Comparison of Different Egocentric Pointing Methods for 3D Sound Localization Experiments

Hélène Bahu, Thibaut Carpentier, Markus Noisternig, Olivier Warusfel

Acta Acustica united with Acustica
Vol. 102 (2016) 107 – 118

DOI 10.3813/AAA.918928

Comparison of Different Egocentric Pointing Methods for 3D Sound Localization Experiments

Hélène Bahu, Thibaut Carpentier, Markus Noisternig, Olivier Warusfel
IRCAM, CNRS, Sorbonne Universités, UPMC Univ Paris 06, UMR 9912 STMS, 75004 Paris, France.
helene.bahu@hotmail.fr, markus.noisternig@ircam.fr

Summary
This study evaluates several methods for reporting the perceived location of real sound sources. It is well known that the method used for collecting judgments in auditory-localization experiments has a strong influence on the accuracy of a subject's response. Previous works on auditory-localization tasks revealed that egocentric pointing methods (which are based on a body-centered coordinate system) allow for more accurate judgments than verbal reporting or exocentric pointing techniques (which are based on a 2D or 3D reporting device). Three different egocentric methods are compared: the most commonly applied "manual pointing" and "head pointing" methods, and the "proximal pointing" method, which forces the participants to indicate the apparent direction by pointing in the proximal region of the head with a marker held at the fingertips. The two first methods involve a rotation of the body of the participant, whereas the third method only involves movements of the arm(s) and hand(s) with a fixed head. Sound stimuli were presented randomly over 24 loudspeakers that were uniformly distributed on the upper hemisphere around the subject. The merits of the different methods are compared and discussed with regard to localization errors and to practical considerations. Although they show similar trends, each of the different methods affects the pointing accuracy in a specific way. The proximal pointing method, for example, is more accurate for sources located at high elevation angles. However, at rear locations close to the median plane an increased bias appears due to difficulties in performing the motor task to reach these positions. The proximal pointing method shows faster response times, which may be advantageous when planning 3D sound localization experiments.

PACS no. 43.66.Qp

1. Introduction

The evaluation of human sound localization performances typically relies on measuring the bias and variability between the perceived and the actual sound location. The precision and reproducibility of such an evaluation not only depend on the subject's ability to localize sound sources, but also on the accuracy of the method used for collecting the responses.

Several reporting methods were proposed and discussed in literature. Although frequently used, the paradigm of verbal report (in which the subject indicates the apparent spatial position in terms of azimuth and elevation angles [1, 2]) was found to be unintuitive for unexperienced subjects, and thus yielding inaccurate judgments [3].

Exocentric reporting techniques were also investigated. In [4, 5, 6, 7], subjects reported the perceived sound location on a 2D/3D graphical interface, which schematically represents different views of the environment. Another reporting technique, the "God's Eye Localization Pointing" method (GELP, see *e.g.* [8, 9]), uses a rigid sphere that represents the auditory space and the subject indicates the perceived source direction by pointing with a stick onto the sphere. This method allows for fast response times. However, a systematic judgment error was observed for such exocentric methods, as subjects typically have difficulties to "project" their own auditory space onto the reporting device, *i.e.* the rigid sphere or the graphical user interface [5, 9].

Egocentric pointing methods have been shown to be more precise than exocentric methods [5, 8, 9]. The most common egocentric methods use manual pointing (*e.g.*, with the finger or with a tracked object held in the hand) [5, 9, 10, 11] or head pointing [9, 10, 12, 13]. As has been shown in Frank *et al.* [14], the advantage of the manual pointing method (implemented with a toy gun in their study) is that it is very intuitive and technically reliable. The different methods are sometimes used along with some visual feedback that indicates the pointed direction (*e.g.*, a laser pointer mounted on the head or handheld pointing device) [14, 15, 16, 17, 18, 19, 20], as it has been shown that a visual feedback improves the pointing accuracy [12, 18]. Seeber *et al.* [21, 22] introduced a reporting method, where the subject indicates the perceived

Received 26 February 2015,
accepted 11 October 2015.

direction with a trackball-controlled laser pointer, while the head is maintained in a fixed position. This method reduces the complex interaction effects between the different sensorial modalities (*i.e.*, vestibular, visual, and auditory), but is restricted to the field of vision (*i.e.* frontal targets). Other methods use an acoustic pointer that has to be aligned with a reference sound [23, 24, 25]. Those methods have also been omitted, as they are not very well adapted to 3D localization tests with real sound sources.

This study focuses on 3D localization tests with real sound sources (which are regularly distributed around the listeners) and only involves the auditory and proprioceptive modality. It compares three different egocentric pointing methods: (i) The "head pointing" method, where the participant is asked to turn his body towards the perceived sound source, and then to point to its direction with the head. When applied in previous studies (see *e.g.* [9, 10, 12]), this method showed many advantages over other methods, despite a difficulty to reach highly elevated targets. (ii) The "manual pointing" method is investigated using a toy-gun held in one hand. This method is also referred to as "gun pointing". The participant has to rotate on a swivel chair before aiming at the perceived sound direction with the arm stretched out. Majdak *et al.* [10] compared this method to the head pointing method. Both methods showed similar results, except for a higher accuracy of the manual pointing method for sources at high elevations. (iii) The present study investigates the benefits of a particular manual pointing method, which does not involve any head or torso movements. The participants have to indicate the perceived source direction by placing a marker held at the fingertips in the proximal region of the head. This method will be referred to as "proximal pointing" method throughout the remainder of this article.

The proximal pointing method is closely related to the finger pointing method reported in Djelani *et al.* [9], but allows for the use of both hands. It thus overcomes the systematic pointing error that is associated to movement restrictions when the subject points to sound sources in the hemispace opposite to the hand he is using for pointing. Contrary to [9], in this experiment the subject remained still during the stimulus playback in order to avoid that dynamic localization cues influence the test results. In addition, more source directions were evaluated than in Djelani *et al.*'s study, and real sound sources (*i.e.* sound stimuli played back over loudspeakers) were used instead of virtual sound sources (*i.e.* sound stimuli played back over headphones).

The motivation for studying the proximal pointing method is fourfold: (i) when a subject reports the perceived source direction with the fingertips in the proximal region of her head, being allowed to freely choose either the left or the right hand for pointing, she should easily point to any direction in 3D space; (ii) the proximal pointing method does not require any head or torso movements (not even for rear directions) and thus allows for fast response times; (iii) the measurement apparatus can be used for localization tests with both stimuli presentations over loudspeakers and headphones; (iv) this method may also be used for collecting judgments on the perceived distance of virtual sound stimuli played back over headphones in binaural sound reproduction systems. We however expect several limitations linked to limits of the human motor control. For instance, source directions close to the median plane in the rear hemifield may be difficult to reach with both the left and the right hand. One may also expect that a subject's pointing performance varies greatly depending on if he either responds with the dominant or the non-dominant hand.

This paper is organized as follows: Section 2 presents the experimental apparatus and the different methods. Section 3 examines the results in terms of localization accuracy, dispersion of responses and response time. Section 4 compares the results with reference studies in literature, discusses the respective merits of the different methods, and the limitations and possible future improvements of the proximal method.

2. Methods

2.1. Participants

39 participants (15 female and 24 male), ranging in age from 21 to 55 years, served as paid volunteers. 19 participants were audio professionals and 4 of them had already participated in auditory localization experiments. The experiment consisted of three different conditions (which are detailed in sections 2.4 and 2.5). 13 participants (8 male and 5 female) took part in each condition (in other words, each participant tested only one condition). All participants reported normal hearing. They were informed that the sound could emanate from any direction in 3D space, but had no prior knowledge of the loudspeakers positions. All tests were performed in the same acoustic environment and in total darkness, so that the loudspeakers were invisible for the subjects.

2.2. Experimental Setup

The experiment was conducted in an acoustically damped room with dimensions 7.8 m × 6 m × 4.2 m (experimental studio with all walls and the ceiling covered with porous broadband sound absorption panels, $RT_{60} = 150$ ms at 1 kHz; the floor was covered with additional sound absorbing fabric material) and a background noise level of 27 dBA. The room was equipped with a dome of twenty-four coaxial Amadeus PMX-5 loudspeakers, with an effective frequency range from 80 Hz to 22 kHz. The loudspeakers were uniformly distributed on a hemisphere with a radius of 2.65 m. The spatial distribution was symmetric about the median plane and composed of 3 horizontal rings at elevations of −5°, 26° and 57°, respectively, plus one additional loudspeaker at the zenith of the sphere. The loudspeaker positions are detailed in Table I. In the remainder of this paper, azimuth angles are measured clockwise from the median plane (front to rear), *i.e.* negative and positive azimuth angles are in the left and right hemispace, respectively. Elevation angles are measured from the

horizontal plane, with positive values in the upper hemisphere. All loudspeakers were oriented towards the center of the hemisphere, which also determines the origin of the reference coordinate system.

All participants were seated on a height-adjustable swivel chair and their heads were positioned in the center of the sphere. The position and the orientation of both the participant's head and the pointing devices were captured in real-time (with an update rate of 60 Hz) by the means of four infrared motion capture cameras (A.R.T. Track3). With the motion capture software, several tracking objects were defined, each consisting of a set of reflective markers mounted on a rigid body. For the gun pointing and proximal pointing methods, a toy-gun and small ad hoc objects were used, respectively. For head tracking, several markers were mounted on the inner frame of a safety helmet, which could be easily adjusted to the participant's head in such a way that it did not move or change its shape during the experiment. Prior to each experiment, and after aligning the participant's head with the reference coordinate system, the motion capture system was calibrated. A set of four coincident self-leveling cross-line laser beams was used to center and align the head (the interaural axis, the median plane, and the horizontal plane). This defines the initial position of the head. Then the head was tracked in real-time during the repositioning phase and for calculating the pointed direction (see section 2.5).

A 17-inch computer monitor was placed below the frontal loudspeaker to display an interactive graphical user interface (GUI). The GUI was solely switched on during the repositioning phase, in order to guide the participants to retrieve the initial head position, *i.e.* $(x, y, z) = (0, 0, 0)$ and orientation of 0° in azimuth and elevation. The GUI depicted a top view of the listener's head (left side of the screen) for adjusting the (x, y) position, and the orientation of the listener's head (right side of the screen). A vertical slider guided the participants to adjust their vertical head position. The GUI was refreshed in real time according to the head tracking data and the background color was switched from red to green to indicate that the subject was well positioned.

In order to make sure that none of the participants could see the loudspeakers, they (i) entered the darkened room blindfolded, and (ii) had to wear opaque sunglasses during the entire experiment. A preliminary test showed that, although the computer monitor was switched on during the repositioning phase, none of the loudspeakers was visible. Prior to each trial, the GUI was switched off to avoid visual anchors during the localization and pointing tasks.

2.3. Stimuli

The sound stimuli should be short enough in order to avoid that subjects receive dynamic auditory cues resulting from head movements while the sound is being played. According to Blauert [26], it takes at least 200 ms to initiate head movements. In other words, stimuli with a duration of less than 200 ms will not be affected. Other studies, which investigated the influence of the stimulus duration on the sound localization (see *e.g.* [12, 27, 28]), revealed that a stimulus of at least 80 ms is required for a stable estimation of the target sound elevation. Katz and Parseihian [29] further showed an improved source localization accuracy in azimuthal direction when using successive short bursts instead of a single noise burst. In this work, the target sound stimulus was a train of four 50 ms Gaussian noise bursts with 10 ms raise and fall times ($\cos^2$ slopes), and 10 ms pauses in between successive bursts. Therefore, the total duration of the target sound stimulus was 230 ms. The sound pressure level of each burst was set to 65 dBA, with a maximum deviation of 2 dBA over all loudspeakers.

2.4. Conditions 1 and 2

Condition 1 – head pointing: Participants were instructed to point their nose towards the direction of the perceived sound source. The reported direction is computed from the intersection of the head orientation with a virtual spherical surface, that has the same radius and center as the loudspeaker hemisphere. The computed intersection point is then converted into polar coordinates. Participants sat on a swiveling chair to facilitate body rotations in azimuthal direction. Once facing the perceived direction, they had to validate their response by double-clicking a button on a hand-held device. Participants had the free choice of which hand to use for the response button. In this experiment, one of the participants showed noticeable differences in performance from the overall average. Therefore, we decided to remove her results from the analysis, resulting in a total of 12 participants, thereof 8 male and 4 female.

Condition 2 – gun pointing: Participants were instructed to aim at the perceived location of the sound with a toy-gun (equipped with reflective markers for motion tracking) held in the hand of their choice. As before, the reported direction is computed from the intersection of the toy-gun orientation with a virtual spherical surface, that has the same radius and center as the loudspeakers hemisphere. The computed intersection point is then converted into polar coordinates. Here again, rotating the body on the swiveling chair was allowed. Participants were instructed to aim at the target with a stretched out arm and to position the gun such that it aligns with the arm (*i.e.* to avoid wrist movements). They had to validate their response by pressing twice the trigger of the gun.

In both conditions, participants entered the room blindfolded, guided by the experimenter. Once seated, their head was positioned in the center of the sphere using a set of coincident laser beams. After calibration, all laser beams and disturbing lights were turned off and the participants remained in total darkness. Then the blindfold was removed so that participants could see the graphical user interface. One trial consisted of the following steps: (1) the subject had to return to the initial position using the GUI; (2) once the position was reached (within a tolerance range of ±1° in azimuth and elevation, and ±2 cm in x, y and z) and maintained for two seconds, the participant was instructed not to move and the GUI was switched off; (3) two seconds later, the sound stimulus was played back

Table I. Loudspeaker coordinates (azimuth and elevation in degrees).

Elevation (°)	Azimuth (°)								
-5	-160	-120	-80	-40	0	40	80	120	160
26	-20	-60	-100	-140	180	140	100	60	20
57	-144		-72		0		72		144
90					0				

over a randomly selected loudspeaker; (4) the subject had to turn his body and head or the gun towards the location of the sound, and then to validate his reported direction by pressing twice the button or gun trigger.

2.5. Condition 3

Condition 3 – proximal pointing: Participants had to place their head on a neck-rest during the experiment in order to avoid head movements. In each hand, they held a rigid body with the reference marker at the fingertips. Participants were instructed to indicate the direction of the sound by placing the fingertips in the perceived direction within the proximal region of the head (see Figure 1). They were free to use the hand of their choice according to the most comfortable way to indicate the corresponding direction. No constraints about the distance of pointing from the head were imposed. Although the participants were instructed not to move their heads, the tracking data showed unconscious head movements during the pointing phase (average position and angular displacements in the order of 1 cm and 1°, respectively). Hence, the reported direction was evaluated between the position of the fingertips and the actual position of the head.

Here again, participants entered the room blindfolded, guided by the experimenter. They were seated on a non-swiveling height-adjustable chair. The position of the neck-rest was adjusted in order to position the participant's head at the center of the loudspeaker hemisphere. As for the other conditions, participants remained in total darkness, and the repositioning GUI was used to ensure the exact positioning of the head just before the stimulus playback. One trial consisted of the following steps: (1) the subject had to adjust the position and orientation of her head according to the GUI; (2) once well placed during 2 seconds, the subject was asked to put the arms onto the armrests; (3) once the arms remained well positioned for at least 1 second, the GUI was switched off; (4) two seconds later, the sound stimulus was played back over a randomly selected loudspeaker; (5) the participant had to indicate the perceived source direction by placing one of the two rigid bodies close to the head, while keeping the other hand on the arm rest, and had to confirm the response by pressing a foot-pedal.

2.6. Training and experiment

The experiment consisted of 192 trials (8 repetitions for each of the 24 source locations), with a total duration of about 40 minutes. No feedback was given with regard to the localization performance, neither during the training nor during the experiment.

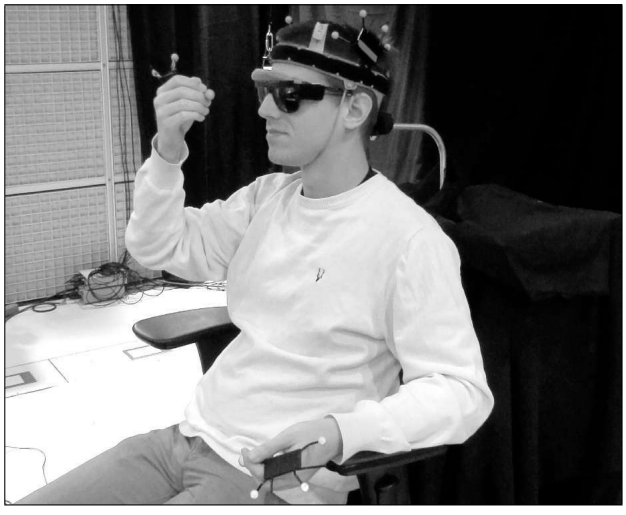

Figure 1. Photo of the proximal pointing task. The participant holds a rigid body in each hand, aligning the reference marker with his fingertips. The head is supported by a neck-rest to prevent head movements, and the head position and orientation were monitored in real-time using the motion capture system (inner frame of a helmet with reflective markers). The subjects were free to use the hand of their choice according to the most comfortable way to indicate the direction of the sound source, which is computed with regard to the current head position.

The subjects were trained to get familiar with the respective pointing method using the same stimuli as for the experiment. A training session consisted of 10 different target sound directions that were identical for all subjects and conditions. The different training directions were selected such that at least one target sound direction per quadrant was tested.

3. Data analysis and results

3.1. Definition of errors

In order to analyze the data, measures used to calculate the localization error and the coordinate system need to be defined. The choice may be determined from practical considerations (such as the spatial distribution of the sound sources), the hypothesized auditory localization model, or the reporting modalities. Wightman and Kistler [1] used a single pole coordinate system, where the azimuth and elevation angles correspond to the longitude and latitude, respectively. Data were analyzed in terms of the average angle of error (*i.e.* the mean of the unsigned angles between

the judgment vector and the target vector), the direction of the judgment centroid, and the associated dispersion. Morimoto and Aokata [30] and Majdak *et al.* [10] used a lateral/polar coordinate system. In lateral directions (*i.e.* the arc length between the judgment and the vertical median plane), errors were analyzed in terms of a lateral bias and a lateral precision. In polar directions (*i.e.* the elevation angle within a circle of constant lateral angle), errors were analyzed in terms of quadrant errors [10]. The use of the lateral angle is in good accordance with auditory localization models that depend on the interaural time and level differences (ITD, ILD).

However, in order to assess the localization error, the sensory-motor constraints (auditory and proprioceptive) of a pointing task should be considered [31, 32]. For instance, the use of the lateral angle is less suitable for head pointing and gun pointing tasks, as they require body rotations (given by the rotation axis of the swiveling chair) and a head and/or arm movement. For this reason, a single pole coordinate system is used in the present study and the localization errors are analyzed as in Makous and Middlebrooks [12] and Gilkey *et al.* [8]. The vertical error is given by the angular distance in between the elevation of the reported direction and the target sound source. As illustrated in Figure 2, the horizontal error is given by the angle between the target vector and a vector from the center of the sphere to a point on the surface of the sphere, whose longitude and latitude are equal to the judgment longitude and target latitude, respectively [8]. This error definition takes the azimuth compression at the zenith into account. For a source at the zenith, the horizontal error becomes zero. Consequently, this target source will not be considered when analyzing the horizontal localization performance.

3.2. Data analysis

Both the signed and unsigned localization errors are computed: unsigned errors only consider the magnitude of the error and thus reflect the localization accuracy, whereas signed errors provide information about the direction of the bias with reference to the target source position. A positive elevation error corresponds to a pointed direction higher than the target source. A positive horizontal error indicates that the subject rotated his body farther than the target position, the zero reference being the initial position, *i.e.* the body and head orientated towards (az, el)=(0°,0°).

No head movements were allowed during the stimulus playback, which results in an increase of front-back confusions [33]. During data analysis, front-back confusions were detected as described in Wightman and Kistler [1]: if the absolute angle between the target direction and the judgment gets smaller when the judgment is mirrored about the frontal plane, it is considered as a front-back confusion. Sources which are close to or on the frontal plane are excluded from the detection of confusions [34]. This applies to targets located in the following areas: (az, el) = (±80°, −5°); (az, el) = (±100°, 26°); (az, el) = (±72°, 57°), and (az, el) = (0°, 90°).

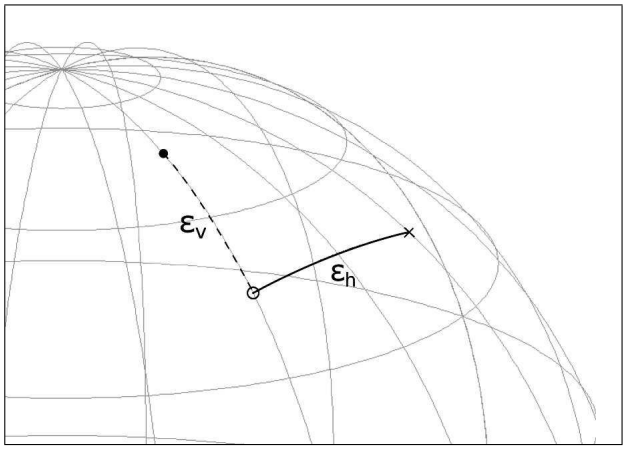

Figure 2. Illustration of the horizontal error angle (ϵ_h) and vertical error angle (ϵ_v) measured on the unit sphere subtended at the center of the head as defined in Makous and Middlebrooks [12] and Gilkey *et al.* [8]. The grey lines represent a part of the sphere's latitude and longitude lines, drawn with 20° increments; the filled circle corresponds to the response direction and the cross represents the stimulus direction. The point represented by the open circle is defined by the longitude of the response direction and the latitude of the stimulus direction.

As the proximal pointing method involves both hands, we examined the differences in the responses given with the dominant and non-dominant hands. This effect will be referred to as lateral dominance with items "dominant hand" and "non-dominant hand". It is worth noting that the participants always used the hand corresponding to the hemispace where the target source was located (*e.g.*, the right hand to indicate a source in the right hemispace); for sources located on the median plane, the hand differed over repetitions, for a given source and the same participant.

A repeated measures ANOVA (analysis of variance) was performed with the pointing condition (head pointing, H; gun pointing, G; proximal pointing, P) as between-subjects factor, and with several within-subject factors, such as (i) the repetition index ranging from 1 to 8; (ii) the hemifield of the target, with the categories "Front" associated to targets ranging clockwise from azimuth −80° to +80°, and "Back" associated to targets ranging clockwise from azimuth +100° to −100° (*i.e.* all targets on the frontal plane are excluded); (iii) the target elevations (−5°, 26°, 57°, and 90°); (iv) the elevation range, with items "Mid" combining the elevations −5° and 26°, and "High" for elevations 57° and 90°.

Following the recommendations given in Johnson [35], a tested effect was considered significant when the p-value was below 0.005, and only marginally significant for values $0.005 < p < 0.05$. Post-hoc tests were performed using the Tukey-Kramer method [10].

3.3. Results

3.3.1. Front-back confusions

Most of the time, subjects were rather consistent in their judgments. We found that the judgments were distributed

as follows: for a given target, the subjects either pointed in the wrong hemifield for all trials (see Figure 3, left subfigure) or just for a minor part of the repeated trials (see Figure 3, right subfigure). This behavior shows a very consistent perception of the target direction (possibly on the wrong hemifield), rather than a dispersion of pointing directions. Consequently, response directions for which confusions were detected were resolved prior to further analysis by mirroring the judgments about the frontal plane.

A repeated measures ANOVA was performed on the front-back reversal rates, with the pointing condition (H, G, P) as between-subjects factor, and the repetition (1-8), the hemifield (Front, Back), and the target elevation (−5°, 26°, 57°) as within-subjects factors. No significant trend was noticed over repetitions ($F(7, 245) = 1.53$; $p = 0.16$) and no significant effect of the condition was found ($F(2, 35) = 0.46$; $p = 0.63$). This last observation may be related to the fact that front-back confusions reflect perceptual ambiguities, and thus should not be affected by the pointing method itself. It further ensures that the applied reversal treatments do not affect the comparison between the different pointing methods. Confusion rates appeared to be significantly higher in the rear hemifield than in the frontal hemifield ($F(1, 35) = 63.54$; $p < 0.001$), which means that most of the confusions were back-to-front confusions. Furthermore, the target elevation has a significant effect on the reversal rate ($F(2, 70) = 49.75$; $p < 0.001$). The interaction effect between target elevation and hemifield was found to be significant too ($F(2, 70) = 52.07$; $p < 0.001$). This is illustrated in Figure 4. A post-hoc test showed that, only for an elevation of 57°, the reversal rates are significantly higher in the rear hemifield than in the front hemifield ($p < 0.001$).

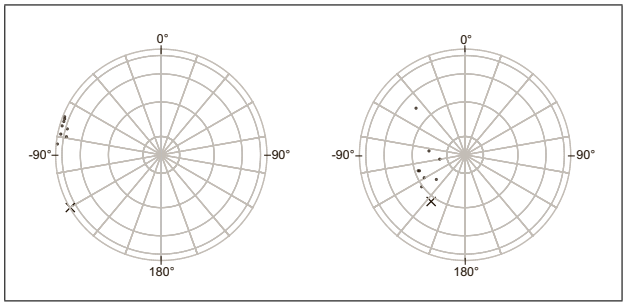

Figure 3. Top views of a schematic representation of the upper sphere (on which the loudspeakers are located). The black crosses represent the target locations and the grey points represent the pointed directions. The shown examples highlight the two different kinds of judgment distributions that reflect front-back confusions. Judgments presented here are raw data (*i.e.* the front-back confusions are not yet treated).

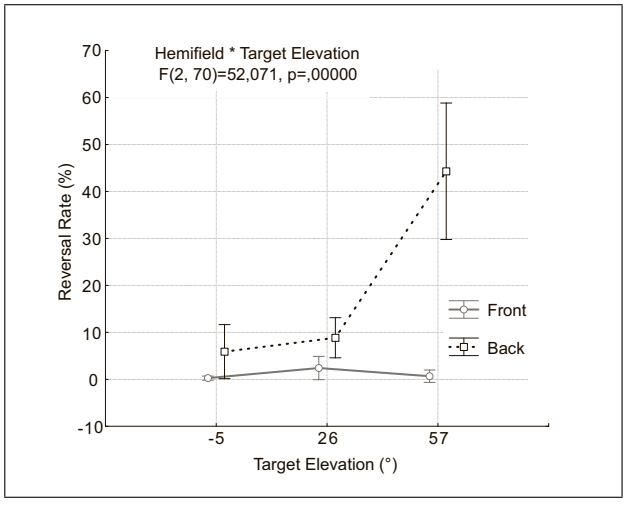

Figure 4. Reversal rate (%) observed on all subjects and all targets located in the front (solid grey line) and back (dashed black line) hemifields. Significance of the interaction effect between source elevation and hemifield is indicated on top of the figure by F and p values. Vertical bars represent standard errors.

3.3.2. Horizontal errors

A repeated measures ANOVA was performed on the unsigned horizontal errors, with the pointing condition (H, G, P) as between-subjects factor, and the repetition (1-8) and the hemifield (Front, Back) as within-subjects factors. Firstly, the effects of the repetition and of the interaction between condition and repetition were not significant ($F(7, 245) = 1.40$; $p = 0.21$ and $F(14, 245) = 1.08$; $p = 0.37$, respectively). Secondly, the condition appeared to be significant ($F(2, 35) = 11.63$; $p < 0.001$). A post-hoc test highlighted significant differences between the proximal pointing method and the other methods (P vs. H: $p = 0.004$; P vs. G: $p < 0.001$), whereas no significant differences could be observed between the head pointing and the gun pointing ($p = 0.58$). The proximal pointing condition showed, on average, larger horizontal errors than the other conditions.

In Figure 5 (left subfigure), a progressive increase of the unsigned horizontal errors from frontal to rear target source locations can be observed for all conditions. According to the ANOVA, the effect of the hemifield was shown to be significant ($F(1, 35) = 167.95$; $p < 0.001$). A significant interaction effect between condition and hemifield was also observed ($F(2, 35) = 6.37$; $p = 0.004$) and a post-hoc test revealed that the significantly greater rate of errors for the proximal pointing condition only appears for targets located in the back hemifield.

The right subfigure in Figure 5 depicts the systematic bias between the target directions and the reported directions. A repeated measures ANOVA was performed on the horizontal bias (*i.e.* the signed horizontal errors), using similar between-subjects and within-subjects factors. A significant effect of the repetition was noticed ($F(7, 245) = 2.93$; $p = 0.006$), which is linked to a slight reduction of the horizontal bias over repetitions of about 2°. The interaction effect between condition and repetition was found not to be significant ($F(14, 245) = 0.69$; $p = 0.78$). Here again, the ANOVA revealed a significant effect of the condition ($F(2, 35) = 9.91$; $p < 0.001$), and a post-hoc test showed that the results from the proximal method significantly differ from the results of other methods (P vs. H: $p = 0.005$, and P vs. G: $p < 0.001$). The hemi-

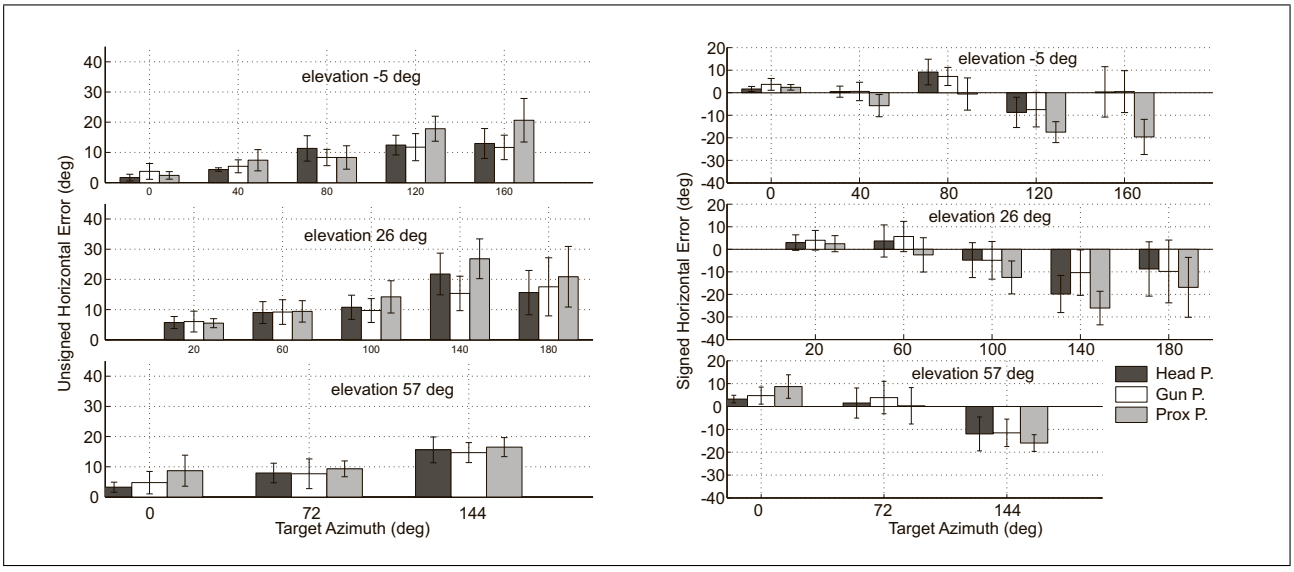

Figure 5. Unsigned (left) and signed (right) horizontal localization errors for each target elevation as a function of the target's azimuth position. Data were averaged over repetitions and subjects for each condition. For improved readability, data are combined for the left and right hemispaces. Vertical black lines show the standard deviation, and thus the dispersion of subject responses about the mean errors. It is important to note that this figure only gives a global overview over the distribution and magnitude of errors as the different averages were not taken over the same number of observations (*e.g.*, the average over the left and right hemispaces is not applied to loudspeakers located on the median plane).

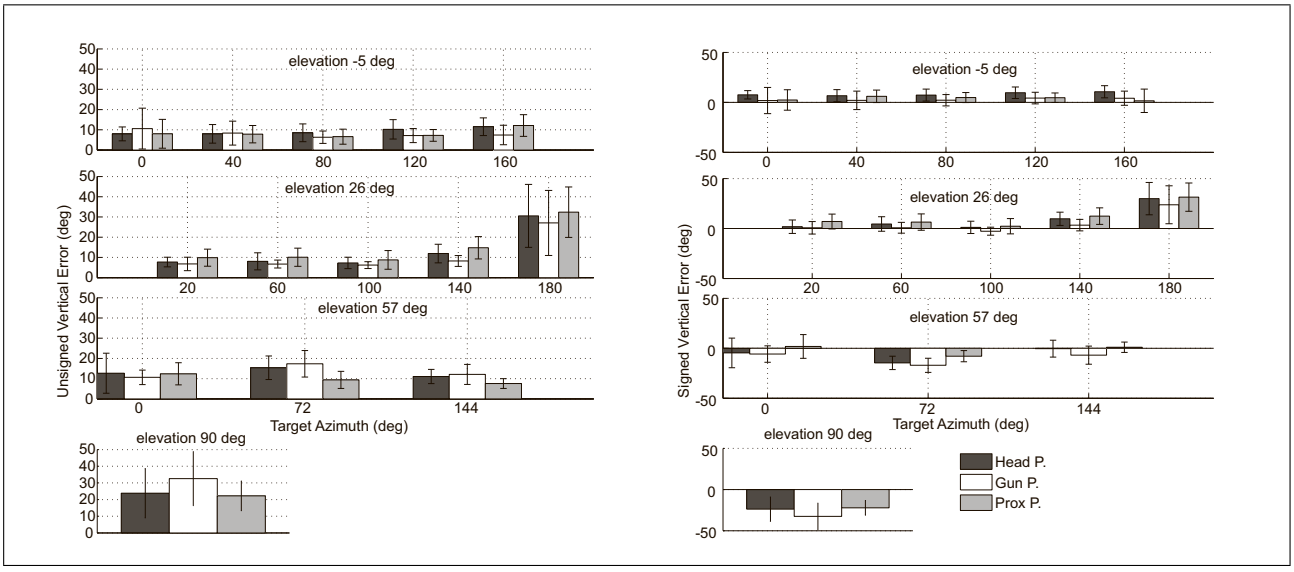

Figure 6. Unsigned (left) and signed (right) vertical localization errors for each target elevation as a function of the target's azimuth position. See legend of Figure 5 for more details.

field has a significant effect on the signed horizontal errors ($F(1, 35) = 281.48$; $p < 0.001$), as well as the interaction effect between condition and hemifield ($F(2, 35) = 6.90$; $p = 0.003$). According to a post-hoc test, the larger systematic horizontal bias of the proximal pointing method compared to the other methods only holds for targets in the back hemifield. Figure 7 shows almost no systematic bias for the proximal pointing method for frontal targets, whereas a positive systematic bias can be observed for the other methods. For rear targets, a negative systematic bias was found for all conditions, which is significantly higher for the proximal pointing method.

3.3.3. Vertical errors

A repeated measures ANOVA was performed on the unsigned vertical errors, with the pointing condition (H, G, P) as between-subjects factor, and the repetition (1-8) and the elevation range (Mid, High) as within-subjects factors. Neither the effect of the repetition ($F(7, 245) = 1.67$; $p = 0.12$) nor the interaction between condition and repetition ($F(14, 245) = 1.50$; $p = 0.11$) were found to be significant. Unsigned vertical errors do not differ significantly between the conditions ($F(2, 35) = 1.35$; $p = 0.27$), but they increase significantly from mid to high eleva-

tions ($F(1, 35) = 17.37$; $p < 0.001$). The interaction effect between condition and elevation range appeared to be marginally significant ($F(2, 35) = 4.68$; $p = 0.02$). According to the post-hoc test, and as shown in Figure 8 (left subfigure), the significant increase of the unsigned vertical errors between mid and high elevations only holds for the gun pointing method ($p < 0.001$). The proximal pointing method shows almost no difference in vertical pointing accuracy. Moreover, the proximal pointing method is more accurate in vertical pointing at high elevations than the gun pointing method ($p = 0.02$).

The same analysis was performed on the signed vertical errors. No significant effect of the repetition could be found ($F(7, 245) = 1.55$; $p = 0.15$); however, the interaction effect between repetition and condition appeared to be significant ($F(14, 245) = 4.69$; $p < 0.001$). This is linked to the fact that we observed symptoms of fatigue for the gun pointing condition. Subjects gradually decreased the pointing height with ongoing repetitions (about 6° in total). The effect of the condition was found to be marginally significant ($F(2, 35) = 3.95$; $p = 0.03$): a global negative vertical bias was noticed for the gun pointing method, whereas the average bias remained close to zero for the head pointing and proximal pointing. The dependence of the signed vertical errors on the elevation range was found to be significant ($F(1, 35) = 322.93$; $p < 0.001$): an average positive bias of +6° was found for mid elevations as well as an averaged negative bias of −10° for high elevations. In other terms, the vertical range of the pointed elevation is compressed compared to the actual elevation range of the target sources. The interaction effect between condition and elevation range is also marginally significant ($F(2, 35) = 3.64$; $p = 0.037$): a marginally significant difference in vertical pointing accuracy is noticed at the high elevation range between the proximal and gun pointing methods ($p = 0.015$). As shown in Figure 8 (right subfigure), the compression effect is lower for the proximal pointing than for the two other conditions.

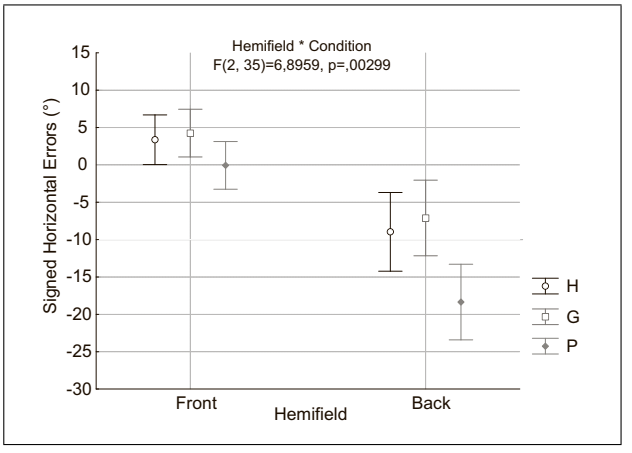

Figure 7. Signed horizontal errors averaged over subjects and repetitions. Results are presented separately for front and back hemifields and for each experimental condition (H:head, G:gun, P:proximal).

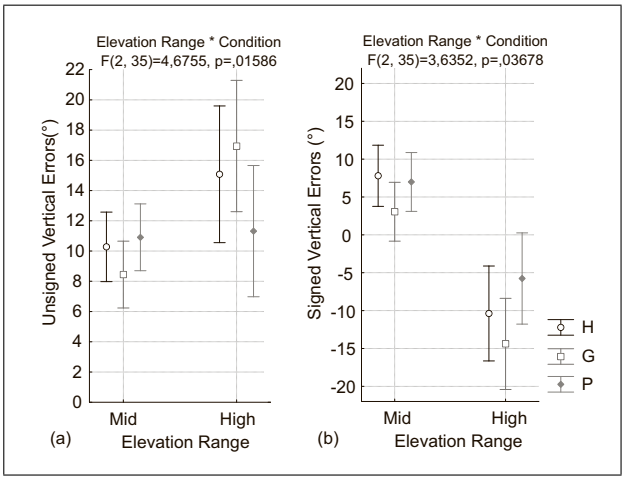

Figure 8. Unsigned (a) and signed (b) vertical errors as a function of the target elevation range. "Mid" includes the target elevations −5° and +26°, and "High" is related to 57° and 90°. Data are displayed separately for each condition (H:Head, G:Gun, P:Proximal).

3.3.4. Dispersion of responses

Following [1], the dispersion parameter κ^{-1} was investigated in order to summarize the consistency of the subjects' responses, regardless of the horizontal and vertical dimensions. The parameter κ is estimated from the length of the sum of all unit-length vectors corresponding to the directions pointed by a given subject for a given target.

$$\kappa = (N - 1)^2 / N(N - R),$$

where N is the number of repetitions and R is the length of the resultant vector. The dispersion parameter κ^{-1} varies from 0 (no dispersion) to 1 (full dispersion, *i.e.* the reported directions are randomly distributed on the sphere).

A repeated measures ANOVA was performed on the dispersion, with the pointing condition (H, G, P) as between-subjects factor, and the target elevation (−5°, 26° and 57°) and the hemifield (Front, Back) as within-subjects factors. The results are depicted in Figure 9. The condition effect was found not to be significant ($F(2, 35) = 0.75$; $p = 0.48$). The dispersion appeared to be significantly larger for rear targets than for frontal targets ($F(1, 35) = 40.05$; $p < 0.001$), and to significantly increase from low to high elevations ($F(2, 70) = 16.94$; $p < 0.001$). However, a more detailed analysis of the interactions reveals that the results of the proximal pointing are more homogeneous. A marginally significant interaction effect between condition and target elevation was found ($F(4, 70) = 2.72$; $p = 0.04$). The associated post-hoc test revealed, that the significant increase of the dispersion between frontal and rear targets only holds for the head and gun pointing methods. Moreover, the interaction effect between condition, hemifield and target elevation appeared to be significant ($F(4, 70) = 9.43$; $p < 0.001$). According to a post-hoc test, the proximal pointing method shows a significantly lower dispersion at an elevation of 57° in

the rear hemifield compared to the gun pointing method ($p = 0.002$). Note that the target elevation of 90° was excluded from the current statistical analysis because it doesn't belong to any hemifield. Then, in order to evaluate the dispersion effect for the different pointing methods at the highest elevation, an ANOVA was performed with condition as between-subjects factor and target elevation as within-subjects factor (including elevation 90°). However, no significant difference between pointing methods emerged from the associated post-hoc test at this elevation.

3.3.5. Response time

We further compared the different pointing methods with respect to the response time, *i.e.* the time delay between the playback of the stimulus and the (validated) response of a participant. In a localization test with open loop conditions (*i.e.* where the stimulus playback is stopped before the participant points at the target), the response time cannot be considered as a relevant indicator of the localization performance. The observed time differences will be solely determined by the motor task itself. However, the response time is an important parameter that may be taken into account when planning and designing a localization experiment. For this reason, this parameter was studied in the present work. A repeated measures ANOVA was performed on the response time, with the pointing condition (H, G, P) as between-subjects factor, and the repetition (1-8) and the absolute target angle as within-subjects factors. The absolute target angle refers to the magnitude of the angular distance between the frontal direction (*i.e.* the resting position) and the target position on the sphere. Firstly, the analysis results revealed a marginally significant effect of the pointing condition on the response time ($F(2, 35) = 5.46$; $p = 0.009$): we obtained average response times of 3.5 s, 3.1 s, and 2.5 s for the head pointing, gun pointing, and proximal pointing methods, respectively. Secondly, the response time decreases significantly over the repetitions ($F(7; 245) = 15.22$; $p < 0.001$). However, according to the interaction effect between condition and repetition ($F(14, 245) = 2.80$; $p < 0.001$), this was only observed for the proximal and the head pointing methods (see Figure 10). No improvement on the reaction time can be noticed for the gun pointing method.

As expected, the absolute target angle has a significant effect on the response time ($F(12, 420) = 43.54$; $p < 0.001$): there is an increase in the response time as a function of the absolute target angle. According to the interaction effect between condition and absolute target angle ($F(24, 420) = 6.40$; $p < 0.001$), this effect is well pronounced for the gun pointing method and, in particular, for the head pointing method, whereas no clear trend could be observed for the proximal condition.

3.3.6. Effect of the hand used to point

For the proximal pointing method, we examined the responses given with the participant's dominant or non-dominant hands. Note that the analysis considers only one pointing condition, which eliminates the between-subjects factor. A two-way repeated measures ANOVA

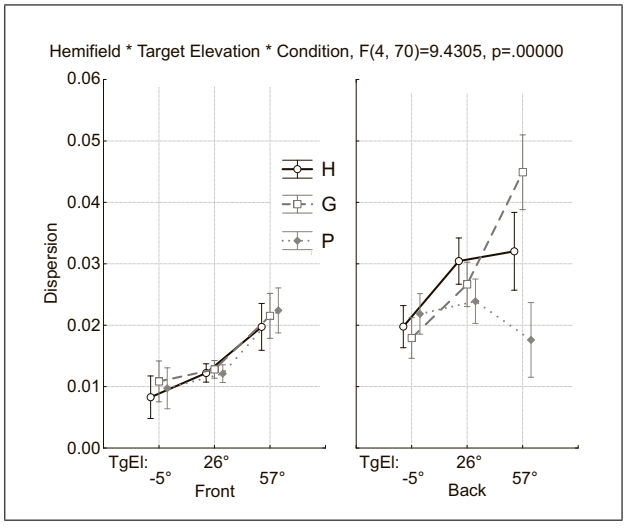

Figure 9. Response dispersion κ^{-1} as a function of the experimental condition (H:Head, G:Gun, P:Proximal), hemifield (Front, Back), and target elevation (−5°, 26°, 57°).

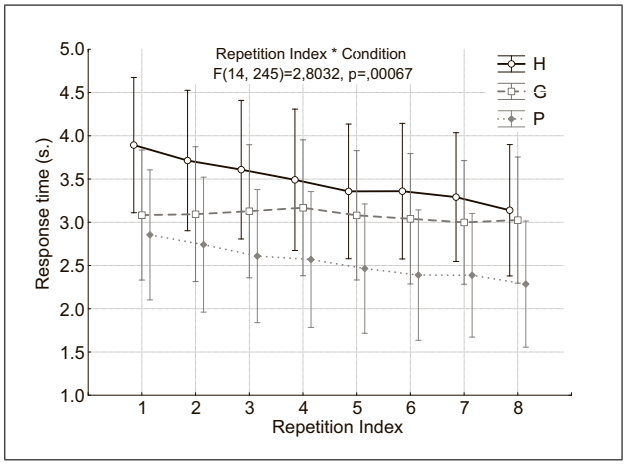

Figure 10. Evolution of the response time ($s.$) for the same target source over the repetitions. The data are displayed separately for each experimental condition (H:Head, G:Gun, P:Proximal).

was performed with the lateral dominance (Dominant, Non-dominant) and the hemifield (Front, Back) as within-subjects factors. The effect of the lateral dominance on the signed and unsigned vertical errors is not significant ($F(1, 12) = 0.24$; $p = 0.64$ and $F(1, 12) = 0.37$; $p = 0.55$, respectively). Therefore, the analysis was focused on the horizontal dimension and showed a significant effect of the lateral dominance on the unsigned horizontal errors ($F(1, 12) = 14.78$; $p = 0.002$) as well as of the hemifield ($F(1, 12) = 103.4$; $p < 0.001$). The interaction effect between those factors was found not to be significant ($F(1, 12) = 2.06$; $p = 0.18$). Regarding the signed horizontal errors, the effect of the hemifield is significant ($F(1, 12) = 188.6$; $p < 0.001$), but the lateral dominance effect ($F(1, 12) = 4.34$; $p = 0.06$) and the interaction effect between the two factors ($F(1, 12) = 0.81$; $p = 0.39$) are not significant. To summarize, the choice of the hand affects the horizontal pointing accuracy only, regardless of

the hemifield of the sound source. The averaged unsigned horizontal error is 12.3° for the dominant hand compared to 14.8° for the non-dominant hand.

4. Discussion

The primary goal of this study is to compare the proximal pointing method to more conventional pointing methods, such as the head and gun pointing. The localization accuracy was analyzed on a relatively small number of spatial directions (mostly in the upper hemisphere). It is thus essential to verify the results by comparing them to the results of previously published reference studies.

The study of Makous and Middlebrooks [12] has been selected for several reasons: Firstly, because it is based on a head pointing task for reporting the localization of real sound sources distributed over a sphere. Secondly, some of the tested positions are very close to those used in this study. Thirdly, the definition of localization errors is identical. Finally, the results of Makous and Middlebrooks' study were grounded on observations gathered over a relatively dense grid of tested directions and after a long training of the participants. For similar reasons, the results from Gilkey *et al.* [8] were used for comparisons, although they used an exocentric reporting method (GELP) in their study.

Figure 11 presents the unsigned horizontal and vertical errors observed in the two reference studies compared to the different conditions of the present experiment. A very similar trend is observed for all studies with increasing errors in the back hemifield, especially from azimuth 110° to 160° for the unsigned horizontal errors as already pointed out in [11]. Unsigned errors found in the study of Gilkey *et al.* are systematically greater than those of Makous and Middlebrooks, which confirms the advantage of egocentric pointing methods compared to exocentric ones [8]. On average, the results gathered in the different conditions of the present experiment are lying in between the two reference methods. In particular, the unsigned horizontal errors observed for the gun pointing and the head pointing conditions perfectly match the corresponding curve of the study of Makous and Middlebrooks [12], except for the head pointing condition at 140°. This result provides confidence to the present study, especially when considering the very limited training of our participants (10 directions, completed in approximately 2 minutes) compared to the study of Makous and Middlebrooks (1000 directions, with a total duration of the training period up to two hours).

The main exception to this behavior are the higher unsigned horizontal errors for targets located in the back hemifield (starting from azimuth 120°) observed for the proximal method, as previously noticed.

With regard to signed horizontal errors, an average positive bias was found in the frontal hemifield as well as an average negative bias in the back hemifield (see section 3.3.2). This general extra-lateralization of the reported directions has already been underlined in [11, 34, 10]. Moreover, the particular positive vertical bias observed for

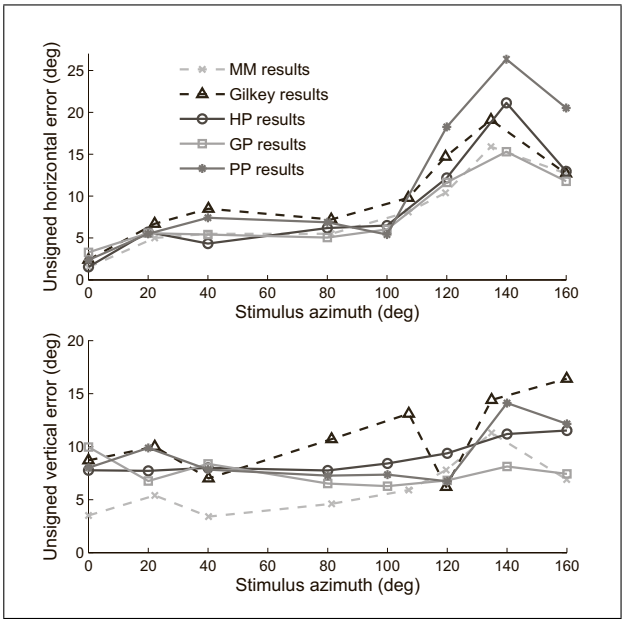

Figure 11. Unsigned horizontal errors (top) and unsigned vertical errors (bottom) as a function of target azimuth (in degrees). Data from the present study are compared to the results from Makous and Middlebrooks' open-loop experiment [12] (based on the head pointing method) and the results from Gilkey *et al.* [8] (using the GELP technique). All data were averaged over the left and right hemispaces and confusions were treated in the same way (*i.e.* removed from the data).

highly elevated sources located in the rear has been also pointed out by Oldfield and Parker [11].

Regarding the confusions, the important reversal rate observed for targets located at high elevations in the rear hemifield, is consistent with literature [12, 34]. More generally, in the present study higher confusion rates were noticed in the rear hemifield, which corresponds to a prevailing occurrence of back-to-front confusions. This observation is in accordance with a localization test conducted by Bronkhorst [36], in which he was using a reporting method similar to that one used in this study (*i.e.* head pointing and blindfolded subjects). Bronkhorst shows a predominant occurrence of front-to-back and diagonal confusions for virtual sound sources (*i.e.* the playback of binaural signals on headphones), but a predominant occurrence of back-to-front confusions for real sound sources.

Another comparison can be done with the results reported in Djelani *et al.* [9], who compared a finger pointing method with both a head pointing and a GELP pointing method, although they only provide global values of the absolute angle error and the dispersion coefficient κ^{-1} averaged over 12 directions and eight subjects. The higher accuracy provided by the egocentric methods (finger and head pointing) against the exocentric GELP technique is in good agreement with the observations reported in this article. However, some differences can also be noticed. The averaged dispersion coefficient κ^{-1} measured in their study is much higher (between 0.04 and 0.08 depending on the method) than what we observed in the present study

(between 0.01 and 0.04 according to the pointed directions and methods). A possible explanation is that their study was based on virtual sound sources rendered over headphones, whereas real sources were used in this study. Finally, we found larger horizontal errors in the back hemifield for the proximal pointing method compared to the head pointing method, which was apparently not the case in the study of Djelani *et al.* [9]. However, this may originate from an important difference in the pointing task compared to the present study. They used a closed-loop condition (*i.e.* continuous stimulus during the pointing task) for the GELP technique and the finger pointing method, whereas an open-loop condition (*i.e.* stimulus stopped before pointing) was used for the head condition. In contrast, we used an open-loop condition for all methods in order to allow for a more consistent comparison of the data. The choice of a continuous stimulus was excluded in our experiment since it would mean that under the head pointing and gun pointing conditions, the localization task would always end up with the sound target located in front of the participant [9]. Nevertheless, results of Djelani *et al.* give an indication that the relative weakness of the proximal method observed in the back hemifield could probably be overcome or at least reduced in a closed-loop condition as expected from other auditory pointing tasks [37]. It could as well benefit from a longer training.

According to the observations made in Section 3.3, the proximal method shows a lack of accuracy in the horizontal dimension for rear locations close to the median plane. The negative horizontal bias observed for these locations suggests a motor task difficulty to indicate directions in the back hemifield (task described in [9] as "uncomfortable"). In contrast, for the head and gun pointing methods, the horizontal error increase observed at rear locations may rather come from the inaccuracy induced by the large body rotations.

In comparison with the proximal pointing method, the vertical pointing accuracy of the gun pointing method is lower at high elevations, and the dispersion of judgments for rear targets located at high elevations is higher. This observation reflects the difficult motor task that combines large arm movements with large body rotations. There is also a tendency that the subjects point too low when reporting the perceived sound position with the head or a gun. This is probably caused by the physical strain experienced when pointing to high elevations with these pointing methods [9].

It was shown that the response time of the proximal pointing method is 30% and 20% shorter than that of the head and gun pointing methods, respectively. Although this advantage was partly compensated in the present study by the additional time needed for the complex repositioning procedure (note that the percentage is divided in half when considering the repositioning time), it can be taken into account when designing localization experiments. Shorter response times help to reduce the total duration of experiments, which limits the fatigue of the participants. With shorter response times, more responses can be collected with the same test duration. This strengthens the statistical results.

5. Conclusion

Egocentric pointing tasks for reporting a perceived sound source position involve moving different body segments, and the motor control may affect the localization performance. This study shows similar trends for the spatial distribution of localization errors for the three methods under investigation: a progressive increase of the horizontal errors with the azimuth, a negative horizontal bias for rear locations, and a negative vertical bias for sources located at high elevations. It further highlights some limitations specific to each pointing method. For sources located at high elevations, the head and gun pointing methods (which both involve a rotation of the whole body) exhibit a larger dispersion and a larger vertical bias than the other method. For sources located in the rear hemifield close to the median plane, the proximal pointing method presents a larger horizontal bias. However, this method has several practical advantages and interesting improvement perspectives for future 3D sound localization experiments. The proximal pointing method's shorter response times allow for collecting a larger number of judgments for a given test duration. In contrast to other pointing methods, it is also well suited for being used in a closed-loop condition, i.e., where the sound stimuli remain switched on during the pointing task. Exploiting the auditory-motor loop, combined with an extensive training, may help to improve the pointing accuracy, especially for rear locations and for responses collected with the non-dominant hand. The proximal pointing method may also be extended to assess the perceived distance of virtual sound sources played back over headphones, *e.g.*, to investigate the externalization phenomenon in binaural synthesis.

Acknowledgement

This work was funded by the French FUI project BiLi ("Binaural Listening", www.bili-project.org, FUI-AAP14) with support from Cap Digital Paris Region.

References

[1] F. L. Wightman, D. J. Kistler: Headphone simulation of free-field listening. ii: Psychophysical validation. The Journal of the Acoustical Society of America **85** (1989) 868–878.

[2] E. Wenzel, M. Arruda, D. Kistler, F. Wightman: Localization using nonindividualized head-related transfer functions. The Journal of the Acoustical Society of America **94** (1993) 111–123.

[3] M. J. Evans: Obtaining accurate responses in directional listening tests. 104th Audio Engineering Society Convention, May 1998.

[4] V. Larcher: Techniques de spatialisation des sons pour la réalité virtuelle. Dissertation. Université de Pierre et Marie Curie, Paris VI, 2001.

[5] J.-M. Pernaux, M. Emerit, R. Nicol: Perceptual evaluation of binaural sound synthesis: the problem of reporting local-

ization judgments. 114th Audio Engineering Society Convention, Amsterdam, March 2003.

[6] D. R. Begault, E. M. Wenzel, M. R. Anderson: Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J. Audio Eng. Soc **49** (2001) 904–916.

[7] M. Schoeffler, S. Westphal, A. Adami, H. Bayerlein, J. Herre: Comparison of a 2d- and 3d-based graphical user interface for localization listening tests. Proc. of the EAA Joint Symposium on Auralization and Ambisonics, April 2014.

[8] R. Gilkey, M. Good, M. Ericson, J. Brinkman, J. Stewart: A pointing technique for rapidly collecting localization responses in auditory research. Behavior Research Methods, Instruments, & Computers **27** (1995) 1–11.

[9] 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** (November 2000) 1046–1053.

[10] P. Majdak, M. 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.

[11] S. R. Oldfield, S. P. A. Parker: Acuity of sound localisation: a topography of auditory space. ii. pinna cues absent. Perception **13** (1984) 601–617.

[12] J. C. Makous, J. C. Middlebrooks: Two-dimensional sound localization by human listeners. The Journal of the Acoustical Society of America **87** (1990) 2188–2200.

[13] V. Best, S. Carlile, C. Jin, A. van Schaik: The role of high frequencies in speech localization. The Journal of the Acoustical Society of America **118** (2005) 353–363.

[14] M. Frank, L. Mohr, A. Sontacchi, F. Zotter: Flexible and intuitive pointing method for 3-d auditory localization experiments. Audio Engineering Society Conference: 38th International Conference: Sound Quality Evaluation, June 2010.

[15] S. Perrett, W. Noble: The contribution of head motion cues to localization of low-pass noise. Perception & Psychophysics **59** (1997) 1018–1026.

[16] R. L. Martin, K. I. McAnally, M. A. Senova: Free-field equivalent localization of virtual audio. J. Audio Eng. Soc **49** (2001) 14–22.

[17] H. Wierstorf, A. Raake, S. Spors: Localization of a virtual point source within the listening area for wave field synthesis. 133th Audio Engineering Society Convention, Oct 2012.

[18] V. Tabry, R. J. Zatorre, P. Voss: The influence of vision on sound localization abilities in both the horizontal and vertical planes. Frontiers in Psychology **4** (2013).

[19] T. Ashby, R. Mason, T. Brookes: Head movements in three-dimensional localization. 134th Audio Engineering Society Convention, May 2013.

[20] P. Paukner, M. Rothbucher, K. Diepold: Sound localization performance comparison of different hrtf-individualization methods. Dissertation. Technische Universität München, Lehrstuhl für Datenverarbeitung, April 2014.

[21] B. Seeber: A new method for localization studies. Acta Acustica united with Acustica **83** (1997) 1–2.

[22] B. Seeber: Untersuchung der auditiven lokalisation mit einer lichtzeigermethode (development and test of a new method to study auditory localization and application of it to virtual acoustics). Dissertation. Technical University of Munich, Munich, 2003.

[23] 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 **4-99** (2013) 642–657.

[24] V. Pulkki, T. Hirvonen: Localization of virtual sources in multichannel audio reproduction. IEEE Transactions on Speech and Audio Processing **13** (2005) 105–119.

[25] E. H. A. Langendijk, A. W. Bronkhorst: Contribution of spectral cues to human sound localization. The Journal of the Acoustical Society of America **112** (2002) 1583–1596.

[26] J. Blauert: Spatial hearing: The psychophysics of human sound localization. MIT Press, 1997.

[27] J. Vliegen, A. J. Van Opstal: The influence of duration and level on human sound localization. The Journal of the Acoustical Society of America **115** (2004) 1705–1713.

[28] P. M. Hofman, A. J. Van Opstal: Spectro-temporal factors in two-dimensional human sound localization. The Journal of the Acoustical Society of America **103** (1998) 2634–2648.

[29] B. F. G. Katz, G. Parseihian: Perceptually based head-related transfer function database optimization. The Journal of the Acoustical Society of America **131** (2012) EL99–EL105.

[30] M. Morimoto, H. Aokata: Localization cues of sound sources in the upper hemisphere. The Journal of the Acoustical Society of Japan (E) **5** (1984) 165–173.

[31] M. Aytekin, C. F. Moss, J. Z. Simon: A sensorimotor approach to sound localization. Neural Comput. **20** (2008) 603–635.

[32] I. Viaud-Delmon, O. Warusfel: From ear to body: the auditory-motor loop in spatial cognition. Frontiers in Neuroscience **8** (2014) 283.

[33] F. L. Wightman, D. J. Kistler: Resolution of front–back ambiguity in spatial hearing by listener and source movement. The Journal of the Acoustical Society of America **105** (1999) 2841–2853.

[34] S. Carlile, P. Leong, S. Hyams: The nature and distribution of errors in sound localization by human listeners. Hearing Research **114** (1997).

[35] V. E. Johnson: Revised standards for statistical evidence. Proceedings of the National Academy of Sciences **110** (2013) 19313–19317.

[36] A. W. Bronkhorst: Localization of real and virtual sound sources. The Journal of the Acoustical Society of America **98** (1995) 2542–2553.

[37] E. O. Boyer, B. M. Babayan, F. Bevilacqua, M. Noisternig, O. Warusfel, A. Roby-Brami, S. Hanneton, I. Viaud-Delmon: From ear to hand: the role of the auditory-motor loop in pointing to an auditory source. Frontiers in Computational Neuroscience **7** (2013).