



HAL
open science

Perceptual evaluation of an ambisonic auralization system of measured 3D acoustics

Simon Fargeot, Adrien Vidal, Mitsuko Aramaki, Richard Kronland-Martinet

► **To cite this version:**

Simon Fargeot, Adrien Vidal, Mitsuko Aramaki, Richard Kronland-Martinet. Perceptual evaluation of an ambisonic auralization system of measured 3D acoustics. *Acta Acustica*, 2023, 7, pp.56. 10.1051/aacus/2023052 . hal-04360874

HAL Id: hal-04360874

<https://hal.science/hal-04360874v1>

Submitted on 22 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Perceptual evaluation of an ambisonic auralization system of measured 3D acoustics

Simon Fargeot^{*} , Adrien Vidal, Mitsuko Aramaki , and Richard Kronland-Martinet 

Aix Marseille Univ, CNRS, PRISM, Marseille, France

Received 5 January 2023, Accepted 29 September 2023

Abstract – This paper presents a perceptual experiment aimed at assessing the spatial quality of acoustic environment rendering using a 4th order ambisonic auralization system. A novel test protocol is developed for this purpose, based on comparing the perceived spatial attributes of sound sources in both real (in-situ) and virtual listening conditions (loudspeaker-based ambisonic auralization of measured SRIRs). The perceptual evaluation is conducted using a specific reporting method combined with a virtual reality interface, enabling simultaneous assessment of perceived distance, angular position, and apparent width of sound sources. The test is conducted in three “office like” rooms, varying in reverberation properties and size. The results highlight differences in spatial perception between (a) real rooms and (b) their reproduction through the auralization system. Overall, localization performance is worse in auralized conditions than in real conditions, as evidenced by a clear increase in localization errors in azimuth and elevation, along with an increase in reported source width. This study also reveals that the spatial accuracy of the auralization depends on the rooms being auralized.

Keywords: Auralization, Spatial hearing, Room acoustics, VR

1 Introduction

It has been shown that the acoustic environment plays a decisive role on perception of spatial attributes of sound sources, in particular, on the perception of distance [1, 2], apparent width [3–5] and to a lesser extent on the perception of their angular position [6–8]. However, the grouped study of the influence of acoustics on these three major spatial attributes of the source has been scarcely explored, due to the lack of a suitable experimental protocol.

In general, perceptual evaluation of spatial acoustic attributes is commonly conducted through a virtual rendering process, commonly referred to as auralization. This approach is often preferred to avoid the tediousness of in-situ experiments that would require participants to physically move between different locations. A straightforward method for auralization utilizes the Higher Order Ambisonics (HOA) codec, which involves measuring Spatial Room Impulse Responses (SRIRs) using microphone arrays. The measured acoustic spaces can then be rendered through headphones using various ambisonic to binaural conversion techniques (examples can be found in [9–12]). However binaural rendering presents several drawbacks. Firstly, it relies on Head-Related Transfer Functions (HRTF) which need to be individualized for a better listening experience [13, 14]. Additionally, binaural rendering

may lead to internalization issues, where sound sources are perceived inside the head, and front-back confusions. This can be problematic when investigating source localization performances and distance perception. SRIR rendering can also be achieved through 3D loudspeaker array systems, using HOA formalism. While several studies have reported satisfactory spatial quality in HOA rendering [15–17], its suitability for studying the influence of acoustics on the perception of spatial attributes of sound sources has yet to be investigated.

In this paper the two following issues are thus jointly addressed:

1. A need for new methodological tools to investigate spatial perception of sound sources in reverberant environments, allowing a more global description of sound source spatial attributes. To do so, we first present a novel test protocol based on a VR interface allowing the simultaneous description of three major spatial attributes of sources, namely the angular position, the distance and the apparent width.
2. A perceptual characterization of potential spatial degradations induced by a loudspeaker based 4th order ambisonics auralization of measured SRIRs. This is addressed by comparing source localization performances in several acoustic environments (i.e. rooms) under both real and auralized conditions, by means of the previously introduced VR reporting method.

^{*}Corresponding author: fargeot@prism.cnrs.fr

1.1 Reporting methods for localization studies

Asking a listener to indicate the position of a sound source may seem trivial. However, there are many ways to perform this task. A first method consists in describing orally or through an interface the position of the source, by indicating the two angles of the spherical coordinates, azimuth and elevation. This method, known as the absolute judgment method [18], has the advantage of being able to indicate the position of sources located behind the listener but requires extensive training. Manson et al. also suggest that verbal elicitation to indicate the position of an object is subject to greater variability in responses compared to non-verbal elicitation methods [19].

Among the non-verbal elicitation methods, pointing methods consist in pointing with the finger, the head or by manipulating objects (e.g. a stick, a gun) [20–22], in the perceived direction of the sound source. This is a widely used method, yet it can induce biases depending on the pointer used. Finger pointing, for instance, may introduce biases based on the hand used [23], while pointing with the head or nose induces a bias for high elevation directions due to the physiological limitations of head orientation [24]. These methods are suitable for closed-loop localization tasks (i.e. localization task performed during sound stimulus presentation) and when the listener is allowed to turn towards the source. On the other hand, they may be less suitable for brief stimuli or source positions far from the frontal position, as they require the listener to memorize the stimulus position and subsequently turn towards the memorized direction.

Other egocentric pointing methods have been developed to overcome some of these biases. They involve indirect pointing toward the source by manipulating a visual or acoustic pointer. Methods using a visual pointer (e.g. laser pointer, positioned on the head or in the hand of the listener), have been used in several studies [25–27]. They consist of projecting the perceived position of the source onto a surface, mostly surrounding the listener, using the visual pointer. This technique thus proposes a visual deferral of auditory perception, which, for sources located in the listener’s field of view, proves to be a method with a low bias [25]. However, it does not allow reporting the position of sources located outside one’s field of view and is therefore inappropriate for experiments where the subject is not allowed to move his head.

Methods based on the manipulation of an acoustic pointer aim to avoid mixing two modalities in the localization task [28–30]. They consist in using a speaker mounted on an arm, whose position can be controlled. User is therefore asked to match the position of the acoustic pointer with that of the source to be located. This method requires a rather heavy instrumentation and is most of the time used for the localization of sources located in the azimuthal plane. On the other hand it has the advantage of allowing the subject to report their judgment for source positions at 360° without needing to turn around.

Finally, exocentric methods consist in reporting perceived position of sound sources on an object or graphical

interface representing auditory space [31–33]. One of the most cited examples is the GELP (God’s Eye Localization Pointing) method which consists in plotting the position of the source on a 20 cm diameter sphere placed in front of the listener [31]. Another example proposed by Hassager et al. allows participants to report both perceived source position and width on a touch-screen representing auditory space. They achieve this by placing a circle on the screen and adjusting its position and radius [34]. In a study investigating spatial hearing with incongruent visual or auditory room cues, Gil-Carvajal et al. introduced a method for assessing source position, width and distance. This was achieved using three subjective scales, displayed on a 2D representation of auditory space [35]. These methods offer the advantage of quick reporting of sound source spatial attributes. In contrast, they are generally less precise than egocentric pointing methods. When comparing the GELP method to finger and head pointing methods, Bahu et al. revealed a systematically higher localization error induced by the exocentric method [22].

In summary, there are many methods for reporting the perceived angular position of sound sources: verbal or non-verbal, egocentric or exocentric, with visual or acoustic pointers. The methods that induce the least bias are those that use a visual pointer (e.g. a laser pointer), which are based on a multi-sensory integration of space (i.e. visual confirmation of the perception of an auditory event). These methods are well adapted to the localization of sources located in the listener’s field of view or when the listener is allowed to orient themselves in the sound scene. However, they are not suitable for blindfolded localization tests and for sources located in the whole auditory sphere, when the listener must remain static.

1.2 Use of virtual reality

Virtual reality (VR) technologies have opened up new avenues for exploring perception [36, 37]. With respect to spatial auditory perception, innovative interfaces for source localization in VR have emerged and proven to be effective. For instance, Majdak et al. studied source localization via a virtual reality interface [26]. Participants were asked to point towards the source using their head or a hand-held controller, with or without visual feedback from the virtual environment. The study reveals enhanced localization performance when visual feedback was available, regardless of the pointing method used (head or hand), compared to blind localization. VR also facilitates the implementation of new protocols for giving a more global description of the perceived spatial image of sound scenes, as proposed by [38]. For this study, a virtual reality interface was employed to assess the spatial degradations resulting from a spatial remix of sound sources obtained through a source separation process. The localization task consisted in surrounding the sources composing the sound scene, by drawing on the surface of a virtual sphere surrounding the listener, using a visual pointer. This method enabled participants to define perceived source position in space as well as its perceived shape and size. It has been shown

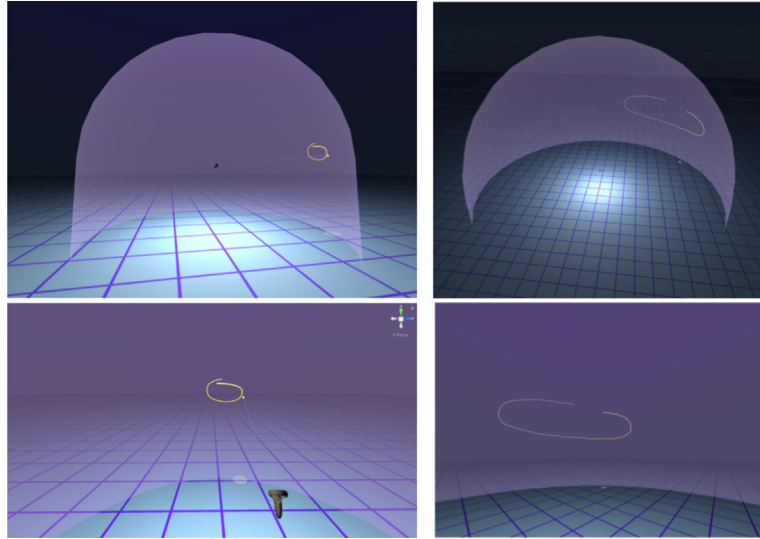


Figure 1. VR interface for reporting spatial attributes of sound sources (i.e. angular position, distance and apparent width). Left: reporting of a close source, of small size. Right: reporting of a distant source, perceived as large. Top: spectator point of view (from outside the sphere). Bottom: user point of view.

to effectively highlight spatial degradations that are characteristic of a remix involving separated sources (e.g. phantom sources, source widening or spatial instability).

On the other hand, wearing a head-mounted device, such as a VR headset, can have an impact on the HRTF of the listener wearing it [39]. Recently, Huisman et al. investigated the localization performance with a VR headset for 2D ambisonic sound reproduction at different orders, with and without visual feedback [40]. In their study, the localization task was performed in an open-loop fashion (i.e. performing localization task after sound stimulus presentation). The results showed that wearing a headset introduced a bias in the perceived position of the sources, with an average lateralization bias of approximately 2° for sources located on either side of the frontal plane. In contrast, no effect was observed for sources around the frontal position. These findings suggest that using a virtual reality (VR) headset in closed-loop localization tasks (where the listener is free to rotate their head during stimulus presentation) could be beneficial in reducing the bias observed in the study by Huisman et al.

1.3 Proposed interface

In order to study potential effects of 3D ambisonics auralization system on spatial attributes of sound sources, namely angular position, distance and apparent width, a novel reporting method utilizing a VR interface has been designed. The proposed interface is based on the protocol from Fargeot et al. study [38]. In addition to reporting source angular position and size, the new interface enables participants to report the perceived source listener distance. It is presented as follows. The user is immersed in a virtual space illustrated in Figure 1, consisting of an infinite tiled floor (tiled with one-meter squares) and a semi-transparent half-sphere surrounding the listener. Using a controller held

in the right hand, the user manipulates a pointer like he would manipulate a laser pointer. A beam of light is emitted from the controller toward the pointed direction, and the intersection between this beam and the surface of the half-sphere is marked by a small shining ball. The localization task is decomposed into two steps. First, users are instructed to report the perceived distance of the sound source. For that, they have to adjust the radius of the sphere that surrounds them with the controller's joystick. It is important to note that participants are informed about the 1-meter length of the tiles comprising the floor in the VR environment. Once the distance is set and validated, they are asked to report the perceived position and width of the source by drawing on the surface of the half-sphere a shape within which they believe the source is located (visual examples of the localization task can be seen in Fig. 1). The wider the source is perceived or the more difficult it is to locate, the wider the surrounding area should be. In practice, shapes are drawn by manipulating the pointer while holding down the controller's trigger.

The present interface has been developed using the Unity game engine and compiled as an Android application, allowing its use on a mobile headset (here, Oculus Quest). The interest of using a mobile headset is to create a simple and light workflow and thus be able to perform the localization test in different real acoustic environments, without the need for a third-party computer.

2 Methods

An experiment has been designed to highlight potential perceived degradation of spatial attributes of sound sources caused by 4th order ambisonic auralization system. This is achieved by comparing localization performances of sound sources in three different acoustic environments, both in

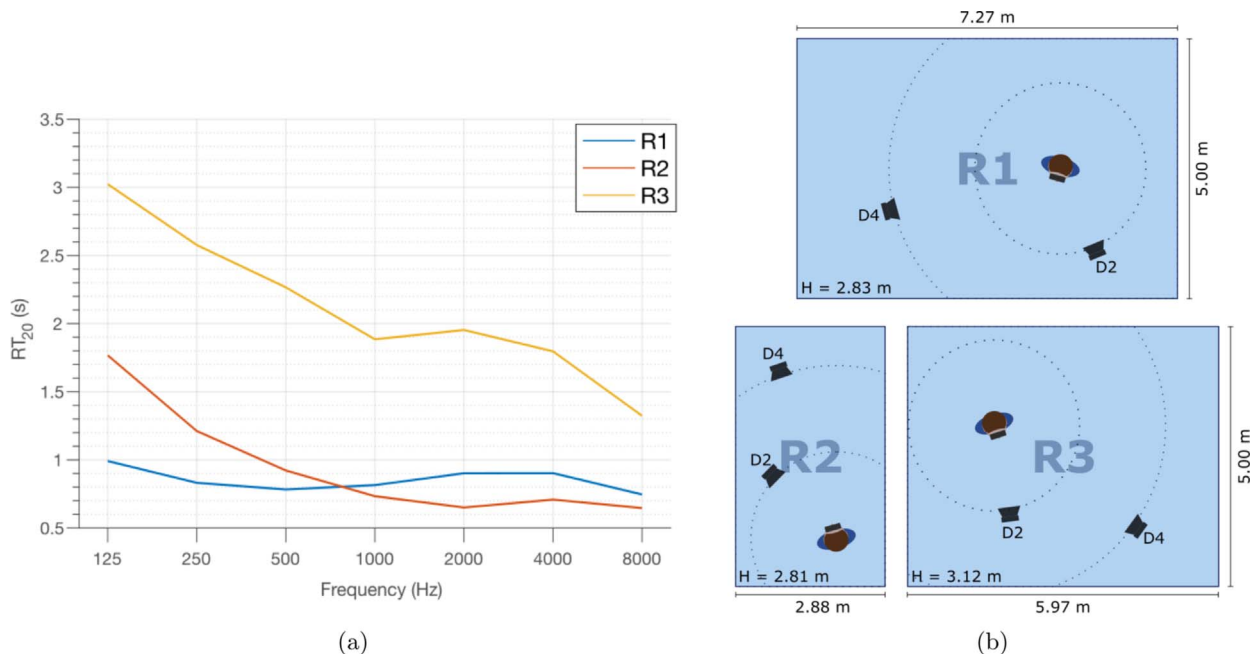


Figure 2. Acoustical and dimensional properties of the rooms under study, denoted R1, R2 and R3. All rooms are located in the same buildings and are “office” type rooms. R1 and R2 have ceiling acoustic treatment, whereas R3 has not. This explains the particularly high value of RT_{20} for R3. (a) Reverberation times by octave bands of the three rooms, calculated on omnidirectional component of measured SRIRs, using ITAToolbox [41]. (b) Dimensions and spatial configurations of the sources and listening positions in the three rooms. H value indicates room height. D2 and D4 represents the two loudspeakers, respectively placed at 2 and 4 m from the listener. In each room, the two source positions remain fixed and identical across all participants. Critical distances of the three rooms at 1 kHz were respectively 0.64 m, 0.43 m and 0.37 m. The calculation of critical distances is based on an omnidirectional source.

real (in-situ) and auralized listening conditions. Participants were asked to assess their perception of the sources in terms of position (azimuth, elevation), width and distance through the VR interface previously described, in both listening conditions.

2.1 Acoustic conditions

Three real acoustic environments were selected for this experiment, based on three criteria, namely acoustical diversity, size diversity and geographical proximity. Three empty rooms noted R1, R2 and R3 were chosen within the same building at PRISM laboratory, for the sake of simplicity and comfort for the participants. These rooms were also located near the ambisonic auralization system so that the participants could carry out both sessions (RE and V conditions) on the same day. Figure 2a shows the reverberation times per octave band of the three selected acoustic environments. In order to examine the impact of source/room balance (i.e direct-to-reverberant energy ratio, denoted DRR) on localization performances, two loudspeakers (Genelec 8020C) were positioned in each room at distances of 2 and 4 m from the listening position, noted as D2 and D4, respectively. The distance D4 was chosen as the largest distance possible to test within the smallest room. The angular positions of the sources were chosen arbitrarily to avoid expectations on the spatial configuration in the different rooms. Figure 2b provides an overview of room geometric properties and loudspeakers placement.

2.2 Auralization framework

The auralization is performed through a Spatial Room Impulse Responses (SRIRs) measurement and playback framework. Acoustic measurements consisted in collecting the SRIRs of the 6 configurations: 3 rooms \times 2 distances. The signal used for the measurements was an exponential swept-sine with a duration of 10 s and a frequency range of [20 Hz; 22 kHz] as proposed by [42]. Measurements were conducted using a 32-microphones spherical array (mh-acoustics Eigenmike™, also denoted em32). Raw recordings were encoded to 4th-order HOA format with ACN channel ordering and SN3D normalization scheme, using EigenUnits encoder plug-in from mh-acoustics. Details about the encoding framework can be found in the EigenUnits datasheet [43]. Playback is performed through a 42-loudspeakers (Genelec 8020C) spherical array distributed on a 3.8 m diameter geodesic structure, represented in Figure 3. The playback system is located inside a semi-anechoic chamber ($9 \times 4 \times 4$ m³, cutoff frequency of 80 Hz). The loudspeakers of the playback system were individually time-synchronized and level equalized. Moreover, the overall frequency response was equalized using a minimal phase Finite Impulse Response filter in such a way the frequency response of the measured stimuli was the same (at ± 3 dB) than their rendering through the virtual system from 80 Hz to 7 kHz. Sounds were processed using Max/MSP and the spat5 library [44]. The choice for decoding method and optimization scheme may depend,



Figure 3. Spatialization system at PRISM Laboratory (Marseille, France). It consists of a spherical array of 42 loudspeakers (Genelec 8020C) with a diameter of 3.8 m, situated in a semi-anechoic chamber. Auralization is performed by a 4th order HOA rendering (decoder: energy-preserving with basic optimization) of SRIRs measured with mh-acoustics em32.

among other matters, on speaker array layout and acceptable sweet-spot width. For regular and dense speaker layouts, as the one used in this experiment, decoding method doesn't seem to make a great difference on rendering quality. Energy-preserving method proposed by [45] is known to avoid sudden variations in the decoded energy while preserving apparent width of virtual sources [46]. For that reason, energy-preserving decoding method was chosen here. As far as decoding optimization is concerned, informal listening sessions showed no great difference between *basic* and *maxRE* optimization while *in-phase* optimization had critical impact on source timbre. Therefore, *basic* optimization was chosen here.

2.3 Stimuli

Three sound excerpts were considered: a speech stimulus (duration: 3 s), identified as “Speech” and a classical guitar extract (duration: 1 min) identified as “Guitar”. The third stimulus was a train of white noise bursts (duration: 1 s), identified as “Burst” which is broadly used in source localization experiments [47]. For the RE condition, stimuli were directly played through the loudspeakers in the room, and for the V condition, stimuli were convolved with the measured SRIRs as described in Section 2.2.

2.4 Participants

A total of 21 participants (15 males, 6 females) agreed to participate in this study. Their average age was 28.6 years (SE = 6.5 years), and they reported no hearing problems. All participants were unaware of the spatial configuration of the acoustic sources and environments, both in the virtual and real conditions. Prior to the start of the

experiment, participants were provided with an instruction sheet explaining the nature and procedure of the study, and they signed a consent form.

2.5 Procedure

The experiment is composed of three sessions: a short familiarization session and two test sessions defined by the two listening conditions, real (RE) and virtual (V). For every trial in all three sessions, the task to be carried out is a source localization task as depicted in Section 1.3. As a reminder, it consists of two steps: (1) report the perceived distance of the source, by adjusting the radius of a half-sphere surrounding the listener, (2) surround as precisely as possible the area in which the source is heard, using the pointer. If the source is large or difficult to locate, participants are encouraged to circle a large area that should circumscribe the source. For both test sessions, the localization task is open-loop meaning that sound excerpt is played in a loop, until localization task is completed and that participants are free to rotate their head during stimulus presentation. To avoid session order bias, one half of the participants started with the RE session, while the other half started with the V session. Within session V, all conditions were presented randomly. In contrast, the RE session was organized into 3 sub-sessions related to the three acoustic environments R1, R2, R3. The order of the conditions within each sub-session as well as the order of the sub-sessions were determined randomly for each participant.

The familiarization session is an informal session, carried out at the beginning of the experiment, where the participant is brought to discover the interface in order to become familiar with the task to be carried out and the various controls proposed by the interface. For that purpose, the

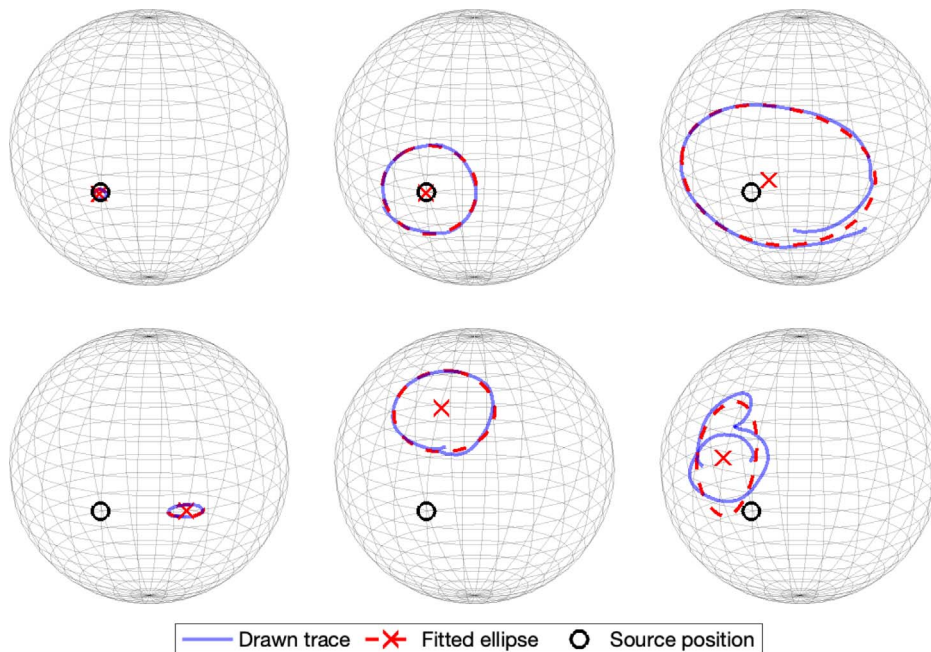


Figure 4. Examples of participants’ traces (blue line) for different conditions and approximation of the plots by an ellipse (red dotted line). The angular position of the source is marked by a black circle. The center of the ellipse is marked by a red cross. These examples illustrate the diversity of plots in terms of angular error and reported width as well as the relevance of approximating plots with an ellipse. The example at the bottom right illustrates the limits of the approximation by an ellipse.

participant is positioned on a chair, in a room which is not part of the acoustic environments under study and is equipped with the HMD displaying the test interface. The experimenter move to an arbitrary position in the room and the participant must then locate the experimenter’s voice and validate his or her response. The trial is repeated until the participant feels comfortable with the task.

For the real listening session, the participants are led, blindfolded, into each of the test rooms and are accompanied to a swivel chair placed at the listening position. Once comfortably installed on the chair, the participant is equipped with the VR headset, on which the application presenting the interface of the experiment has been previously launched. At this point, the participants are allowed to open their eyes and are asked to adjust the position of the headset on the head for a comfortable wearing and so that the presented interface is visible and clear. For each trial, a sound excerpt is played on one of the two loudspeakers present in the room. Localization task is performed sequentially for all 6 possible combinations “source position \times sound excerpt” within the room and with no repetition. The order of presentation for these six conditions is randomly determined for each participant. At the end of the session in the first room, the subject is escorted outside, blindfolded. The operation is repeated in the other two rooms selected for the test.

For the virtual listening session, the participants are invited to sit on a swivel chair placed in the middle of the auralization system. The swivel chair is adjusted individually for each participant to ensure that their head is positioned within the sweet-spot of the rendering system.

Once comfortably seated, participants are asked to put on the VR headset and adjust it to their liking. Finally, the testing phase consists of performing the localization task for the 18 auralized conditions, with no repetition. To prevent any perceptual expectation regarding the position of the source, a rotation in the azimuthal plane is applied to the sound scene so that the source to be localized is presented with a random azimuthal incidence, ranging from -90° to 90° .

2.6 Data processing

For each trial, the Cartesian coordinates of the points of all the drawn traces were collected. Some typical examples of traces are shown in Figure 4. We then computed the perceived angular position in azimuth θ_p and elevation ϕ_p , the distance R_p and the apparent source width as follows. Each plot is firstly fitted by an ellipse in the azimuth – elevation plane, using linear least squares ellipse fitting from Matlab *fitellipse* function by Richard Brown [48]. The angular position is given by the polar coordinates of the center of the ellipse. The apparent width of the ellipse is characterized by an equivalent solid angle S_{eq} given by the following formula:

$$S_{eq} = 4\pi \times \sin\left(\frac{A}{2}\right) \times \sin\left(\frac{B}{2}\right),$$

with A and B , respectively the small and large radii of the ellipse, expressed in radian. The unit for solid angles is the steradian, noted (sr). The perceived distance R_p is thus

Table 1. Summary of the variables analyzed for the experiment. The log trans. column indicates the variables for which the data have been logarithmically transformed to meet the assumption of normal distribution, which is necessary for conducting statistical analysis using linear models.

Quantity	Variable	Expression	Log trans.
Azimuth error (\pm)	ε_θ ($^\circ$)	$\theta_p - \theta_0$	
Azimuth error (abs.)	$ \varepsilon_\theta $ ($^\circ$)	$ \theta_p - \theta_0 $	✓
Elevation error (\pm)	ε_ϕ ($^\circ$)	$\phi_p - \phi_0$	
Elevation error (abs.)	$ \varepsilon_\phi $ ($^\circ$)	$ \phi_p - \phi_0 $	✓
Perceived distance	R_p (m)	$\ \text{OP}_1\ $	✓
Apparent source width	S_{eq} (sr)	$4\pi \times \sin\left(\frac{\alpha}{2}\right) \times \sin\left(\frac{\beta}{2}\right)$	✓

given by the radius of the sphere, i.e. the norm of the vector between the listening position (at the center of the sphere) and the position of the first point of the trace. From the angular positions reported by the participants (azimuth θ_p , elevation ϕ_p), and knowing the theoretical angular positions of the sources in the rooms (θ_0 ϕ_0), the errors in azimuth $\varepsilon_\theta = \theta_p - \theta_0$ and in elevation $\varepsilon_\phi = \phi_p - \phi_0$ are calculated. The absolute values of these errors $|\varepsilon_\theta|$ and $|\varepsilon_\phi|$ are also calculated. The signed errors are indicative of a dissymmetry in the angular perception of the sources while the absolute error represents the average amount of error made during the localization task, independently of the sign of the error.

2.7 Statistical analysis

The statistical analyses were conducted on the both signed and absolute errors in azimuth (ε_θ and $|\varepsilon_\theta|$) and elevation (ε_ϕ and $|\varepsilon_\phi|$) as well as on the perceived distance (R_p) and apparent source width (S_{eq}). Table 1 summarizes the variables considered in the statistical analysis. In order to satisfy the normality assumption required for the use of linear models, the data for some of these variables had to undergo log transformation. The data were analysed by mixed linear model, considering three fixed effects and two random effects. The listening condition (RE, V), the acoustic environment (R1, R2, R3), the distance from the source (D2, D4), along with their interactions were treated as fixed effects. The sound stimulus (Speech, Guitar, Burst) as well as the participants were treated as random effects. A repeated measures ANOVA was performed to measure the significance of the fixed effects. The magnitude of the fixed effects was estimated using η^2 , which measures the amount of explained variance. Post-hoc analyses were performed with a Tukey adjustment. Outliers were defined as responses with an absolute error in azimuth greater than 45° or absolute errors in elevation greater than 60° and were removed from the analysis of angular errors data. Thus, 1.2% of responses were excluded from the azimuthal data analysis, and 3.6% were excluded from the elevation data analysis. These analyses were carried out using the ‘‘lmerTest’’ library in the R software [49]. The use of this analysis methodology was inspired by a statistical analysis conducted by Ahrens et al., on sound source localization data under different audio-visual conditions [50].

3 Results

Note that the signed azimuth localization error has been excluded from the analysis, as it did not yield clear and interpretable results relevant to the current research question. Results pertaining to the other variables are presented below.

3.1 Effects on azimuth localization

As shown in Figure 5, the absolute azimuth localization error $|\varepsilon_\theta|$ is significantly affected by listening condition COND: $F(1, 706.59) = 173.0953$, $p < 0.0001$, $\eta^2 = 0.20$ (Fig. 5a), acoustic environment ROOM: $F(2, 706.69) = 8.5683$, $p = 0.0002$, $\eta^2 = 0.02$ (Fig. 5b). More precisely, the absolute error of localization in azimuth was on average 1.9° in RE condition while it was 5.9° in V condition, leading to an average increase of the error of 4° . The interaction COND \times ROOM is also significant: $F(2, 706.50) = 7.4978$, $p = 0.0006$, $\eta^2 = 0.02$ (Fig. 5c), showing that in RE condition, the absolute error in azimuth is not dependent on acoustics while in the V condition, the absolute error is significantly lower for room R1 than in the other two rooms. Conversely, the distance from the source also seems to have a significant effect on the localization: $F(1, 706.63) = 7.5523$, $p = 0.0061$, $\eta^2 = 0.01$ (Fig. 5d). The significant interaction between listening condition and source distance COND \times DIST: $F(1, 706.41) = 9.7136$, $p = 0.0019$, $\eta^2 < 0.01$ (Fig. 5e), indicates that the effect of source distance on $|\varepsilon_\theta|$ is only observed in RE condition. Indeed, in real conditions, participants on average committed a higher localization error for sources at 4 m (2.4°), than for those placed at 2 m (1.5°), while no effect of distance is observed in auralized conditions.

3.2 Effects on elevation localization

A significant difference between the listening conditions is observed in the signed elevation error ε_ϕ COND: $F(1, 695.58) = 356.8744$, $p < 0.0001$, $\eta^2 = 0.34$, as well as for the absolute error $|\varepsilon_\phi|$ COND: $F(1, 695.81) = 359.5922$, $p < 0.0001$, $\eta^2 = 0.34$. Figure 6 illustrates these results and shows in Figure 6a a large upward skewness of the signed error, with sources perceived at an elevation of 15.7° against 0.7° for the real conditions. The signed error ε_ϕ is very slightly affected by the acoustic environment

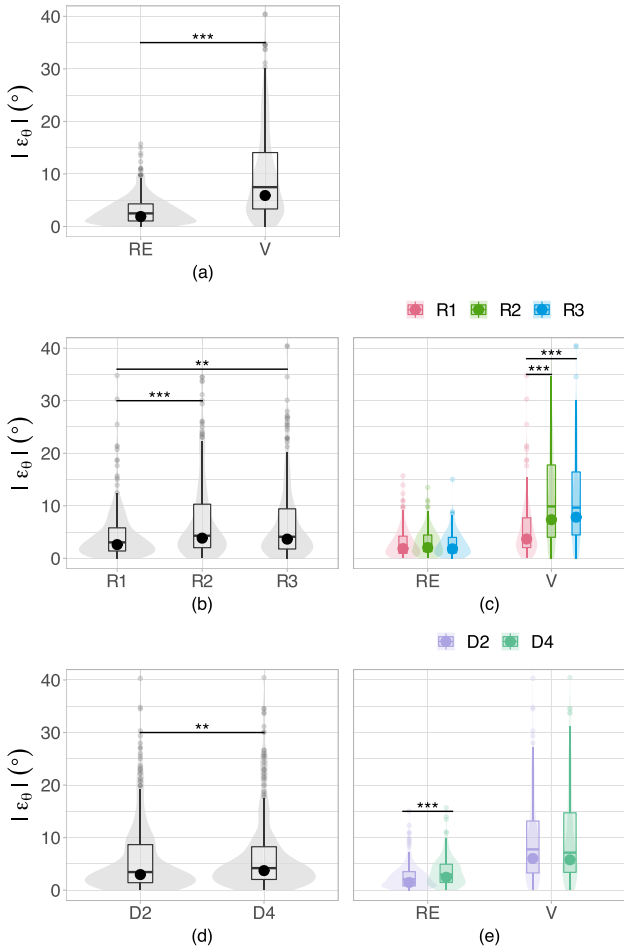


Figure 5. Absolute azimuth localization error $|\varepsilon_\theta|$, as a function of (a) the listening condition COND (RE: real listening, V: HOA4 auralization), (b) the acoustic environment ROOM (R1, R2, R3), (c) the interaction COND \times ROOM between listening condition on x -axis and acoustic environment in color, (d) the source distance DIST (D2 and D4), (e) the interaction COND \times DIST between the listening condition on the x -axis and the source distance in color. Post-hoc results are indicated by stars: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

ROOM: $F(2, 695.07) = 4.1102$, $p = 0.0168$, $\eta^2 = 0.01$. Post-hoc tests reveal that sources are perceived slightly but significantly higher in room R2 than in room R1 (Fig. 6b).

A significant difference between real and virtual listening conditions is also observed in the absolute elevation localization error $|\varepsilon_\phi|$ COND: $F(1, 695.81) = 359.5922$, $p < 0.0001$, $\eta^2 = 0.34$. Figure 6c shows a higher absolute error in the virtual (8.5°) than in the real condition (1.9°).

3.3 Effects on perceived distance

The perceived distance of the sources R_p is mainly dependent on their actual distance DIST: $F(1, 722) = 126.6416$, $p < 0.0001$, $\eta^2 = 0.15$. Not surprisingly, the sources positioned at 2 m (D2) were perceived closer than those positioned at 4 m (D4), with an average reported distance of 4.06 m for the sources at 2 m against 5.71 m for

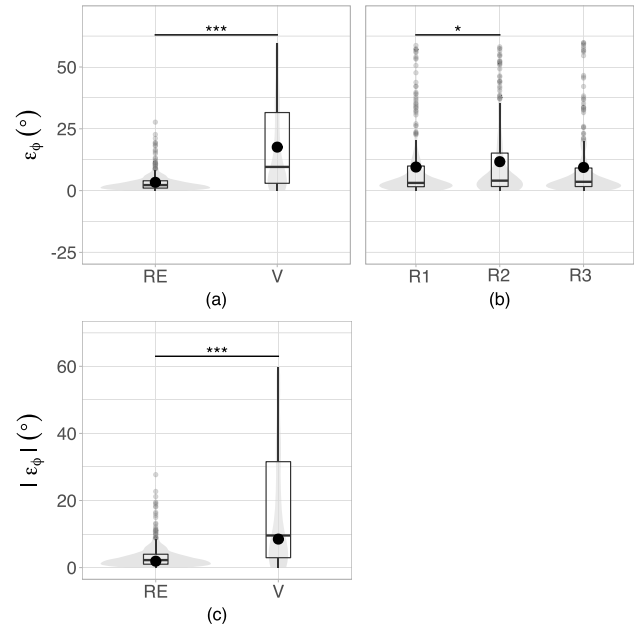


Figure 6. Elevation localization errors ε_ϕ as a function of (a) the listening condition COND (RE: real listening, V: HOA4 auralization), (b) the acoustic environment ROOM (R1, R2, R3). (c) Absolute elevation localization errors $|\varepsilon_\phi|$ as a function of the listening condition COND. Post-hoc results are indicated by stars: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

those at 4 m (Fig. 7d). The perceived distance is also affected, to a lesser extent, by the acoustics environment in which the sources are located, ROOM: $F(2, 722) = 15.8712$, $p < 0.0001$, $\eta^2 = 0.04$ (Fig. 7a), with larger perceived distance in room R3 than in the two other ones. In addition, the sources in room R2 were on average perceived slightly closer than those in room R1. Although the listening condition factor does not appear to have an overall influence on distance judgement, it significantly interacts with the room COND \times ROOM: $F(2, 722) = 11.5848$, $p < 0.0001$, $\eta^2 = 0.03$, as illustrated in Figures 7b and 7c. The results of the post-hoc tests reveal that in the RE condition, the room did not have a significant influence on the distance judgement while in the V condition, the sources were on average perceived further away in the most reverberant room (R3) than in the other two rooms ($p < 0.0001$) and slightly less far away in room R2 than in room R1 ($p = 0.02$) (Fig. 7c). Post-hoc tests also reveal that the listening condition had no effect on the evaluation of the distance in room R1. On the other hand, the sources in room R2 were perceived closer in the auralized condition than in the real condition ($p = 0.0073$) while those in room R3 were perceived further away in the auralized condition than in the real condition ($p = 0.0001$) Figure 7b. The analysis finally highlight the presence of a weak interaction between the listening condition and the actual distance from the source COND \times DIST: $F(1, 722) = 7.5482$, $p = 0.0062$, $\eta^2 = 0.01$ (Fig. 7e). Indeed, the distance is on average correctly reproduced by the auralization for the source at 4 m (D4), while the sources placed at 2 m (D2)

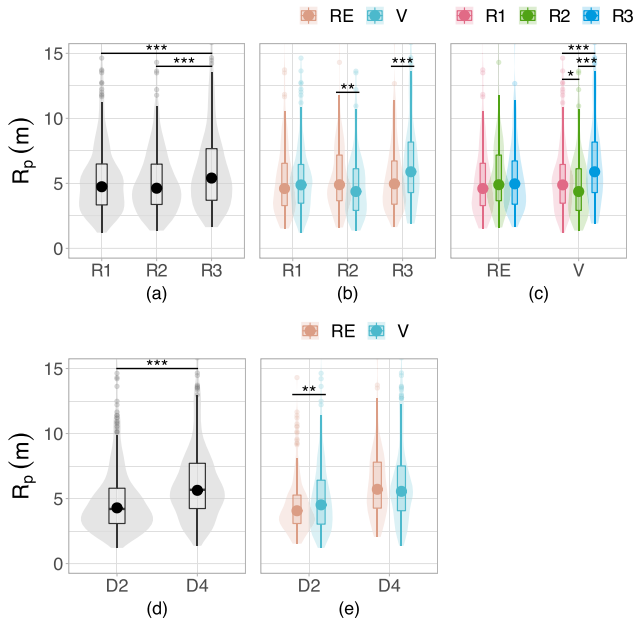


Figure 7. Perceived distance as a function of (a) the acoustic environment ROOM (R1, R2, R3), (b) the interaction COND \times ROOM between the listening condition COND (RE: real listening, V: HOA4 auralization) on the x -axis and the acoustic environment in color, (c) the interaction COND \times ROOM between the acoustic environment on the x -axis and the listening condition in color, (d) the real distance of the source DIST (D2: source at 2 m, D4: source at 4 m), (e) the interaction COND \times DIST between the real distance of the source on the x -axis and the listening condition in color. Post-hoc results are indicated by stars: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)

were perceived significantly farther away in the auralized condition (4.52 m on average) than in the real listening condition (4.06 m on average).

3.4 Effects on apparent source width

The perceived source width is largely affected by the listening condition COND: $F(1, 722) = 362.6654$, $p < 0.0001$, $\eta^2 = 0.33$, shown in Figure 8a. The reported apparent width is significantly higher in the virtual conditions than in real listening. A smaller effect of room is also observed ROOM: $F(2, 722) = 5.9941$, $p = 0.0026$, $\eta^2 = 0.02$, in Figure 8b. A significant interaction between listening condition and room COND \times ROOM: $F(2, 722) = 6.2536$, $p < 0.0022$, $\eta^2 = 0.02$ is also observed (Fig. 8c). Post-hoc tests reveal that the effect of acoustics on the reported apparent width is present only during auralization with smaller source widths for room R1 than for the other two rooms. Finally, the analysis reveals an effect of source distance DIST: $F(1, 722) = 8.2625$, $p < 0.0042$, $\eta^2 = 0.01$, as shown in Figure 8d, with lower apparent source width for sources at 4 m (0.018 sr) than that reported for sources at 2 m (0.026 sr). A significant interaction between the listening condition and the distance from the source was also found COND \times DIST: $F(1, 722) = 6.1858$, $p < 0.0131$,

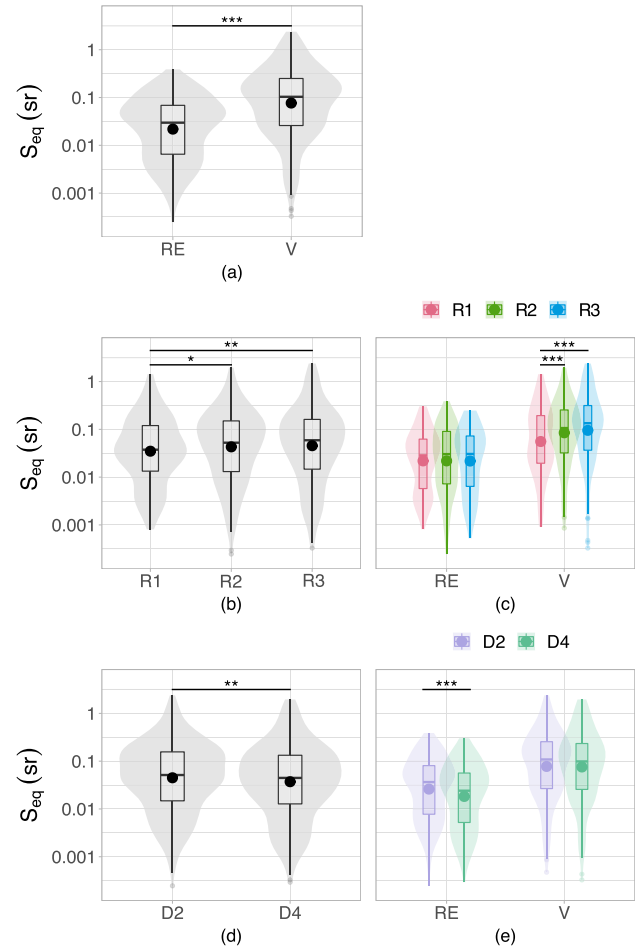


Figure 8. Apparent width reported as a function of (a) listening condition COND (RE: real listening, V: HOA4 auralization), (b) acoustic environment ROOM (R1, R2, R3), (c) the interaction COND \times ROOM between the listening condition on the x -axis and acoustic environment in color, (d) the distance from source DIST (D2: source at 2 m, D4: source at 4 m), (e) the interaction COND \times DIST between listening condition on the x -axis and source distance in color. Post-hoc results are indicated by stars: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)

$\eta^2 < 0.01$ (Fig. 8e). In particular, in the case of real listening, the reported apparent source width is smaller for sources at 4 m than for those at 2 m, whereas in the virtual condition, no difference in apparent width is observed.

Note that the distributions shown in Figures 8c and 8e display, in certain cases, two lobes. This phenomenon could potentially be attributed to interactions with random factors, specifically the sound stimulus (STIM: Speech, Guitar, Burst) and the participants.

4 Discussion

4.1 Source perception in real listening condition

The results revealed that in real listening condition, the localization performances for sources placed in different acoustic environments were globally good and comparable

to those obtained in related studies [51]. As far as angular localization is concerned, the participants perceived the sources close to the actual position of the sources, with low average errors. Therefore, the acoustics of the room do not seem to disturb the angular localization performance in real condition. By contrast, we observed that the distance of the source were overestimated for both distances D2 and D4. This diverges from most related studies which instead indicate a tendency to underestimate the distance for sources located beyond 1 m [2]. This difference may be due to a bias induced by the VR interface and is discussed in more detail in Section 4.3. We found that the performance of localization in azimuth decreased as the source-listener distance increases, although the error difference was small ($<1^\circ$). Given that the sound level of the source at the listener's position decreases as the distance increases, we assume that the localization performance in azimuth is impacted by the ratio of direct field energy to room reverberant energy, called the Direct-to-Reverberant Ratio (DRR). Actually, since the experiment was conducted in untreated rooms and considering that the level of the room response to source excitation is not very sensitive to source position, the DRR decreases as the source moves further away. Therefore, the lower DRR, the more difficult the source is to locate, and the greater the location error. This assumption is in agreement with previous findings [7, 52]. Finally, we observed that the apparent source width reported by the participants was lower for sources at 4 m than for those at 2 m. This suggests that participants perceived and reported the effective width of the source (i.e. real width of the sound source, in contrast with its apparent width), which can be calculated by multiplying the average reported apparent width \bar{S}_{eq} by the average reported distance \bar{R}_p , for each of the two distances: $\bar{S}_{\text{eq}}(D2) \times \bar{R}_p(D2) = 0.106 \text{ m}^2$ and $\bar{S}_{\text{eq}}(D4) \times \bar{R}_p(D4) = 0.103 \text{ m}^2$. It is approximately the same value for both distances and corresponds to the area of a disk of about 9 cm radius.

4.2 Source perception in auralized listening conditions

The results clearly demonstrates that the spatial perception of sound sources is influenced by the auralization of the measured acoustic environments. On average, localization performance is degraded in the auralized cases compared to the real cases. These degradations are manifested through an increase of localization errors in azimuth and elevation as well as an increase of the apparent source widths. These results are comparable to those found in the literature for different ambisonic systems [16, 17, 40]. Hence, the degradation of azimuthal localization may be attributed to a poor reproduction of binaural cues (ILD and ITD), which are essential for localization in the azimuthal plane. An objective measurement of these cues in the auralized condition, would provide further insights.

Additionally, when examining the signed errors in elevation localization in auralized condition, a clear upward attraction of the sources is observed, i.e. sources were on average perceived 15.7° higher than their actual position. It is worth mentioning that this phenomenon is well known

from the scientific community and has been mainly observed in the case of half-sphere 3D ambisonic rendering systems [16]. However, to our knowledge, there has been limited research on this particular phenomenon. As elevation perception relies on monaural frequency cues within the 4–16 kHz frequency range, the spatial aliasing at high frequencies, imposed by the characteristics of the microphone array (above 5 kHz for the em32) can explain the difficulty in locating the sources along the vertical plane. Moreover, while wearing a VR headset seems to have a limited impact on azimuthal localization performance during 2D ambisonic rendering [40], it may have a more significant effect on the perception of source elevation.

Finally, the auralization system had an impact on the apparent width of the sources reported by the participants. Specifically, the sources were perceived as significantly larger in the auralized conditions compared to the real conditions (see Fig. 8). This result raises several hypotheses. Firstly, it suggests that the rendering of interaural cross-correlation (IACC), which is known to be correlated with apparent source width, may not be accurately reproduced by the system. Secondly, the strong correlation observed between the reported source width and azimuthal localization error indicates that participants in virtual conditions may also experience localization blur. This localization blur, in conjunction with the nature of the localization task and the instructions given to participants, eventually leads to an increase in reported source width. The interpretation of these results, in relation to the reporting method, will be discussed in more detail later (see Sect. 4.3).

4.2.1 Degradations depending on source distance

We noticed that most of the differences perceived between the two source distances in real listening are not observed in the auralized cases. This is particularly the case for the absolute error of localization in azimuth and the apparent width of the sources. The difficulty in virtual conditions to locate the sources precisely seems to prevail and leads to (1) a large increase in the average value of these two quantities and (2) a large variability in the report of these quantities. It is worth mentioning that the analysis of the apparent source width in relation to the perceived distance in the real listening condition suggested that participants were capable of accurately perceiving the effective source width (see Sect. 4.1). However, this observation did not hold true for the auralized condition. One possible explanation for this disparity is that the apparent source widths were considerably larger in the auralized condition, potentially making it more challenging for participants to accurately judge the effective width of the sources.

On the other hand, the overall impression of distance from the sources seems to be fairly well reproduced by the system. Indeed, an increase of the perceived distance, when the real distance increases, was observed in both listening conditions. However, while the reporting of the perceived distance for the sources at 4 m was approximately identical in the two listening conditions, the analysis revealed that the sources at 2 m were perceived significantly farther in

the auralized cases than in the real cases. According to the literature on auditory perception of distance, the most salient acoustic cues for judging the distance of a source in a reverberant environment are on the one hand the direct field sound level and on the other hand the DRR, both of which decrease with increasing distance from the source. We can hypothesize that these two attributes are not consistently rendered by auralization among all distances. This hypothesis is supported by a recent study by Lee and Johnson [53], showing noticeable differences in measured DRR across several 3D microphone arrays, suggesting that the energy balance between direct-sound and reverberated field is also influenced by HOA encoding/decoding considerations. On the other hand, these observations may also be influenced by inconsistent reproduction of spatial cues such as Interaural Level Difference (ILD), Interaural Time Difference (ITD) and Interaural Cross-Correlation (IACC) induced by HOA technique. These limitations can critically affect the ability to perceptually differentiate between localized events (i.e. direct-sound) and diffuse field effects (i.e. room effect). As a consequence, this could lead to increased challenges in localizing sound sources (localization blur) and in estimating source distance in auralized reverberant environments. However, these hypotheses need further objective measurements to be investigated (e.g. measuring sound level of direct sound and DRR as well as spatial cues such as ILD, ITD and IACC for both real and auralized conditions).

4.2.2 Degradations depending on the acoustic environments

While the study did not reveal any significant effect of the acoustic environment on the participants' reports in real listening conditions, significant differences between the rooms were observed in virtual conditions. First, the localization performances in terms of absolute errors in azimuth and apparent source width were significantly better in room R1 than in the other two rooms. From an acoustic point of view, room R1 differs from the other two by its shorter reverberation time, especially at low frequency. We can then hypothesize that the spatial precision of the restitution is disturbed by the quantity of diffuse field of the environment to be auralized.

Additionally, from a global perspective, the perception of source distance does not show significant differences across the listening conditions (RE, V). The reported distance when averaged across the factors in RE condition did not differ from those averaged across all the factors in auralized condition. This lack of difference would suggest that the cues for judging distance (notably direct sound level and DRR) were generally well reproduced by auralization system. However, statistical interaction between listening condition (RE, V) and acoustical environment (R1, R2, R3) reveals that in virtual condition only, perceived distance was impacted by the auralized room. As shown in Figure 7b, the auralized version of R2 demonstrates smaller perceived source-listener distances compared to the real listening condition, while the auralized version of R3 tends to increase

perceived source-listener distances. This result suggests the quality of reproduction of the distance cues (direct sound level and DRR) may be also influenced by acoustical properties of the auralized room. As mentioned in previous section, these findings could also be imputed to an inconsistent reproduction of binaural cues induced by HOA acquisition/rendering systems. To go further, it is necessary to complement these observations with objective measurements of distance and binaural cues in auralized conditions. This will be a part of our future work.

4.3 Feedback on the reporting method

The chosen reporting method in this experiment allows to characterize the spatial perception of sound sources according to different attributes. Actually, the present results highlighted a certain number of biases and confusions induced by the experimental device.

Firstly, it was shown that wearing a VR headset has an impact on the HRTF of the listener wearing it [39]. As mentioned in Section 4.2, although VR headset seems to have a limited impact for azimuthal localization of sound sources [40], when performing closed-loop localization task, it may have a much more critical impact regarding elevation localization.

Secondly, there may be confusion regarding the interpretation of the apparent width of the sources due to the reporting method. Apparent source width is reported by instructing participants to surround the sound source. As a reminder, the instruction for the task was: "Precisely surround, using the pointer, the area in which the source is heard." However based on the present results, one can question the meaning behind participants responses. In real cases, it appears that the reported source width is indeed related to the perceived source width, as the observed differences in apparent width between the two source distances lead to the same effective source width (see Sect. 4.1). On the other hand, in virtual conditions, the strong correlation between the reported source width and azimuth localization error indicates that participants responses are rather a reflection of blurriness in source localization. This confusion is inherent to the reporting method, which does not allow for distinguishing between reporting source width or localization blur. One potential approach to disambiguate this issue would be to ask participants to rate, for each trial, their confidence in their response. This way, low confidence levels would be interpreted as a blurry situations while high confidence levels would be interpreted as precise reporting of source width.

Finally, we observe an overall overestimation of the distance by the participants regardless of the listening condition (Fig. 9). These results differ from those reported in the literature on auditory distance perception, which indicate that sound sources located beyond 1 m are generally underestimated, with greater underestimation as the distance increases [1, 2, 54]. Several studies comparing visual perception of distance in real and virtual conditions have revealed an underestimation of visual distance in virtual reality [55–57]. Therefore, participants may have visually

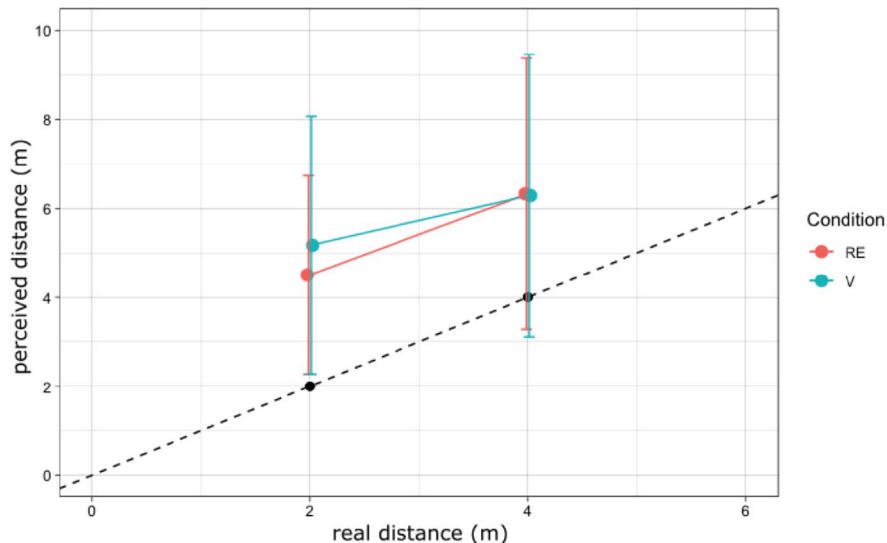


Figure 9. Mean distances reported for real distances of 2 and 4 m, in real (RE) and virtual (V) listening conditions. The whiskers represent the 95% confidence interval. The black dotted line represents the real distance of the sources. All points above this line correspond to an overestimation of the distance. All points below this line correspond to an underestimation of the distance.

underestimated the distance of the pointer used to report the perceived distance, resulting in an overestimation of the actual distance. Understanding this phenomenon of distance underestimation in VR remains an important research topic in the field of virtual reality. A meta-analysis of no less than 40 studies on this topic, proposed by Feldstein et al. [58], suggests a clear improvement in virtual technologies and a reduction of this bias over the past 10 years. Specifically, it has been shown that distance perception is more accurate in virtual scenes representing closed environments compared to open environments [56]. The realism of the virtual environment also appears to influence distance evaluation [58], particularly factors such as the realism of the ground or the presence of an avatar representing the user's body. Additionally, the presence of elements that depict perspective, such as a convergence point, can enhance distance judgement. These studies provide valuable insights for improving virtual environments to facilitate more accurate reporting of perceived distance.

5 Conclusions and future works

In this study, localization performances in different acoustic environments were evaluated under real and auralization conditions using a 4th order 3D HOA system. A novel method was proposed for reporting the spatial image of the sources through a VR interface, allowing for simultaneous assessment of angular position, distance, and apparent width of sound sources. The study indicated degraded localization performances in auralized conditions, particularly in terms of the angular accuracy of the sources. It is hypothesized that the reporting of the apparent source width not only represented the perceived source width but also reflected the degree of blur in localization. The amount of degradation was variable and appeared to be generally higher for reverberant acoustics. Furthermore, while the

overall perception of distance seemed to be accurately reproduced by the system, the results revealed that in the auralization condition, both the acoustics and the actual distance of the source had an impact on the perceived distance. Finally, a systematic overestimation of the perceived distance was observed. This result differs from those presented in the literature on the perception of the distance of sound sources, which reveals the presence of a bias in the distance reporting method, probably induced by a poor visual evaluation of distance in virtual reality.

Based on these results, three perspectives are considered to guide future work. Firstly, it is observed that the nature of spatial image degradations appears to be dependent on the acoustic properties of the measured and auralized environment. Hypotheses were formulated regarding the potential causes of observed degradations. To progress further, it is now crucial to meet the present results with objective data. For instance, the degradation of angular localization performances could be treated through the study of the quality of restitution of the localization cues (ILD, ITD and IACC). Secondly, the proposed methodology can be deployed to characterize other auralization systems such as hybrid HO-SIRR [59] or SDM [60], which were designed to render with better accuracy the spatial properties of ambisonic signals, or binaural auralization methods with head-tracking [61, 62]. Thirdly, for practical reasons, the present experiment has been conducted on a limited set of acoustic conditions (three rooms, two distances per room). However, the portability and lightweight nature of the present VR protocol enabled the acquisition of precise information regarding actual perception within both real and virtual listening conditions. Moving forward, new in-situ studies can be carried out with this device, allowing to diversify the acoustic conditions and leading to a broader characterization of the spatial perception of sources in reverberant environments.

Conflict of interest

Author declared no conflict of interests.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon request and at https://www.prism.cnrs.fr/publications-media/ACTAACUS23_Fargeot [63]. In addition, a demo version of the proposed VR interface for reporting spatial attributes of sound sources (see Sect. 1.3), is available at <https://gitlab.prism.cnrs.fr/fargeot.prism.cnrs.fr/VRLoc-Toolkit/> [64].

References

- P. Zahorik, D.S. Brungart, A.W. Bronkhorst: Auditory distance perception in humans: a summary of past and present research. *ACTA Acustica united with Acustica* 91, 3 (2005) 409–420.
- A.J. Kolarik, B.C.J. Moore, P. Zahorik, S. Cirstea, S. Pardhan: Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics* 78 (2016) 373–395.
- A. Ihlefeld, B.G. Shinn-Cunningham: Effect of source spectrum on sound localization in an everyday reverberant room. *Journal of the Acoustical Society of America* 130 (2011) 324–333.
- J. Käsbach, A. Wiinberg, T. May, M.L. Jepsen, T. Dau: Apparent source width perception in normal-hearing, hearing-impaired and aided listeners, DAGA, Nürnberg, 2015.
- P. Wang, Z. Lin, X. Qiu: Influence of interaural cross-correlation coefficient and loudness level on auditory source width at different frequency. *Applied Acoustics* 162 (2020) 107198.
- W.M. Hartmann: Localization of sound in rooms. *Journal of the Acoustical Society of America* 74, 5 (1983) 1380–1391.
- B.G. Shinn-Cunningham: Localizing sound in rooms, in: *ACM/SIGGRAPH and Eurographics Campfire: Acoustic Rendering for Virtual Environments*, 2001, pp. 1–6.
- M. Rychtáriková, T.V. den Bogaert, G. Vermeir, J. Wouters: Binaural sound source localization in real and virtual rooms. *Journal of the Audio Engineering Society* 57, 4 (2009) 205–220.
- J. Meyer, G. Elko: A highly scalable spherical microphone array based on an orthonormal decomposition of the sound-field, in: *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, IEEE, 2002, p. II–1781.
- M. Noisternig, A. Sontacchi, T. Musil, R. Holdrich: A 3D ambisonic based binaural sound reproduction system, in: *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, Audio Engineering Society, 2003.
- B. Bernschütz, A.V. Giner, C. Pörschmann, J. Arend: Binaural reproduction of plane waves with reduced modal order. *Acta Acustica united with Acustica* 100, 5 (2014) 972–983.
- G. Routray, P. Dwivedi, R.M. Hegde: Binaural reproduction of HOA signal using sparse multiple measurement vector projections, in: *2021 National Conference on Communications (NCC)*, IEEE, 2021, pp. 1–6.
- E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman: Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America* 94, 1 (1993) 111–123.
- G. Parsehian, B.F. Katz: Rapid head-related transfer function adaptation using a virtual auditory environment. *Journal of the Acoustical Society of America* 131, 4 (2012) 2948–2957.
- S. Bertet, J. Daniel, E. Parizet, O. Warusfel: Influence of microphone and loudspeaker setup on perceived higher order ambisonics reproduced sound field, in: *Proceedings of Ambisonics Symposium*, 2009.
- S. Braun, M. Frank: Localization of 3D ambisonic recordings and ambisonic virtual sources, in: *1st International Conference on Spatial Audio*, Detmold, 2011.
- S. Bertet, J. Daniel, E. Parizet, O. Warusfel: Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources. *Acta Acustica united with Acustica* 99, 4 (2013) 642–657.
- F.L. Wightman, D.J. Kistler: Headphone simulation of free-field listening. II: psychophysical validation. *Journal of the Acoustical Society of America* 85, 2 (1989) 868–878.
- R. Mason, N. Ford, F. Rumsey, B. De Bruyn, Verbal and non-verbal elicitation techniques in the subjective assessment of spatial sound reproduction, vol. 5225. *Audio Engineering Society Preprint*, 2000.
- L. Haber, R.N. Haber, S. Penningroth, K. Novak, H. Radgowski: Comparison of nine methods of indicating the direction to objects: data from blind adults. *Perception* 22, 1 (1993) 35–47.
- M. Gröhn, T. Lokki, T. Takala: Localizing sound sources in a cave-like virtual environment with loudspeaker array reproduction. *Presence: Teleoperators and Virtual Environments* 16, 2 (2007) 157–171.
- H. Bahu, T. Carpentier, M. Noisternig, O. Warusfel: Comparison of different egocentric pointing methods for 3d sound localization experiments. *Acta acustica united with Acustica* 102, 1 (2016) 107–118.
- M. Goupell, B. Laback, P. Majdak, M. Mihocic: The accuracy of localizing virtual sound sources: effects of pointing method and visual environment, in: *Audio Engineering Society Convention 124*, Audio Engineering Society, 2008.
- T. Djelani, C. Pörschmann, J. Sahrhage, J. Blauert: An interactive virtual-environment generator for psychoacoustic research II: collection of head-related impulse responses and evaluation of auditory localization. *Acta Acustica united with Acustica* 86, 6 (2000) 1046–1053.
- B. Seeber: A new method for localization studies. *Acustica-Stuttgart* 88, 3 (2002) 446–449.
- P. Majdak, M.J. Goupell, B. Laback: 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics* 72 (2010) 454–469.
- F. Winter, H. Wierstorf, S. Spors: Improvement of the reporting method for closed-loop human localization experiments, in: *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.
- E.H. Langendijk, A.W. Bronkhorst: Collecting localization response with a virtual acoustic pointer. *Journal of the Acoustical Society of America* 101 (1997) 3106–3106.
- V. Pulkki, T. Hirvonen: Localization of virtual sources in multichannel audio reproduction. *IEEE Transactions on Speech and Audio Processing* 13, 1 (2004) 105–119.
- S. Bertet: *Formats audio 3D hiérarchiques: caractérisation objective et perceptive des systemes ambisonics d'ordres supérieurs*. PhD thesis 2009.
- R.H. Gilkey, M.D. Good, M.A. Ericson, J. Brinkman, J.M. Stewart: A pointing technique for rapidly collecting localization responses in auditory research. *Behavior Research Methods, Instruments, & Computers* 27, 1 (1995) 1–11.

32. J. Braasch, K. Hartung: Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data. *Acta Acustica United with Acustica* 88, 6 (2002) 942–955.
33. M. Schoeffler, S. Westphal, A. Adami, H. Bayerlein, J. Herre: Comparison of a 2D-and 3D-based graphical user interface for localization listening tests, in: *Proceeding of the EAA Joint Symposium on Auralization and Ambisonics*, vol. 3, 2014, p. 5.
34. H.G. Hassager, A. Wiinberg, T. Dau: Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment. *Journal of the Acoustical Society of America* 141, 4 (2017) 2556–2568.
35. J.C. Gil-Carvajal, J. Cubick, S. Santurette, T. Dau: Spatial hearing with incongruent visual or auditory room cues. *Scientific Reports* 6 (2016) 37342.
36. F. Rumsey: Perceptual evaluation: listening strategies, methods, and VR. *Journal of the Audio Engineering Society* 66, 4 (2018) 301–305.
37. G.C. Stecker: Using virtual reality to assess auditory performance. *The Hearing Journal* 72, 6 (2019) 20–22.
38. S. Fargeot, O. Derrien, G. Parsehian, M. Aramaki, R. Kronland-Martinet: Subjective evaluation of spatial distortions induced by a sound source separation process, in: *EAA Spatial Audio Signal Processing Symposium, 2019*, pp. 67–72.
39. R. Gupta, R. Ranjan, J. He, G. Woon-Seng: Investigation of effect of VR/AR headgear on Head related transfer functions for natural listening, in: *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.
40. T. Huisman, A. Ahrens, E. MacDonald: Sound source localization in virtual reality with ambisonics sound reproduction, *PsyArXiv*, 2021.
41. M. Berzborn, R. Bomhardt, J. Klein, J.-G. Richter, M. Vorländer, The ITA-toolbox: an open source MATLAB toolbox for acoustic measurements and signal processing, in: *Proceedings of the 43th Annual German Congress on Acoustics*, Kiel, Germany, vol. 2017, 2017, pp. 6–9.
42. A. Farina: Simultaneous measurement of impulse response and distortion with a swept-sine technique, in: *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.
43. Datasheet: Eigenbeam data, specifications for eigenbeams. Tech. Rep. 1.4 mh acoustics, LLC, 2016.
44. T. Carpentier: A new implementation of Spat in Max, in: *15th Sound and Music Computing Conference (SMC2018)*, 2018, pp. 184–191.
45. F. Zotter, H. Pomberger, M. Noisternig: Energy-preserving ambisonic decoding. *Acta Acustica United with Acustica* 98 (2012) 37–47.
46. T. Carpentier, M. Noisternig, O. Warusfel: Twenty years of Ircam Spat: Looking back, looking forward, in: *41st International Computer Music Conference (ICMC)*, 2015, pp. 270–277.
47. M.J.-M. Macé, F. Dramas, C. Jouffrais: Reaching to sound accuracy in the peri-personal space of blind and sighted humans, in: *International Conference on Computers for Handicapped Persons*, Springer, 2012, pp. 636–643.
48. R. Brown, Fitellipse.m, 2023. <https://fr.mathworks.com/matlabcentral/fileexchange/15125-fitellipse-m>.
49. A. Kuznetsova, R.H. Christensen, C. Bavay, P.B. Brockhoff: Automated mixed ANOVA modeling of sensory and consumer data. *Food Quality and Preference* 40 (2015) 31–38.
50. A. Ahrens, K.D. Lund, M. Marschall, T. Dau: Sound source localization with varying amount of visual information in virtual reality. *PloS One* 14 (2019) e0214603.
51. J. Blauert: *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT press, 1997.
52. J.M. Buchholz, V. Best: Speech detection and localization in a reverberant multitalker environment by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America* 147 (2020) 1469–1477.
53. H. Lee, D. Johnson: 3D microphone array comparison: objective measurements. *Journal of the Audio Engineering Society* 69 (2021) 871–887.
54. D.H. Mershon, J.N. Bowers: Absolute and relative cues for the auditory perception of egocentric distance. *Perception* 8 (1979) 311–322.
55. J.M. Loomis, J.M. Knapp: Visual perception of egocentric distance in real and virtual environments. *Virtual and Adaptive Environments* 11 (2003) 21–46.
56. C. Armbrüster, M. Wolter, T. Kuhlen, W. Spijkers, B. Fimm: Depth perception in virtual reality: distance estimations in peri- and extrapersonal space. *Cyberpsychology & Behavior* 11, 1 (2008) 9–15.
57. J.W. Kelly, L.A. Cherep, B. Klesel, Z.D. Siegel, S. George: Comparison of two methods for improving distance perception in virtual reality. *ACM Transactions on Applied Perception (TAP)* 15, 2 (2018) 1–11.
58. I.T. Feldstein, F.M. Kölsch, R. Konrad: Egocentric distance perception: a comparative study investigating differences between real and virtual environments. *Perception* 49, 9 (2020) 940–967.
59. L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, M. Marschall: Higher-order spatial impulse response rendering: investigating the perceived effects of spherical order, dedicated diffuse rendering, and frequency resolution. *Journal of the Audio Engineering Society* 68 (2020) 338–354.
60. S. Tervo, J. Pätynen, A. Kuusinen, T. Lokki: Spatial decomposition method for room impulse responses. *Journal of the Audio Engineering Society* 61, 1/2 (2013) 17–28.
61. P. Stitt, E. Hendrickx, J.C. Messonnier, B.F. Katz: The role of head tracking in binaural rendering, in: *29th Tonmeister-tagung, International VDT Convention*, 2016.
62. M. Romanov, P. Berghold, M. Frank, D. Rudrich, M. Zaunschirm, F. Zotter: Implementation and evaluation of a low-cost headtracker for binaural synthesis, in: *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.
63. S. Fargeot, A. Vidal, M. Aramaki, R. Kronland-Martinet: Stimuli for perceptual evaluation of an ambisonic auralization system of measured 3D acoustics [Data set], 2023. https://www.prism.cnrs.fr/publications-media/ACTAACUS23_Fargeot.
64. S. Fargeot, A. Vidal, M. Aramaki, R. Kronland-Martinet: VRLoc-Toolkit: a set of tools for investigating sound source localization in VR [Code], 2023. <https://gitlab.prism.cnrs.fr/fargeot.prism.cnrs.fr/VRLoc-Toolkit>.