

Experimental Study

Experiments on localization accuracy with non-individual and individual HRTFs comparing static and dynamic reproduction methods

Josefa Oberem¹⁾ , Jan-Gerrit Richter¹⁾ , Dorothea Setzer¹⁾,
Julia Seibold²⁾ , Iring Koch²⁾ , Janina Fels¹⁾ 

¹⁾ Teaching and Research Area of Medical Acoustics,
Institute of Technical Acoustics, RWTH Aachen University, Germany

²⁾ Institute of Psychology, RWTH Aachen University, Germany

Corresponding Author: josefa.oberem@akustik.rwth-aachen.de

Abstract: Binaural reproduction can be used in listening experiments under real-life conditions to achieve a high realism and good reproducibility. In recent years a clear trend to more individual reproduction can be observed as the ability to measure individual head-related-transfer-functions (HRTFs) is becoming more widespread. The question of the accuracy and reproduction methods needed for a realistic playback however has not been sufficiently answered. To evaluate an appropriate approach for binaural reproduction via headphones different head-related-transfer-functions (HRTFs) and reproduction methods were compared in this paper. In a listening test eleven explicitly trained participants were asked to localize eleven sound sources positioned in the right hemisphere using the proximal pointing method. Binaural stimuli based on individually measured HRTFs were compared to those of an artificial head in a static reproduction of stimuli and in three dynamic reproduction methods of different resolutions (5° , 2.5° and 1°). Unsigned errors in azimuth and elevation as well as front-back-confusions and in-head-localization were observed. Dynamic reproduction of any resolution applied turned out fundamental for a reduction of undesired front-back-confusions and in-head-localization. Individually measured HRTFs showed a smaller effect on localization accuracy compared to the influence of dynamic sound reproduction. They were mainly observed to reduce the front-back-confusion rate.

Keywords: localization, individual head-related transfer functions, dynamic reproduction, front-back-confusion, in-head localization

1 Introduction

Parts of this study were presented at the conference DAGA, Munich, Germany, 2018 [1] and are published in the PhD-thesis [2].

The progress of individual binaural reproduction of recent years has brought the capability to study complex cognitive processes which require real-life conditions and high level of plausibility under a controlled and reproduce-able environment. The accuracy and plausibility of these binaural reproductions is highly dependent on several factors. These factors range from, for example, the set of head-related-transfer-functions (HRTFs) used (e.g. individuality, resolution, section) as well as the type of reproduction (e.g. dynamic, static). The importance of each of these components individually has been discussed for several years. The benefit of the use of individual HRTFs for localization tasks has been discussed for several decades with some of the

earliest work by Wightman, Kistler and Wenzel [3], [4]. A gain from small head movements was described first in Wallach [5] and could be reproduced several times [6]–[8].

However, no study has systematically looked at the influence of the *resolution* of high resolution individual HRTFs and tried to quantify the influence of *both* high resolution individual HRTFs as well as dynamic binaural reproduction. Consequently, when dealing with a reproduction problem that requires real-life conditions, one must assume that all factors will benefit equally and are crucially important.

As taking all these factors into account is a time-consuming and expensive procedure the goal of this publication is to find a HRTF dataset which, combined with a reproduction method (static / dynamic), will result in a performance comparable to performance achieved with real sources. To this end localization



experiments were conducted since localization performance delivers a rather general measure to indicate the quality of binaural auditory displays, not limited by any paradigm conditions. In the experiment individually measured HRTFs were compared to those of an artificial head in a static reproduction of stimuli and in three dynamic reproduction methods of different resolutions (5° , 2.5° and 1°).

2 State of the Art

Several studies investigated localization performance for real sources in comparison with virtual sound sources using a binaural headphone reproduction. There was a broad agreement that results nearly as accurate as results for real sources could also be achieved by using individual HRTFs for sound reproduction [3], [9]–[11]. Headphone reproduction, however, featured a generally higher rate of front-back-confusions (e.g. [3]). Investigations on experiments using individual and non-individual HRTF measurements demonstrated the necessity to use individual HRTFs in order to attain results nearly equivalent to real sources. Non-individual HRTF-measurements did not only show an overall reduced localization performance, but were particularly prone to front-back-confusions [4], [9], [10]. Møller and colleagues reported about more errors located in the median plane for dummy head HRTFs [11]. Additionally, Andéol and colleagues observed a higher rate of front-back-confusions also for individual HRTFs, but stated greater errors in elevation for the non-individual reproduction via headphones [12].

Dynamic reproduction, allowing for head movements, significantly improved spatial localization in terms of a reduced reversal error [9], [13]–[16]. Indeed, head movements seemed to play an essential role for a realistic sound perception and have been subject to a wide range of studies (e.g. [6], [17]–[19]).

Most of the published work investigating the benefit of dynamic reproduction in combination with individually measured HRTFs suffer from two drawbacks. Either the resolution of the individual HRTFs is rather low [8], [20], [21], which would artificially inflate the measured localization errors, or the tested positions are all located in the horizontal plane [8] which would reduce the expected gain from using individual HRTFs in the first place. Sandvad indeed looked at the influence of HRTF resolution, but used interpolation from relatively low resolution to achieve high resolution samples [20]. An extensive measurement of individual HRTFs was done in [9] but the authors voice concerns about the measurement quality themselves. Indeed, only the recent development of measurement systems that are capable of acquiring high resolution individual HRTFs in a short period of time (e.g. [22]) enable a more thorough investigation.

Similar to Begault and colleagues [8], the present investigation combined and compared individual and

non-individual binaural stimuli used in static and dynamic reproduction. Additionally, a minimum resolution necessary for dynamic reproduction was to be estimated, since expenditure of time constitutes a relevant factor for feasibility of individual HRTF-measurements. To cover the range of nearly all audible frequencies, broadband noise pulses were used instead of speech stimuli [8].

Based on previous findings (e.g. [10]), individual HRTFs were expected to provide better results for localization ability. The same holds for dynamic sound reproduction, assumed to be superior to static sound reproduction (e.g. [4]) with an increased localization performance for high HRTF-resolutions.

3 Methods and equipment

This chapter will give an overview of all used systems and algorithms.



Figure 1: The HRTF Arc with 64 loudspeakers, exemplarily an artificial head is positioned in the center [2].

3.1 Binaural measurements and equalization method

One of the aims of the experiment was to determine the influence of the HRTF resolution of dynamic reproduction on the localization accuracy. To this end, three

different HRTF resolutions were needed: 5° , 2.5° and 1° in azimuth and elevation, respectively (see section 4).

3.1.1 HRTF measurement

The HRTF measurements took place in a semi-anechoic chamber ($l \times w \times h = 9.2 \times 6.2 \times 5.0 \text{ m}^3$) with a lower boundary frequency limit of 200 Hz.

The measurement system was specifically developed for fast individual HRTF measurements with a small impact on the measurement signal [22]. It used 64 loudspeakers (1 inch) arranged in an incomplete semi-circle from elevation 0° to 160° with 2.5° degree resolution (see Figure 1). The participants had two Sennheiser KE3 microphones inserted into their ears at the entrance of the blocked ear canals. They were positioned standing at two meters ear height in the center of the measurement arc on a turntable. The measurement signal was an interleaved sweep of all 64 loudspeakers with a frequency range from 500 Hz to 22,050 Hz with a time delay between loudspeaker starts of 40 ms and an overall length of 3.38 s.

To reduce unnecessary head movement, the participant's head was positioned against a head rest. The turntable moved the participant in azimuth angle of 2.5° . The overall measurement duration was ten minutes.

The dummy head was measured in the same measurement setup with identical measurement signal.

While the 5° resolution could be achieved by simply only using every second azimuth step of the 2.5° measurement data, an interpolation was needed to achieve the 1° resolution.

3.1.2 Interpolation

As shown by Duraiswaini and colleagues as well as Richter and colleagues [23], [24], the spherical harmonic decomposition could be used to interpolate HRTF datasets. To this end, each frequency bin from all measured directions could be regarded as a frequency dependent directivity, measured on a spherical surface. This data was then transformed to spherical harmonic coefficients. The interpolation depended on the maximum order chosen for this transformation. The lower the maximum order was chosen, the more interpolation would occur. It should be noted that this interpolation would not preserve the measured data points as they were also affected.

For this publication, a maximum order of 50 was chosen. As the measured data was not acquired on a whole sphere, regularization as presented by Duraiswaini and colleagues [23] with the regularization factor of 10^{-5} was used to reduce unwanted artifacts. After the decomposition into a finite number of coefficients a reconstruction to arbitrary points could be done. These points were chosen with the same elevation values as the measurement but with a higher azimuth resolution of 1° instead of 2.5° .

3.1.3 HpTF measurement

Headphone-transfer-functions (HpTFs) were measured to calculate an adequate robust equalization. After each of in total eight HpTF measurements, headphones were repositioned on the participants head [25]. To give the best comfort, the repositioning was done by the participant himself/herself. Based on Masiero and Fels [25] the equalization was calculated using the mean of the HpTF measurements. Since phase information was lost at this process, minimum phase was used. Furthermore, notches in the high frequency range were smoothed [25], [26].

3.2 Stimuli

Participants were asked to localize a train of pulsed white Gaussian noise. The frequency domain ranged from 100 Hz to 20 kHz providing high-frequency pinna cues essential for localization [27]. The use of white noise was in accordance with previous studies and strongly recommended for the investigations of localization with head movements as broadband signals will excite all changes in monaural in binaural information during tiny head motion (e.g. [16], [28]–[30]).

Head movements were hardly witnessed in case of short sound events smaller than 200 ms, but best observable for minimum duration of 2 s [31]. For this purpose, the total stimulus length was set to 3.7 s and was composed of five alternating 0.3 s- and 1.2 s-bursts of constant sound level with on- and offset ramps (50 ms rise and 50 ms fall time each), interrupted by small pauses of 100 ms.

3.3 Real-time binaural synthesis

The stimuli had to be convolved with both HpTF and HRTF filters. As participants were free to move their head, different HRTF filters had to be used, depending on the current head orientation. These filter changes lead to the requirement of a real-time controlled playback that could adjust the playback with no noticeable delay. To this end, a real-time binaural synthesis application and a tracking system to monitor head movement was needed.

3.3.1 Auralization

The real-time auralization was realized with the software Virtual Acoustics (VA) which was developed at the Institute of Technical Acoustics [32]. It allowed for fast exchange of filters as a result of head movement and could be controlled by various tracking systems. It used uniform partitioned convolution [33] and a block size of 128 samples with sampling rate of 44100 Hz. Neighboring HRTFs were selected on a nearest neighbor basis. The latency of the auralization software is in the worst case two block sizes. One internal buffer for the convolution and one hardware buffer of the soundcard. This would calculate to a worst case latency of approximately 6 ms.

3.3.2 Tracking

An optical tracking system was used to monitor several parts of the experiment. The system used four infrared cameras mounted in the room. The system operated with a sampling frequency of 120 Hz and was capable of monitoring several tracking bodies simultaneously. The first monitored motion was the participants head movement. To this end, a tracking body was fastened centrally on the headphone. This point was be monitored throughout the experiment. As this tracking body is not located in the center of the head, the movement of this tracking body did not correspond in all axes to the actual head movement. To correct this offset, a head calibration was done individually. For this calibration, the position of the two ears was marked at the beginning of the test with two additional tracking bodies. From these two positions and the position on top of the head, the center point of the head could be approximated more closely. This movement data was given directly to the auralization software to update HRTF filters if needed. The second use of the tracker was its implementation into the pointing method described in detail in Section 3.4. It involved a marker to indicate spatial positions and an input device, both of which were equipped with tracking markers to monitor the two devices. The latency of the tracker is given by the manufacturer as approximately 8 ms. Combined with the software latency the overall system latency is at worst 14 ms.

3.4 Pointing Method

Any investigation into localization accuracy has to be viewed in the context of the used pointing method as the method can heavily influence the achievable accuracy. Over the years, many different approaches to pointing have been proposed and evaluated. Notable methods can be divided into exocentric methods, in which the participant indicates the auditory event on a sphere representing auditory space (e.g GELP (God's eye view localization point) [34], 2D/3D interfaces [8], [35]) and egocentric methods, in which the participants own center is used as a reference point and pointing is done with an extension of the body (e.g laser pointing [36], [37] or head pointing [21], [38], [39]).

Egocentric methods are generally regarded as more precise [21], [40], [41] and head pointing methods in particular are recommended by [21]. For closed-loop localization however, in which the sound plays during the pointing task, this method will always reflect localization accuracy in the front, as the participant will move their head towards the source.

In this publication, proximal pointing is used as a pointing method. The egocentric method was introduced by Bahu and colleagues [42] and is recommended for closed-loop localization tasks. The listener indicates spatial positions by placing a hand-held marker in the region of the head where (s)he perceives the incident sound (see Figure 2). For this technique the spatial co-

ordinate system of the pointing mechanism is concentric with the coordinate system of the stimulus, since localization is referred to the center of the head, or more precisely, to the midpoint of the interaural axis. Closed-loop localization is enabled, as the method does not require head or body rotation. It is practical for both stimuli presentations over headphones and loudspeakers. In this pointing method, only auditory and proprioceptive modalities are involved, neglecting potential visual cues if the hand is positioned in the field of view. According to Bahu and colleagues, proximal pointing yielded more accurate results for high elevations than head and manual pointing. Fast response times may prove advantageous with regard to fatigue and motivation.

Localization of sound sources by proximal pointing is a matter of aligning three points of reference: The center of the pointer is positioned onto the connecting line between sound source and the center of the head. To indicate the direction of the perceived auditory stimulus, only the center of the pointer is relevant, its orientation is of no further consequence. Participants were asked to always hold the pointer in the right hand. This limitation ensures an unchanged localization precision without effects due to handedness. As a consequence, source positions were limited to the right hemisphere. To confirm the indicated sound source position, participants were asked to push a button on a second device held in the left hand. The position of this input device was also observed. To indicate an in-head-localization (compare Section 4.2.4) participants were asked to raise their left hand holding the input device from default position up to the height of their head.

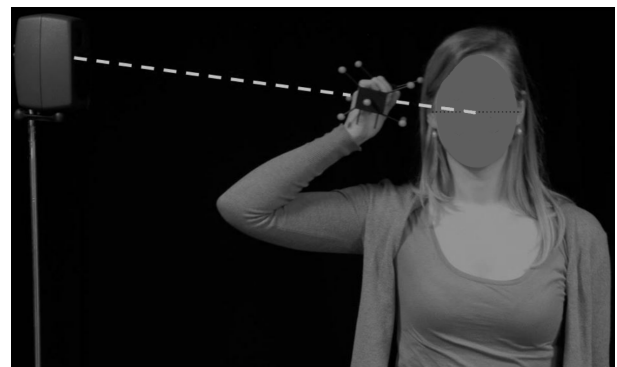


Figure 2: Participant localizing with proximal pointing method.

3.5 Participants

11 right-handed, student participants aged between 20 and 27 years with an average of (23.6 ± 2.4) years completed the localization experiment. Participants were equally divided in male and female listeners and the experiment was remunerated with 40 Euro. All participants were inexperienced in hearing with HRTF-data and not trained in localization prior to the

experiment, however they all achieved a localization accuracy of $< 30^\circ$ in the training session when listening to real sources (see Section 3.7.1). Hearing sensitivity (< 15 dB HL) has been verified with the aid of a pure tone audiometry between 125 Hz and 10 kHz.

3.6 Room setup

The listening tests took place in a hearing booth ($l \times w \times h = 2.3 \times 2.3 \times 1.98$ m³) to ensure a quiet environment during the test. Lights were turned off during the listening test to direct the focus to the aural sense other than the visual sense [43].

3.7 Experimental schedule

The experiment was split up in six sessions of about 45 min. each for every participant. While in session 1 individual HRTFs and HpTFs were measured (see Section 3.1) and audiometric data was collected, session 2, 3 and 4 were dedicated to training. Data used for analysis were only taken from session 5 and 6, which were performed within three days. 2 HRTFs \times 4 reproduction methods \times 11 positions \times 4 repetitions + 48 getting-warm trials = 400 trials were collected for data analysis.

3.7.1 Training

To get used to the proximal pointing method, three training sessions with optical feedback on a screen were performed. To be able to give feedback regarding the localization of the source and avoid learning how to localize with a specific HRTF set the training was done with real sources (loudspeakers: Genelec 6010). The training sessions were held in a highly damped room ($l \times w \times h = 8 \times 5 \times 2.65$ m³, reverberation time: 0.15 s) with real sources positioned in the right hemisphere of the participant. These source positions were different from the positions used in the listening test itself. Split up into blocks of 25 trials each in total 600 trials were delivered in the three training sessions.

4 Experimental design

4.1 Independent factors

The independent variables were HRTF (HRTF: individual vs. non-individual), reproduction method (RepMeth: static vs. dynamic (5°) vs. dynamic (2.5°) vs. dynamic (1°)) and position (Pos: Front vs. Side vs. Back) as within-subject variables.

4.1.1 HRTF

HRTFs used in the localization experiment were either individual or measured with an artificial head and therefore non-individual (see Section 3.1 for more information).

4.1.2 Reproduction Method

A static reproduction (i.e. head movements of the participant do not cause any change in binaural reproduction; the same HRTF is used throughout the presentation of the stimulus) was compared to a dynamic reproduction (i.e. head movements of the participant cause a change in binaural reproduction from a certain angle of movement; HRTFs are adjusted throughout the presentation of the stimulus). Three different resolutions of dynamic reproductions were compared: 5° , 2.5° and 1° (see Section 3.1 for more information).

4.1.3 Source position

The factor source position had three levels, categorized from eleven positions scattered in the right hemisphere, all with a distance of 1.5 m from the participant's center of the head (see Table 1 and Figure 3). There were four frontal positions, two on the median plane with difference in elevation of 30° as well as two positioned close to the median plane ($\varphi = 350^\circ$). Sources in back were designed to be symmetrically identical with those in front. However, the source with the lower elevation ($\vartheta = 120^\circ$) was not added due to limitations of the pointing method (see Section 3.4). Sources to the side were also designed to be symmetrically identical and to be positioned on one cone of confusion.

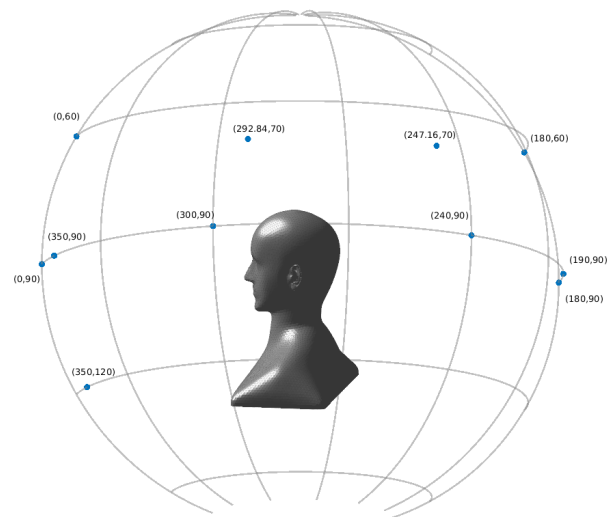


Figure 3: Source positions around listener in localization experiment.

4.2 Dependent factors

The dependent variables were localization error and in-head-localization. For a more detailed analysis the localization error was split into three different ways of counting the localization error: the unsigned horizontal error, unsigned vertical error and weighted front-back-confusions.

Table 1: Angles of the virtual sound source positions in localization experiment, where φ indicates the horizontal angle ($\varphi = 0^\circ$ indicates the front and increased counterclockwise) and ϑ the vertical angle ($\vartheta = 0^\circ$ was above the participant and $\vartheta = 180^\circ$ was underneath the participant.)

position	φ	ϑ
front	0°	60°
	0°	90°
	350°	90°
	350°	120°
side	300°	90°
	292.84°	70°
	247.16°	70°
	240°	90°
back	190°	90°
	180°	60°
	180°	90°

4.2.1 Unsigned horizontal error

The unsigned error gave a measure of absolute deviation from the correct direction of sound incidence. Deviations from the correct position were added up according to:

$$\overline{X_{unsigned}} = \frac{1}{N} \sum_{i=1}^N |x_i|. \quad (1)$$

Therefore, the unsigned horizontal error described the absolute deviation of the localized sound position to the actual sound source position in azimuth.

4.2.2 Unsigned vertical error

Accordingly, the unsigned vertical error described the absolute deviation of the localized sound position to the actual sound source position in elevation.

4.2.3 Weighted front-back-confusions

Front-back-confusions as one of the “most important impediment[s] for success of the binaural technique” [44] were taken account of by calculation of the rate of front-back-confusions. A front-back-confusion occurred when the actual sound source was positioned in the frontal hemisphere [$270^\circ; 90^\circ$), but localized in the rearward hemisphere [$90^\circ; 270^\circ$). A back-front-confusion described the inverted effect. In the following, there was no distinction drawn between back-front-confusions and front-back-confusions and both were called front-back-confusions. The occurrence of a reversal in a specific trial was denoted with the boolean

$$FBC = \begin{cases} 1, & \text{reversal occurred} \\ 0, & \text{no reversal occurred} \end{cases}, \quad (2)$$

Since front-back-confusions easily occur for lateral source positions, but simply due to localization impres-

sion (e.g. actual source position is at 260° and is localized at 275°), a weighting factor should find a remedy:

$$FBC_W = FBC \cdot \begin{cases} \frac{2 \cdot (\varphi_{act} - 270)}{180}, & \text{if } \varphi_{act} \in [270^\circ; 360^\circ] \\ \frac{2 \cdot (270 - \varphi_{act})}{180}, & \text{if } \varphi_{act} \in [180^\circ; 270^\circ] \end{cases}, \quad (3)$$

The definition was defined to the right hemisphere where φ_{act} was the actual sound source position.

4.2.4 In-head-localization

An “in-head-localization” was present when the perceived distance was smaller than the radius of the head [43]. The rate of in-head-localization, was computed thanks to position data of the input device held in left hand indicating a sound perception in the head by raising the arm (see also Section 3.4).

5 Results

The presented results were based on data from eleven participants.

In case Mauchly’s test indicated that the assumption of sphericity was violated, the Huynh-Feldt correction was applied. However, for a better overview uncorrected degrees of freedom were reported.

5.1 Unsigned horizontal error

Figure 4 shows unsigned horizontal error as a function of the position, reproduction method and the HRTF.

Although there was a non-significant trend towards smaller horizontal errors when listening to individual HRTFs (individual HRTFs: 24.2° vs. non-individual HRTFs: 27.4°), there was no significant effect on HRTFs in the unsigned horizontal error.

The main effect of reproduction method was significant, resulting in a significantly higher unsigned horizontal error for a static reproduction than for all three dynamic reproduction methods (static: 38.7° vs. dynamic (5°): 21.2° vs. dynamic (2.5°): 21.5° vs. dynamic (1°): 21.7°). The interaction of HRTF and reproduction method was significant. Post-hoc test showed that the unsigned horizontal error difference between individual and non-individual HRTFs was significantly greater for the static reproduction than for the dynamic reproductions (static: 9.86° vs. dynamic (5°): 0.20° vs. dynamic (2.5°): 0.64° vs. dynamic (1°): 2.18°).

The main effect of source position was not significant. However, unsigned horizontal errors were smallest for trials where the source was positioned to the side and greatest for trials where the source was positioned in front (Front: 36.8° vs. Side: 16.9° vs. Back: 23.6°). As the error was not corrected for front-back confusions this result is not surprising. The interaction of HRTF and source position did not turn out to be significant, meaning that there was gain from the use of individual HRTFs at specific locations only. The interaction of reproduction method and source position was

Table 2: F and p-Values taken from the ANOVA regarding unsigned horizontal error

HRTF:	$F(1, 10) = 2.50,$	$p > .05$
RepMeth:	$F(3, 30) = 15.85,$	$p < .001$
HRTF \times RepMeth:	$F(3, 30) = 3.70,$	$p < .05$
Pos:	$F(2, 20) = 2.57,$	$p > .05$
HRTF \times Pos:	$F < 1$	
RepMeth \times Pos:	$F(6, 60) = 3.48,$	$p < .05$
HRTF \times RepMeth \times Pos:	$F(6, 60) = 2.18,$	$p > .05$

significant, indicating no significant differences between reproduction methods for source positions to the side, but several significant differences between the static and dynamic reproduction for source positions in front and in back. This is explained by a reduction of the number of front-back confusions with dynamic playback.

The three-way interaction was not significant. All p and F-Values for the analysis can be found in Table 2.

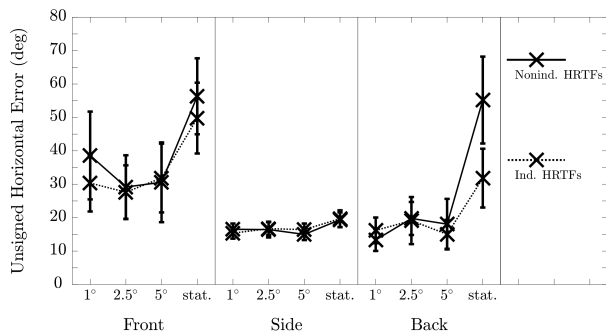


Figure 4: Unsigned horizontal error (in deg) as a function of HRTF, reproduction method and source position (HRTF \times RepMeth \times Pos). Error bars indicate standard errors.

5.2 Unsigned vertical error

Figure 5 shows unsigned vertical error as a function of the position, reproduction method and the HRTF. There was no significant effect in the unsigned vertical error for either the used HRTF, the reproduction method or the interaction between both. This indicates, that there was no gain in the elevation localization accuracy with the use of individual HRTFs regardless of reproduction method. The main effect of source position was significant, indicating significantly smaller unsigned vertical errors for trials where the source was positioned to the side than for trials where the source was positioned in front or in back (Front: 19.2° vs. Side: 10.3° vs. Back: 20.6°). The two-way interactions with source position as well as the three-way interaction did not turn out to be significant. All p and F-Values for the analysis can be found in Table 3.

5.3 Weighted front-back-confusions

Figure 6 shows weighted front-back confusion in percent as a function of the position, reproduction method and the HRTF. The arc-sine transformation of data which is appropriate for the data on proportions is applied before conducting the ANOVA. The ANOVA yielded a

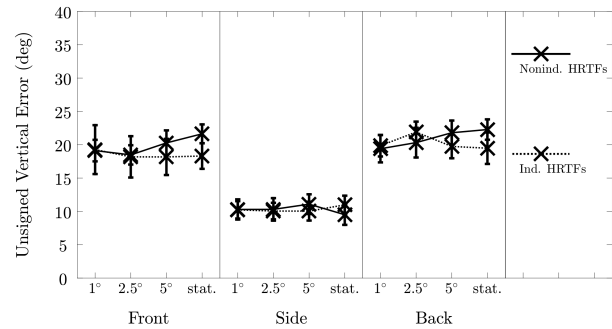


Figure 5: Unsigned vertical error (in deg) as a function of HRTF, reproduction method and source position (HRTF \times RepMeth \times Pos). Error bars indicate standard errors.

Table 3: F and p-Values taken from the ANOVA regarding unsigned vertical error

HRTF:	$F < 1$	
RepMeth:	$F < 1$	
HRTF \times RepMeth:	$F(3, 30) = 2.03,$	$p > .05$
Pos:	$F(2, 20) = 13.65,$	$p < .001$
HRTF \times Pos:	$F < 1$	
RepMeth \times Pos:	$F < 1$	
HRTF \times RepMeth \times Pos:	$F(6, 60) = 1.72,$	$p > .05$

significant effect on HRTFs in the weighted front-back-confusions, indicating less front-back-confusions when listening to individual HRTFs (individual HRTFs: 7.6 % vs. non-individual HRTFs: 10.7 %).

The main effect of reproduction method was also significant, resulting in a significantly higher front-back-confusion for a static reproduction than for all three dynamic reproduction methods (static: 16.5 % vs. dynamic (5°): 6.3 % vs. dynamic (2.5°): 6.8 % vs. dynamic (1°): 7.1 %). The interaction of HRTF and reproduction method was not significant. However, there was a non-significant trend towards a greater front-back-confusion difference between individual and non-individual HRTFs for the static reproduction than for the dynamic reproductions (static: 6.8 % vs. dynamic (5°): 1.7 % vs. dynamic (2.5°): 2.0 % vs. dynamic (1°): 1.6 %). This leads to the interpretation that dynamic reproduction will reduce the gain from individual HRTFs in terms of front-back confusion.

The main effect of source position was not significant. However, weighted front-back-confusions were smallest for trials where the source was positioned to the side and greatest for trials where the source was positioned in front (Front: 14.5 % vs. Side: 4.7 % vs. Back: 8.3 %). The two-way interaction of reproduction method and source position is significant, indicating higher front-back-confusions for a static reproduction in front and in back compared to the side. The two-way interaction of HRTF and source position as well as the three-way interaction did not turn out to be significant. All p and F-Values for the analysis can be found in Table 4.

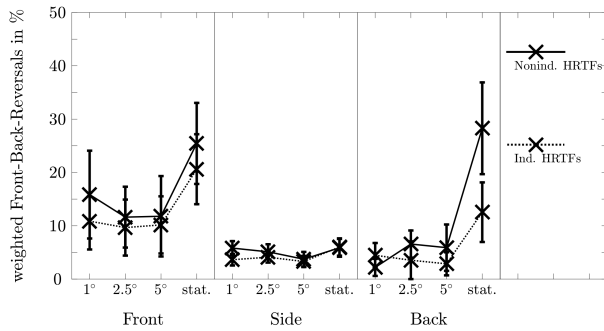


Figure 6: Weighted front-back-confusions (in %) as a function of HRTF, reproduction method and source position (**HRTF** \times **RepMeth** \times **Pos**). Error bars indicate standard errors.

Table 4: F and p-Values taken from the ANOVA regarding weighted front-back-confusions

HRTF:	$F(1, 10) = 6.98,$	$p < .05$
RepMeth:	$F(3, 30) = 11.68,$	$p < .001$
HRTF \times RepMeth:	$F(3, 30) = 1.31,$	$p > .05$
Pos:	$F(2, 20) = 1.31,$	$p > .05$
HRTF \times Pos:	$F(2, 20) = 1.02,$	$p > .05$
RepMeth \times Pos:	$F(6, 60) = 3.54,$	$p < .05$
HRTF \times RepMeth \times Pos:	$F(6, 60) = 2.17,$	$p > .05$

5.4 In-head-localization

Figure 6 shows in-head localization occurrence in percent as a function of the position, reproduction method and the HRTF. The arc-sine transformation of data which is appropriate for the data on proportions is applied before conducting the ANOVA. There was no significant effect on HRTFs in in-head-localization. However, there was a non-significant trend towards a lower rate of in-head-localization when listening to individual HRTFs (individual HRTFs: 1.5 % vs. non-individual HRTFs: 2.6 %).

The main effect of reproduction method was significant, resulting in a greater rate of in-head-localization for the static reproduction than for all three dynamic reproduction methods (static: 6.4 % vs. dynamic (5°): 0.5 % vs. dynamic (2.5°): 0.6 % vs. dynamic (1°): 0.7 %). The interaction of HRTF and reproduction method was not significant. However, there was a non-significant trend towards a greater in-head-localization difference between individual and non-individual HRTFs for the static reproduction than for the dynamic reproductions (static: 3.6 % vs. dynamic (5°): 0.5 % vs. dynamic (2.5°): 0.3 % vs. dynamic (1°): -0.2 %).

The main effect of source position was significant, indicating a significantly lower rate of in-head-localization for trials where the source was positioned to the side than for trials where the source was positioned in front (Front: 3.4 % vs. Side: 0.1 % vs. Back: 2.7 %). The two-way interaction of HRTF and source position did not turn out to be significant. The interaction of reproduction method and source position was significant, indicating more in-head-localization with the static repro-

Table 5: F and p-Values taken from the ANOVA regarding in-head-localization

HRTF:	$F(1, 10) = 1.92,$	$p > .05$
RepMeth:	$F(3, 30) = 8.12,$	$p < .05$
HRTF \times RepMeth:	$F(3, 30) = 1.68,$	$p > .05$
Pos:	$F(2, 20) = 6.59,$	$p < .05$
HRTF \times Pos:	$F < 1$	
RepMeth \times Pos:	$F(6, 60) = 5.36,$	$p < .05$
HRTF \times RepMeth \times Pos:	$F(6, 60) = 1.63,$	$p > .05$

duction for sources positioned in front and back than to the side.

The three-way interaction was not significant. All p and F-Values for the analysis can be found in Table 5.

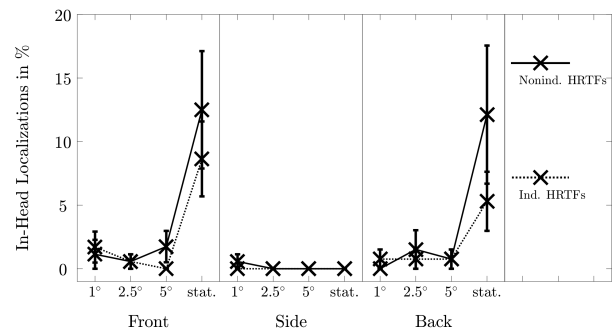


Figure 7: In-head-localization (in %) as a function of HRTF, reproduction method and source position (**HRTF** \times **RepMeth** \times **Pos**). Error bars indicate standard errors.

6 Discussion

In general, lateral sound sources yielded greater localization precision compared to those in front and back (e.g. [3]). This assumption was validated by the present investigation. There was a precision significantly higher by about 10° in elevation for lateral positions compared to frontal and backward positions and almost no and therefore a significantly lower rate of in-head-localization were found at the side. Front-back-confusion rates tended to be less pronounced for positions at the side and a tendency towards generally higher localization accuracy in azimuth for lateral positions was found in the present data.

One of the main investigations of this publication was the influence of the resolution of the HRTF. The hypothesis was that a higher resolution HRTF will increase the localization performance during dynamic reproduction. This hypothesis could not be confirmed. There were no identifiable differences between the different resolutions. Sandvad, conducting listening experiments with interpolated HRTFs with resolution of 1° up to 13°, reported similar findings: “Spatial resolution appears to have very little influence on subject performance. Large angular steps between different HRTF filters introduces audible discontinuities in the sound, but as with reduced update rate, participants seemed to be able to ignore the audible artifacts, and complete the

task with good precision [20].” These artifacts could also be observed: Several participants stated after the present experiment that some trials suffered from audible discontinuities in the sound. Due to the design of the experiment there was no chance to keep track of the specific affected trials. However, the authors themselves also perceived these audible discontinuities in sound during test runs and they could all be reduced to the dynamic reproduction with a resolution of 5° . Thus, a resolution smaller than 5° should be recommended, even though there is no increase in localization performance.

The results of this investigation could confirm that dynamic reproduction enables a more accurate localization compared to static reproduction. Dynamic reproduction was observed to reduce the unsigned horizontal error by 17° and reversal rate could be lowered by more than 9%. This was in accordance with Begault and colleagues who measured confusion rate differences up to 31%, stating: “Head tracking reduced reversals significantly (front-back- and back-front-confusions of the location of the stimuli across the inter-aural axis from a rate of 59% to 28%)[13].” Taking account of front-back-confusions in the unsigned horizontal error by extracting the affected trials, no effect is discernible (Unsigned horizontal error corrected by discarding all front-back-confusion trials: [RepMeth : $F < 1$]). Therefore, the unsigned horizontal error was strongly correlated with the front-back-confusion rate. As a consequence, dynamic reproduction effectively improved sound localization by eliminating, or, at least, reducing the confusion rate. The very same dynamic cues resulted advantageous also for reduction of in-head-localization rates. As a general result, dynamic reproduction seemed to have a great effect on reduction of typical problems arising with virtual acoustic environments such as front-back-confusion and in-head-localization.

One of the major aims of the presented experiment was to figure out, up to which extent individual HRTFs prove advantageous with respect to non-individual HRTFs. A significant main effect was observed only once: On average, front-back-confusions were reduced by about 3% when individual HRTFs were used. Letowski and Letowski, noted slightly higher reversal rates ranging around 7-11% [45]. These findings were not in line with those by Begault and colleagues who did not observe a significant main effect for the HRTF analyzing the front-back-confusion rate [13]. An impact on unsigned horizontal error due to individual HRTFs could be reasonably assumed, however, in the present experiment no significant main effect for the HRTF was observed. The absence of the main effect could be explained by the imbalance of static and dynamic reproduction methods (1:3) and that there is no effect for lateral positions. By reducing the variable of reproduction methods to two levels of static vs. dynamic (5°) and neglecting lateral positions a marginally significant main effect on HRTFs

could be found [HRTF : $F(1, 10) = 4.86, p = 0.052$]. Regarding the accuracy of localizing in elevation no difference between individual and non-individual binaural stimuli could be found. In contrast to that, Andéol and colleagues found considerable differences in the low elevation region [11], [12], substantiating the fact that individual binaural stimuli yield a better localization accuracy in elevation than non-individual.

The experiment was designed to observe interactions between different HRTFs and Reproduction Methods, which turned out to be significant only for the unsigned horizontal error. However, a tendency towards greater differences between individual and non-individual HRTFs in static than in dynamic reproduction could be observed for all dependent variables. These findings were in agreement with the observations by Begault and colleagues [13]. Large differences in static vs. dynamic reproduction appeared to mask potential effects of the HRTF. The same finding was confirmed by Kato and colleagues who report “that individual differences in HRTFs can be overcome by head motion” [18].

7 Conclusion

Aim of this investigation was to find an appropriate setting for the use of individual/non-individual HRTF and static/dynamic reproduction methods to ensure a most realistic auralization of acoustic scenes with binaural technique and headphone reproduction.

Dynamic reproduction of any resolution applied was confirmed fundamental for a reduction of undesired front-back-confusions and in-head-localization. With front-back-confusions dominating the results for the unsigned horizontal error, dynamic reproduction implicitly also had a great impact on the azimuth error. Individually measured HRTFs showed a smaller effect on localization accuracy compared to the influence of dynamic sound reproduction. They were mainly observed to reduce the front-back-confusion rate. In general, individual HRTFs proved indeed beneficial compared with non-individual HRTFs, however, as the benefit tended to be masked in the presence of the more pronounced effect of reproduction method, using non-individual HRTFs in a dynamic reproduction might be a less complex and elaborate way to successfully provide for a realistic auditory perception. The resolution of the used HRTFs did not produce a significant increase in localization performance, however audible artifacts with a resolution of (5°) were present which signified the need for a higher resolution.

Funding

The author is grateful for the provided financing by DFG (Deutsche Forschungsgemeinschaft, FE1168/1-1/2 and KO2045/11-1/2).

ORCID ID

Josefa Oberem  <https://orcid.org/0000-0001-7251-7453>

Jan-Gerrit Richter  <https://orcid.org/0000-0002-2811-5358>
Julia Seibold  <https://orcid.org/0000-0001-8367-8543>
Iring Koch  <https://orcid.org/0000-0001-9664-8220>
Janina Fels  <https://orcid.org/0000-0002-8694-7750>

References

- [1] J. Oberem, J.-G. Richter, D. Setzer, J. Seibold, I. Koch, and J. Fels, “Experiments on localization accuracy with non-individual and individual HRTFs comparing static and dynamic reproduction methods”, in *Fortschritte der Akustik: DAGA 2018*, 2018, pp. 702–705.
- [2] J.-G. Richter, *Fast Measurement of individual Head-related transfer functions: Dissertation*, ser. Aachener Beiträge zur Akustik. Berlin Germany: LOGOS VERLAG BERLIN, 2019, vol. 30.
- [3] F. Wightman and D. Kistler, “Headphone simulation of free-field listening. I: Stimulus synthesis”, *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 858–867, 1989. DOI: [10.1121/1.397557](https://doi.org/10.1121/1.397557).
- [4] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, “Localization using nonindividualized head-related transfer functions”, *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, 1993. DOI: [10.1121/1.407089](https://doi.org/10.1121/1.407089).
- [5] H. Wallach, “The role of head movements and vestibular and visual cues in sound localization”, *J. Exp. Psychol.*, vol. 27, pp. 339–368, 1940. DOI: [10.1037/h0054629](https://doi.org/10.1037/h0054629).
- [6] W. Thurlow, J. Mangels, and P. S. Runge, “Head movements during sound localization”, *J. Acoust. Soc. Am.*, vol. 42, no. 2, pp. 489–493, 1967. DOI: [10.1121/1.1910605](https://doi.org/10.1121/1.1910605).
- [7] I. Pollack and M. Rose, “Effect of head movement on the localization of sounds in the equatorial plane”, *Atten. Percept. Psycho.*, vol. 2, pp. 591–596, 1967. DOI: [10.3758/BF03210274](https://doi.org/10.3758/BF03210274).
- [8] D. Begault, E. Wenzel, A. Lee, and M. 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”, in *108th Audio Eng. Soc. Conv.*, New York, NY, US, 2000.
- [9] A. Bronkhorst, “Localization of real and virtual sound sources”, *J. Acoust. Soc. Am.*, vol. 98, no. 5, pp. 2542–2553, 1995. DOI: [10.1121/1.413219](https://doi.org/10.1121/1.413219).
- [10] J. Middlebrooks, “Individual differences in external-ear transfer functions reduced by scaling in frequency”, *J. Acoust. Soc. Am.*, vol. 106, pp. 1480–1492, 1999. DOI: [10.1121/1.427176](https://doi.org/10.1121/1.427176).
- [11] H. Møller, C. Jensen, D. Hammershøi, and M. Sørensen, “Using a typical human subject for binaural recording”, in *100th Audio Eng. Soc. Conv.*, 1996.
- [12] G. Andéol, S. Savel, and A. Guillaume, “Perceptual factors contribute more than acoustical factors to sound localization abilities with virtual sources”, *Frontiers in Neuroscience*, vol. 8, p. 451, 2014. DOI: [10.3389/fnins.2014.00451](https://doi.org/10.3389/fnins.2014.00451).
- [13] D. R. Begault, *3-D Sound for virtual reality and multimedia*. NASA, 2000.
- [14] D. Brungart and B. Simpson, “Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty”, *J. Acoust. Soc. Am.*, vol. 115, no. 1, pp. 301–310, 2004. DOI: [10.1121/1.1628683](https://doi.org/10.1121/1.1628683).
- [15] J. Pedersen and P. Minnaar, “Evaluation of a 3D-audio system with head tracking”, in *120th Audio Eng. Soc. Conv.*, New York, NY, US, 2006.
- [16] F. Wightman and D. Kistler, “Resolution of front-back ambiguity in spatial hearing by listener and source movement”, *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2841–2853, 1999. DOI: [10.1121/1.426899](https://doi.org/10.1121/1.426899).
- [17] Y. Iwaya, Y. Suzuki, and D. Kimura, “Effects of head movement on front-back error in sound localization”, *Acoust. Sci. Technol.*, vol. 24, no. 5, pp. 322–324, 2003. DOI: [10.1250/ast.24.322](https://doi.org/10.1250/ast.24.322).
- [18] M. Kato, H. Uematsu, M. Kashino, and T. Hirahara, “The effect of head motion on the accuracy of sound localization”, *Acoust. Sci. Technol.*, vol. 24, no. 5, pp. 315–317, 2003. DOI: [10.1250/ast.24.315](https://doi.org/10.1250/ast.24.315).
- [19] S. Perrett and W. Noble, “The contribution of head motion cues to localization of low-pass noise”, *Atten. Percept. Psycho.*, vol. 59, pp. 1018–1026, 1997. DOI: [10.3758/BF03205517](https://doi.org/10.3758/BF03205517).
- [20] J. Sandvad, “Dynamic aspects of auditory virtual environments”, in *100th Audio Eng. Soc. Conv.*, 1996.
- [21] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert, “An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization”, *Acoustica*, vol. 86, pp. 1046–1053, 2000.
- [22] J.-G. Richter, G. Behler, and J. Fels, “Evaluation of a fast hrtf measurement system”, in *140th Audio Eng. Soc. Conv.*, New York, NY, US, 2016, p. 9498.
- [23] R. Duraiswaini, D. N. Zotkin, and N. A. Gumerov, “Interpolation and range extrapolation of hrtfs [head related transfer functions]”, in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 2004, iv–45–iv–48 vol.4. DOI: [10.1109/ICASSP.2004.1326759](https://doi.org/10.1109/ICASSP.2004.1326759).
- [24] J.-G. Richter, M. Pollow, F. Wefers, and J. Fels, “Spherical harmonics based hrtf datasets: Implementation and evaluation for real-time auralization”, *Acta Acust. united Ac.*, vol. 100, no. 4, pp. 667–675, 2014. DOI: [10.3813/AAA.918746](https://doi.org/10.3813/AAA.918746).
- [25] B. Masiero and J. Fels, “Perceptually robust headphone equalization for binaural reproduction”, in *130th Audio Eng. Soc. Conv.*, New York, NY, US, 2011, p. 8388.
- [26] J. Oberem, B. Masiero, and J. Fels, “Experiments on authenticity and plausibility of binaural reproduction via headphones employing different recording methods”, *Appl. Acoust.*, vol. 114, pp. 71–78, 2016. DOI: [10.1016/j.apacoust.2016.07.009](https://doi.org/10.1016/j.apacoust.2016.07.009).
- [27] R. B. King and S. R. Oldfield, “The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays”, *Hu-*

- man Factors*, vol. 39, no. 2, pp. 287–295, 1997. DOI: [10.1518/001872097778543895](https://doi.org/10.1518/001872097778543895).
- [28] T. Ashby, R. Mason, and T. Brookes, “Head movements in three-dimensional localization”, in *134th Audio Eng. Soc. Conv.*, New York, NY, US, 2013.
- [29] B. Seeber, “Untersuchung der auditiven lokalisation mit einer lichtzeigermethode”, Dissertation, Technische Universität München, München, 2002.
- [30] G. Wersenyi, “Effect of emulated head-tracking for reducing localization errors in virtual audio simulation”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 2, pp. 247–252, 2009. DOI: [10.1109/TASL.2008.2006720](https://doi.org/10.1109/TASL.2008.2006720).
- [31] W. Thurlow and J. Mergener, “Effect of stimulus duration on localization of direction of noise stimuli”, *Journal of Speech and Hearing Research*, vol. 13, pp. 826–838, 1970. DOI: [10.1044/jshr.1304.826](https://doi.org/10.1044/jshr.1304.826).
- [32] Institute of Technical Acoustics, RWTH Aachen University, *Virtual acoustics - a real-time auralization framework for scientific research*, 2016.
- [33] F. Wefers, *Partitioned convolution algorithms for real-time auralization: Dissertation*, ser. Aachener Beiträge zur Akustik. Berlin Germany: Logos Verlag Berlin GmbH, 2015, vol. 20.
- [34] K. Hartung, “Messung, verifikation und analyse von außenohrübertragungsfunktionen”, in *Fortschritte der Akustik: DAGA 1995*, 1995, pp. 755–758.
- [35] D. Hammershøi and J. Sandvad, “Binaural auralization. simulating free field conditions by headphones”, in *94th Audio Eng. Soc. Conv.*, 1996.
- [36] J. Lewald, G. J. Dörrscheidt, and W. H. Ehrenstein, “Sound localization with eccentric head position”, *Behav. Brain Res.*, vol. 108, no. 02, pp. 105–125, 2000. DOI: [10.1016/S0166-4328\(99\)00141-2](https://doi.org/10.1016/S0166-4328(99)00141-2).
- [37] K. I. McAnally and R. L. Martin, “Sound localization with head movement: Implications for 3-D audio displays”, *Frontiers in Neuroscience*, vol. 8, p. 210, 2014. DOI: [10.3389/fnins.2014.00210](https://doi.org/10.3389/fnins.2014.00210).
- [38] J. C. Makous and J. C. Middlebrooks, “Two-dimensional sound localization by human listeners”, *J. Acoust. Soc. Am.*, vol. 87, no. 5, pp. 2188–2200, 1990. DOI: [10.1121/1.399186](https://doi.org/10.1121/1.399186).
- [39] M. A. Frens, A. J. V. Opstal, and R. F. V. D. Willigen, “Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements”, *Percept. Psychophys.*, vol. 57, no. 06, pp. 802–816, 1995. DOI: [10.3758/BF03206796](https://doi.org/10.3758/BF03206796).
- [40] R. H. Gilkey, M. D. Good, M. A. Ericson, J. Brinkman, and J. M. Stewart, “A pointing technique for rapidly collecting localization responses in auditory research”, *Behavior Research Methods, Instruments, & Computers*, vol. 27, no. 1, pp. 1–11, 1995.
- [41] J.-M. Pernaux, M. Emerit, and R. Nicol, “Perceptual evaluation of binaural sound synthesis: The problem of reporting localization judgments”, in *114th Audio Eng. Soc. Conv.*, New York, NY, US, 2003.
- [42] H. Bahu, T. Carpentier, M. Noisternig, and O. Warusfel, “Comparison of different egocentric pointing methods for 3d sound localization experiments”, *Acta Acust. united Ac.*, vol. 102, no. 1, pp. 107–118, 2016. DOI: [10.3813/AAA.918928](https://doi.org/10.3813/AAA.918928).
- [43] J. Blauert, *Spatial Hearing - The psychophysics of human sound localization*, 2nd. Cambridge MA: MIT Press, 1997.
- [44] H. Møller, D. Hammershøi, J. Hundebøll, and C. Jensen, “Transfer characteristics of headphones: Measurements on 40 human subjects”, in *92nd Audio Eng. Soc. Conv.*, 1992.
- [45] T. R. Letowski and S. T. Letowski, *Auditory spatial perception: Auditory localization*.