statistical analysis performed on the control group results. Marginal $R^2 = 0.14$; conditional $R^2 = 0.29$. *p < 0.05; **p < 0.01; ***p < 0.001.

	2.5%	97.5%	Estimate	Standard error	df	t value	Pr (>|t|)
(Intercept)	3.37	7.04	5.20	0.94	83.20	5.52	$3.69e^{-07***}$
Environment	3.06	6.79	4.92	0.95	426.00	5.15	$3.90e^{-07***}$
Source type	−4.70	−0.97	−2.84	0.95	426.00	−2.97	0.003**
Angle	0.01	0.06	0.039	0.01	426.00	2.95	0.003**
Environment × source type	−3.15	0.46	−1.34	0.92	426.00	−1.74	0.14
Source type × angle	−0.003	0.05	0.02	0.01	426.00	1.74	0.08
Environment × angle	−0.06	0.03	−0.03	0.01	426.00	−2.16	0.03*

were found between target angle and source type, as well as source type and environment. For the BiHA group, the statistical analysis revealed a significant effect of target angle, but no significant differences in performance between listening environments and sound source type, nor did the analyses reveal an interaction between any of the three variables.

IV. DISCUSSION

The RMSE values reported in this study are consistent with previously reported values found in scientific literature. In particular, a previous study with a near identical loudspeaker setup tested localization of real sound sources and found RMSE values of 6° for 57 young NH participants and 12° for 17 bilateral hearing aid users when measured in a listening room ($RT_{60} < 0.1$ s) (Dorman *et al.*, 2016), which may be compared to the reported 7.5°and 12.9°for the NH and bilateral hearing aid users, respectively, in the listening room condition in the current study. Note that the BiHA users in this previous study also utilized their own devices during testing, and the young NH group and the BiHA group were also not age-matched. The differences in RMSE values are comparatively small, despite the fact that the stimulus utilized in the previous study was a 200-ms white noise burst as opposed to the 2-s speech-shaped noise stimulus used in the present study. Nonetheless, the relatively small differences in RMSE values that occur may be related to the differences in stimuli length, devices and settings, as well as room geometry and acoustic properties. Additionally, variability is inevitably introduced when using clinical devices, which were not programmed in a controlled way for the study. Each individual's hearing aid may have been processing the signals in different ways, and may or may not have been optimized binaurally, as they were recruited from the general public and were often not patients of the clinic. It may also be noted that, while some studies have found that aided hearing led to increased localization error (Van den Bogaert *et al.*, 2011, 2006), others have countered these claims (Drennan *et al.*, 2005; Johnson *et al.*, 2017), which leads to speculation that the variability in performance may be due, at least in part, to the variability present among their devices and the way they were programmed.

Nevertheless, the RMSE values found in the aforementioned study are very similar to the values achieved in this present study for both groups in the listening room setup. The RMSE values of the present study are also in keeping with the findings of an in-depth survey of 29 different localization studies, which utilized different methods, setups, stimuli, and listening environments (Akeroyd, 2014). Twelve of the studies focused on horizontal localization and compared the performance of older listeners with symmetric hearing loss to the performance of young NH listeners, i.e., the groups were not age-matched, in a similar manner to the present study (Abel and Hay, 1996; Best *et al.*, 2011; Best *et al.*, 2010; Brungart *et al.*, 2014; Keidser *et al.*, 2009; Lorenzi *et al.*, 1999; Neher *et al.*, 2011; Noble *et al.*, 1994; Smith-Olinde *et al.*, 1998; Vaillancourt *et al.*, 2011; Van den Bogaert *et al.*, 2011; Van den Bogaert *et al.*, 2006). The mean RMSE for the hearing-impaired older listeners was 14° and the difference in error between the two groups ranged from 5° to 10°, with NH groups having lower error values. Additionally, 11 studies included in the survey looked at aided horizontal localization, and a mean RMSE value of 11° was calculated for broadband stimuli in aided conditions (Best *et al.*, 2010; Brungart *et al.*, 2014; Byrne *et al.*, 1996; Byrne *et al.*, 1998; Drennan *et al.*, 2005; Keidser

TABLE III. Model coefficients for the statistical analysis performed on the BiHA group results. Marginal $R^2 = 0.01$; conditional $R^2 = 0.40$. *p < 0.05; ***p < 0.001.

	2.5%	97.5%	Estimate	Standard error	df	t value	Pr (>|t|)
(Intercept)	10.85	32.56	21.71	5.49	20.56	3.94	0.0007***
Environment	−14.83	1.66	−6.58	4.2	264.00	−1.55	0.12
Source type	−10.57	5.93	−2.32	4.2	264.00	−0.54	0.58
Angle	−0.25	−0.02	−0.13	0.05	264.00	−2.35	0.019*
Environment × source type	−6.95	9.05	1.05	4.11	264.00	0.25	0.79
Source type × angle	−0.12	0.14	0.01	0.06	264.00	0.15	0.87
Environment × angle	−0.014	0.25	0.11	0.06	264.00	1.73	0.08

https://doi.org/10.1121/10.0035828

et al., 2009; Köbler and Rosenhall, 2002; Lorenzi *et al.*, 1999; Noble *et al.*, 1998; Vaillancourt *et al.*, 2011; Van den Bogaert *et al.*, 2011; Van den Bogaert *et al.*, 2006). These RMSE values are not dissimilar from the values reported in the present study.

While previous studies have investigated the localization error with attention to the effect of particular factors, such as microphone placement or variations in processing, the main focus of this study was to determine the effect of the test environment and the specific method of amplitude panning on localization ability. For the NH group, there was a significant effect of both these variables, but interestingly these two factors do not interact with each other. There was also an effect of the target angle, and more importantly an interaction between target angle and environment. This may have been due to reflective surfaces in the room causing an effect at some angles, or due to the room helping in resolving some front-back differences. In effect, these findings imply that, if an anechoic chamber, or a similar free-field environment, is not accessible, a sound-treated room is a feasible option for such tests. However, the error in a given localization test will, on average, increase in the listening room in comparison to free-field, and this increase is consistent regardless of whether real or phantom sources are employed, but can vary based on the azimuth angle.

On an individual basis, some participants from the BiHA group have smaller error values in the listening room environment in comparison to in the free-field environment, which is the opposite behavior observed in the NH group and for many of the other hearing-impaired participants. This was not simply the consequence on the order of testing, which may have been the reason for the resolution of some of the front-back confusions for participants 1 and 6. Half of the participants who have smaller error values in the listening room were randomly assigned to the free-field condition first and half to the non–free-field condition first. The exact explanation for this behavior is unknown, but it may be that the small number of lateral reflections and reverberation that were caused by the walls in the listening room helped with better externalization of the sound sources, which in turn helped them achieve better azimuth localization performance. Previous research has indicated that free-field stimuli are more difficult to externalize, while reverberation helps facilitate externalization (Best *et al.*, 2020). An alternative explanation relates to the adaptive nature of modern hearing aids, as some of the hearing aids worn by participants may have automatically modified the processing approaches to adapt to the different listening environments. Yet another explanation may be differences in the form factor of the devices between participants (e.g., vent style, microphone location), as these variables were not controlled in the present study.

There was no statistically significant effect of the room or sound source type for the BiHA group. The target angle was found to cause a significant difference in their localization performance, but no interactions were found between any factors. Furthermore, the *t* value of the analysis indicates that the effect of the target angle was negative, which contrasts with the positive effect of angle found in the control group's results. In Figs. 4 and 5(b), it can be seen that some BiHA participants displayed larger errors when the target angle approached 45°. This behavior can also be seen in a previous study which investigated the effect of hearing aids on localization (Van den Bogaert *et al.*, 2006). The negative effect of angle for the BiHA group may, therefore, have been due to the hearing devices, i.e., the processing within them and their form factor differences, and their subsequent effect on localization.

Due to the large variation in the BiHA group responses, the small sample size, and the uncontrolled variables of device brand, model, and vent types, caution must be given when interpreting the results of the study. Nevertheless, in regards to the sound source type, the key findings of Ellis and Souza (2020) were similar to the findings reported here (i.e., a lack of difference) when investigating the effect of hearing loss on localization of amplitude panned and real sound sources with unaided hearing-impaired listeners, suggesting that the adaptive nature of the hearing aids may not be the only driver of the individual variability and that the properties of the room itself may have contributed.

Overall, these statistical analysis findings imply that amplitude panning could potentially be used as a clinical tool in sound-treated rooms; but given the small sample size and differences among the devices utilized in the BiHA group, further investigation is warranted. However, it should be noted that the NH group results imply that the localization accuracy would be lower in such conditions than if localization tasks were to be performed in a free-field environment with real sound sources. Therefore, there exists a trade-off between the cost of the setup and the achievable localization accuracy. It remains to be seen whether such a setup can be used for other tasks, such as to train better localization skills in BiHA or other hearing device users, in a similar manner to the ongoing work, which utilizes virtual sounds played over headphones to achieve the same goal (Nisha and Kumar, 2016; Nisha *et al.*, 2023; Syeda *et al.* 2023). Other avenues of future work include expanding the present study to collect data on front-back confusion behavior, as well as median plane localization behavior.

V. CONCLUSION

This article presented the findings of a study which compared the localization error of ten bilateral hearing aid users for two types of sound sources (real and phantom sound sources) in two different listening environments (free-field, anechoic chamber and non–free-field, sound-treated listening room). A control group of 16 NH listeners were also participants in the study. Sound sources were produced over loudspeakers placed 15° apart in both listening environments by playing speech-shaped noise through a loudspeaker directly to create real sound sources, or by pair-wise amplitude panning to create phantom sound sources.

The RMSEs of the two participant groups were calculated from their responses and found to be similar to previously reported values in the literature. These RMSE values were slightly larger for the phantom sources compared to the real sources, and the values calculated for the responses in the listening room environment were found to be larger than those obtained in the free-field environment. These test results were further analyzed via a mixed effect model to compare an individual's localization performance for the different target angles across the two source types and the two listening environments to determine if there were any statistically significant differences in performances across these factors, as well as if there was any interaction between these factors. It was found that neither the sound source type nor the listening environment had a significant effect on the localization performance of the BiHA user group. On the other hand, both of these factors had a statistically significant effect on the localization performance of the NH participants; and there was an interaction between the target angle and the listening environment. Perhaps more interestingly, there were no other significant interactions found in the analysis.

These results support the feasibility of using amplitude panning over loudspeakers in measuring localization accuracy, both in free-field and non–free-field listening room setups. However, it must be noted that using such a method together with hearing devices may lead to a certain level of variability that is, at this stage, unexplainable as the present study did not involve participants that utilized identical device models, used the same venting strategy, nor employed the same processing within their devices. Despite that the localization performance for both phantom and real sound sources was not significantly different between the test environments and across sound source types for the BiHA group, the effect of the room varied substantially from individual to individual. Furthermore, due to the small sample size, caution should be applied when interpreting the results of the statistical analysis. Therefore, it will be important to explore these differences, and the role hearing devices (including factors such as variations in their fitting and processing) may play in creating such differences before introducing such a test clinically.

ACKNOWLEDGMENTS

We thank the reviewers and the editor for their constructive comments and feedback.

AUTHOR DECLARATIONS

Conflict of Interest

There are no conflicts of interest for the authors of this manuscript to disclose.

Ethics Approval

The research procedures were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391), and all participants provided written informed consent for the study procedures and received monetary compensation for their participation in the listening test.

DATA AVAILABILITY

Fully anonymized data from this study may be made available upon reasonable request.

[1]The Institute of Electronic Music and Acoustics plug-in suite's distance compensator may be found at https://plugins.iem.at/.

Abel, S. M., and Hay, V. H. (**1996**). "Sound localization the interaction of aging, hearing loss and hearing protection," Scand. Audiol. **25**(1), 3–12.

Akeroyd, M. A. (**2014**). "An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners," Trends Hear. **18**, 2331216514560442.

Akeroyd, M. A., and Guy, F. H. (**2011**). "The effect of hearing impairment on localization dominance for single-word stimuli," J. Acoust. Soc. Am. **130**(1), 312–323.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**(4), 2169–2180.

Armstrong, C., Thresh, L., Murphy, D., and Kearney, G. (**2018**). "A perceptual evaluation of individual and non-individual HRTFs: A case study of the Sadie II database," Appl. Sci. **8**(11), 2029.

Bates, D. (**2014**). "Fitting linear mixed-effects models using lme4," arXiv:1406.5823.

Besing, J. M., and Koehnke, J. (**1995**). "A test of virtual auditory localization," Ear Hear. **16**(2), 220–229.

Best, V., Baumgartner, R., Lavandier, M., Majdak, P., and Kopčo, N. (**2020**). "Sound externalization: A review of recent research," Trends Heari.**24**, 2331216520948390.

Best, V., Carlile, S., Kopčo, N., and van Schaik, A. (**2011**). "Localization in speech mixtures by listeners with hearing loss," J. Acoust. Soc. Am. **129**(5), EL210–EL215.

Best, V., Kalluri, S., McLachlan, S., Valentine, S., Edwards, B., and Carlile, S. (**2010**). "A comparison of CIC and BTE hearing aids for three-dimensional localization of speech," Intl. J. Audiol. **49**(10), 723–732.

Best, V., Keidser, G., Buchholz, J. M., and Freeston, K. (**2015**). "An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment," Int. J. Audiol. **54**(10), 682–690.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

Brinkmann, F., Dinakaran, M., Pelzer, R., Grosche, P., Voss, D., and Weinzierl, S. (**2019**). "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," J. Audio Eng. Soc. **67**(9), 705–718.

Brungart, D. S., Cohen, J., Cord, M., Zion, D., and Kalluri, S. (**2014**). "Assessment of auditory spatial awareness in complex listening environments," J. Acoust. Soc. Am. **136**(4), 1808–1820.

Brungart, D. S., Cohen, J. I., Zion, D., and Romigh, G. (**2017**). "The localization of nonindividualized virtual sounds by hearing-impaired listeners," J. Acoust. Soc. Am. **141**(4), 2870–2881.

Byrne, D., Noble, W., and Glauerdt, B. (**1996**). "Effects of earmold type on ability to locate sounds when wearing hearing aids," Ear Hear. **17**(3), 218–228.

Byrne, D., Sinclair, S., and Noble, W. (**1998**). "Open earmold fittings for improving aided auditory localization for sensorineural hearing losses with good high-frequency hearing," Ear Hear. **19**(1), 62–71.

Cubick, J., and Dau, T. (**2016**). "Validation of a virtual sound environment system for testing hearing aids," Acta Acust. Acust. **102**(3), 547–557.

Demonte, P. (**2019**). "Harvard speech corpus: Audio recording 2019," available at https://salford.figshare.com/collections/HARVARD_speech_corpus_-_audio_ recording_2019/4437578/1.

Dorman, M. F., Loiselle, L. H., Cook, S. J., Yost, W. A., and Gifford, R. H. (**2016**). "Sound source localization by normal-hearing listeners, hearing-impaired listeners and cochlear implant listeners," Audiol. Neurotol. **21**(3), 127–131.

Drennan, W. R., Gatehouse, S., Howell, P., Van Tasell, D., and Lund, S. (**2005**). "Localization and speech-identification ability of hearing-

impaired listeners using phase-preserving amplification," Ear Hear. **26**(5), 461–472.
Ellis, G. M., and Souza, P. E. (**2020**). "The effect of hearing loss on localization of amplitude panned and physical sources," J. Am. Acad. Audiol. **31**(09), 690–698.
Engel, I., Daugintis, R., Vicente, T., Hogg, A. O., Pauwels, J., Tournier, A. J., and Picinali, L. (**2023**). "The SONICOM HRTF dataset," J. Audio Eng. Soc. **71**(5), 241–253.
Fernandez, J., McCormack, L., Hyvärinen, P., and Kressner, A. A. (**2024**). "Investigating sound-field reproduction methods as perceived by bilateral hearing aid users and normal hearing listeners," J. Acoust. Soc. Am. **155**(2), 1492–1502.
Fernandez, J., Sivonen, V., and Pulkki, V. (**2022**). "Investigating bilateral cochlear implant users' localization of amplitude-and time-panned stimuli produced over a limited loudspeaker arrangement," Am. J. Audiol. **31**, 143–154.
Goverts, S. T., Houtgast, T., and van Beek, H. H. (**2002**). "The precedence effect for lateralization for the mild sensory neural hearing-impaired," Hear. Res. **163**(1), 82–92.
Grimm, G., Ewert, S., and Hohmann, V. (**2015**). "Evaluation of spatial audio reproduction schemes for application in hearing aid research," Acta Acust. Acust. **101**(4), 842–854.
ISO 3382-1:2009 (**2009**). "Acoustics–measurement of room acoustic parameters–part 1: Performance spaces" (International Organization for Standardization, Geneva, Switzerland).
Johnson, J. A., Xu, J., and Cox, R. M. (**2017**). "Impact of hearing aid technology on outcomes in daily life: III. Localization," Ear Hear. **38**(6), 746–759.
Keidser, G., O'Brien, A., Hain, J.-U., McLelland, M., and Yeend, I. (**2009**). "The effect of frequency-dependent microphone directionality on horizontal localization performance in hearing-aid users," Int. J. Audiol. **48**(11), 789–803.
Köbler, S., and Rosenhall, U. (**2002**). "Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification: Localizacíon horizontal y discriminación del lenguaje con adaptación unilateral y bilateral de auxiliares auditivos," Int. J. Audiol. **41**(7), 395–400.
Koski, T., Fazi, F. M., and Pulkki, V. (**2014**). "Beamformer performance in sound fields produced by amplitude panning," in *Proceedings of the 55th International Audio Engineering Society Conference on Spatial Audio*, Helsinki, Finland (August 2014).
Koski, T., Sivonen, V., and Pulkki, V. (**2013**). "Measuring speech intelligibility in noisy environments reproduced with parametric spatial audio," in *Audio Engineering Society Convention 135* (Audio Engineering Society, New York).
Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2017**). "lmertest package: Tests in linear mixed effects models," J. Stat. Softw. **82**, 1–26.
Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in hearing-impaired listeners," J. Acoust. Soc. Am. **105**(6), 3454–3463.
Mansour, N., Marschall, M., May, T., Westermann, A., and Dau, T. (**2021**). "Speech intelligibility in a realistic virtual sound environment," J. Acoust. Soc. Am. **149**(4), 2791–2801.
McCormack, L., and Politis, A. (**2019**). "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio* (Audio Engineering Society, New York).
Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?," J. Audio Eng. Soc. **44**(6), 451–469.
Mueller, M. F., Kegel, A., Schimmel, S. M., Dillier, N., and Hofbauer, M. (**2012**). "Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes," J. Acoust. Soc. Am. **131**(6), 4732–4742.
Neher, T., Laugesen, S., Søgaard Jensen, N., and Kragelund, L. (**2011**). "Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities?," J. Acoust. Soc. Am. **130**(3), 1542–1558.
Nisha, K., and Kumar, U. A. (**2016**). "Effect of localization training in horizontal plane on auditory spatial processing skills in listeners with normal hearing," J. Indian Speech Lang. Hear. Assoc. **30**(2), 28–39.
Nisha, K. V., Uppunda, A. K., and Kumar, R. T. (**2023**). "Spatial rehabilitation using virtual auditory space training paradigm in individuals with sensorineural hearing impairment," Front. Neurosci. **16**, 1080398.
Noble, W., Byrne, D., and Lepage, B. (**1994**). "Effects on sound localization of configuration and type of hearing impairment," J. Acoust. Soc. Am. **95**(2), 992–1005.
Noble, W., Byrne, D., and Ter-Horst, K. (**1997**). "Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing-impaired listeners," J. Acoust. Soc. Am. **102**(4), 2343–2352.
Noble, W., Sinclair, S., and Byrne, D. (**1998**). "Improvement in aided sound localization with open earmolds: Observations in people with high-frequency hearing loss," J. Am. Acad. Audiol. **9**(1), 25–34.
Pulkki, V. (**1997**). "Virtual sound source positioning using vector base amplitude panning," J. Audio Eng. Soc. **45**(6), 456–466.
R Core Team (**2024**). "R: A language and environment for statistical computing," *R Foundation for Statistical Computing*, Vienna, Austria, available at https://www.R-project.org/.
Simon, L. S., Wuethrich, H., and Dillier, N. (**2017**). "Comparison of higher-order ambisonics, vector-and distance-based amplitude panning using a hearing device beamformer," in *Proceedings of the 4th International Conference on Spatial Audio*, Graz, Austria (September 2017).
Smith-Olinde, L., Koehnke, J., and Besing, J. (**1998**). "Effects of sensorineural hearing loss on interaural discrimination and virtual localization," J. Acoust. Soc. Am. **103**(4), 2084–2099.
Syeda, A., Nisha, K., and Jain, C. (**2023**). "Test–retest reliability of virtual acoustic space identification test in school-going children," Am. J. Audiol. **32**(3), 574–582.
Vaillancourt, V., Laroche, C., Giguere, C., Beaulieu, M.-A., and Legault, J.-P. (**2011**). "Evaluation of auditory functions for royal Canadian Mounted Police officers," J. Am. Acad. Audiol. **22**(6), 313–331.
Van den Bogaert, T., Carette, E., and Wouters, J. (**2011**). "Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna," Int. J. Audiol. **50**(3), 164–176.
Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). "Horizontal localization with bilateral hearing aids: Without is better than with," J. Acoust. Soc. Am. **119**(1), 515–526.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**(1), 111–123.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening: I. Psychophysical validation," J. Acoust. Soc. Am. **85**(2), 868–878.
Zenke, K. (**2021**). "Spatial release from masking in children with and without auditory processing disorder in real and virtual auditory environments," Ph.D. thesis, University College London.
Ziegelwanger, H., Kreuzer, W., and Majdak, P. (**2015**). "Mesh2HRTF: Open-source software package for the numerical calculation of head-related transfer functions," in *22nd International Congress on Sound and Vibration*.