KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	David Pelegrín-García, Enzo De Sena, Toon van Waterschoot, Monika Rychtáriková, and Christ Glorieux **Localization of a virtual wall by means of active echolocation by untrained sighted persons** *Applied Acoustics*, vol. 139, pp. 82-92, Oct. 2018.
Archived version	Author manuscript: the content is identical to the content of the submitted paper, but without the final typesetting by the publisher
Published version	https://doi.org/10.1016/j.apacoust.2018.04.018
Journal homepage	https://www.sciencedirect.com/science/journal/0003682X
Author contact	toon.vanwaterschoot@esat.kuleuven.be *+ 32 (0)16 321788*
IR	ftp://ftp.esat.kuleuven.be/pub/SISTA/vanwaterschoot/reports/18-148.pdf

(article begins on next page)

Applied Acoustics 139 (2018) 82–92

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Localization of a virtual wall by means of active echolocation by untrained sighted persons

David Pelegrín-García[a,b,*], Enzo De Sena[b,d], Toon van Waterschoot[b], Monika Rychtáriková[c,e], Christ Glorieux[a]

[a] Laboratory of Acoustics, Division Soft Matter & Biophysics, Dept. Physics & Astronomy, KU Leuven, Belgium
[b] STADIUS-ESAT, Department of Electrical Engineering, KU Leuven, Belgium
[c] KU Leuven, Faculty of Architecture, Hoogstraat 51, 9000 Gent/Paleizenstraat 65, 1030 Brussel, Belgium
[d] Institute of Sound Recording, University of Surrey, Guildford GU2 7XH, UK
[e] STU Bratislava, Faculty of Civil Engineering, Dept. Architecture, Radlinskeho 11, Bratislava 810 05, Slovak Republic

ARTICLE INFO

Keywords:
Human echolocation
Virtual acoustics
Masking
Sound localization

ABSTRACT

The active sensing and perception of the environment by auditory means is typically known as echolocation and it can be acquired by humans, who can profit from it in the absence of vision. We investigated the ability of twenty-one untrained sighted participants to use echolocation with self-generated oral clicks for aligning themselves within the horizontal plane towards a virtual wall, emulated with an acoustic virtual reality system, at distances between 1 and 32 m, in the absence of background noise and reverberation. Participants were able to detect the virtual wall on 61% of the trials, although with large differences across individuals and distances. The use of louder and shorter clicks led to an increased performance, whereas the use of clicks with lower frequency content allowed for the use of interaural time differences to improve the accuracy of reflection localization at very long distances. The distance of 2 m was the most difficult to detect and localize, whereas the furthest distances of 16 and 32 m were the easiest ones. Thus, echolocation may be used effectively to identify large distant environmental landmarks such as buildings.

1. Introduction

The sense of hearing provides relevant information for spatial perception [1]. Audition is particularly important for blind people [2], who lack visual stimuli to build spatial representations of their surroundings. Some blind people have learnt to echolocate [3], i.e. to detect and localize obstacles and environmental features based on the reflections they produce in response to self-generated sounds (active echolocation), typically oral clicks [4], or even to ambient noise [5] (passive echolocation). Echolocation, initially called facial vision because it was believed that sensation arose from pressure sensors on the skin [6], is in fact a purely auditory phenomenon [7]. Sound reflections, or echoes (if perceived as a separate event from the direct sound), arrive at an echolocator with variable attenuation, delay, interaural level and time differences and spectral cues which they exploit [8,9] to infer information about the distance [10,11], angular location [12], size [13], shape [14] and texture [13] of the boundary at which they were generated. At short distances, it is possible to use coloration cues (i.e. a change in the tonal character), arising from the interaction of direct and reflected sounds [15] to detect the latter. However, it is highly unlikely to use coloration cues to detect reflections with delays longer than 10 ms (or arising from obstacles at distances further than 2 m) [16,17]. Localization of reflections is particularly precise due to a partial inhibition of the precedence effect—a collection of phenomena that makes it possible to localize an original sound source in a room despite all reflections [18]—during echolocation [19,20]. This technique represents an active perception mode [21], meaning that the perception of auditory space integrates the auditory sensation at different positions and orientations with the vestibular and proprioceptive feedback [22], and thus head movements are crucial [23] for effective mobility and related tasks like shape perception [14]. In this way, echolocation contributes to enhance the auditory spatial localization of blind people [24].

In addition, echolocation has benefits on the independence of functional echolocators (i.e. people who use echolocation in daily life), namely better mobility in unfamiliar places and access to better salaried jobs [25]. Thus, echolocation is highly relevant for the rehabilitation of persons who have lost sight. Despite its benefits, echolocation is not yet a widespread technique and much research is devoted to understanding

* Corresponding author currently at: Widex A/S, Nymøllevej 6, 3540 Lynge, Denmark.
E-mail address: david.pelegringarcia@kuleuven.be (D. Pelegrín-García).

https://doi.org/10.1016/j.apacoust.2018.04.018
Received 6 September 2017; Received in revised form 4 April 2018; Accepted 13 April 2018
Available online 24 April 2018
0003-682X/ © 2018 Elsevier Ltd. All rights reserved.

the degree to which early and late blind people can profit from echoic information [26]. In the present study, we focus on sighted persons without prior experience in echolocation, and as such their performance can be considered similar to that of potential candidates in rehabilitation programs of orientation and mobility.

Acoustic Virtual Reality (AVR) systems that account for head orientation are regarded useful for the acquisition of auditory space maps [27], for evoking sensations arising in echolocation [10,22] and for the conduction of psychoacoustic tests [20]. When using an AVR system for the conduction of psychoacoustic or active echolocation tests, non-auditory cues such as wind, temperature, or other tactile cues are decoupled from the auditory stimulation. In addition, ambient noise, which could lead to effective passive echolocation in reality, can be minimized or studied separately from active echolocation. In real life, these cues are an integral part of the environment perception and contribute to the detection of obstacles. Furthermore, arbitrarily large surfaces can be simulated with the AVR system at any position. Because large surfaces reflect sound more efficiently than small obstacles, and because the background noise in an AVR is negligible when used in a silent anechoic room, the range of distances where obstacles can be detected with echolocation is expected to be much larger using an AVR system than in other real life experiments.

Sighted participants have been able to learn basic echolocation tasks using an AVR system [11,20]. Therefore, this kind of systems shows a potential to explore effective learning strategies in echolocation, gain further knowledge about its psychophysical mechanisms, and study the relationship between performance and the self-generated oral signals [11,17]. When using an AVR system, the simulated acoustic condition is defined by the Oral-Binaural Room Impulse Response (OBRIR) [28], which contains information on the propagation of sound, including all relevant reflections between a reference point close to the mouth and the entrance of the closed left and right ear canals.

In a previous study [17], we used a static AVR system to determine the detection thresholds for a single reflection up to distances of 16 m, by artificially varying the strength of the reflection in relation to the strength of the direct sound. This relation between strengths was quantified by means of the reflected-to-direct level difference (RDLD) [29].[1] The results closely followed the signs of forward masking—a characteristic of the human auditory system for which one sound can render inaudible another fainter sound closely delayed. When using oral clicks, thresholds improved with increasing distance due to the increased temporal separation between direct and reflected sound, which caused less forward masking. Louder signals resulted in lower (i.e. better) detection thresholds due to the non-linear behavior of masking [17]. These results suggest that it might be possible to use echolocation in a range of distances longer than reported in several previous studies of echolocation with sighted participants in real settings (e.g. 55–65 cm [30], 62–130 cm [31] or less than 1.8 m [7,32]) or in laboratory settings with sighted and blind participants (up to 2 m [9,33]), when using larger obstacles in an otherwise reflection-free environment with low background noise. This was already explored by Schrnich et al. [11] by measuring distance discrimination thresholds in echolocation up to 6.8 m using an AVR. In the present study, we explore echolocation for detection and localization of large obstacles in a range of distances broader than in previous studies, from 1 m to 32 m.

In [17], the task was to discriminate one out of three intervals which was different from the other two. A question that arises is whether untrained sighted individuals would also be able to use echolocation in a more realistic task than the one in [17]. In the present study, we develop and use an AVR system that accounts for head rotations in the horizontal plane. We use this system to simulate a simple outdoor echolocation exercise in which a person has to orient herself towards a simulated large wall (or building) at different distances on a reflecting floor, without competing reflections from other surfaces nor background noise. Thus, participants in the present experiments had firstly and implicitly to detect the reflecting object, and only then, to identify the direction of arrival of the reflection.

Beyond knowing whether untrained sighted individuals can echolocate, we aim at determining the most difficult distance conditions to detect and localize a large obstacle (e.g. a wall) and how performance relates to features of the emitted signals. This knowledge can serve as an inspiration for determining good clicking strategies based on parameters such as intensity, duration, bandwidth and frequency content, and in addition, to point out the most difficult distance conditions in the training of echolocation. We hypothesize that, in absence of background noise and reverberance, the most difficult conditions are those in which the reflection level of the wall is near the audibility thresholds for a reflection determined in [17]. Easier conditions may be those where the reflection is well above threshold. Moreover, in the distance conditions where the reflection level is near the audibility thresholds, it is expected that localization accuracy decrease [34].

2. Method

2.1. Participants

The 21 participants in the experiments, labeled S1 to S21 hereafter, were 7 female and 14 male sighted persons, between 22 and 48 years old, with normal hearing (HL <20 dB from 250 Hz to 8 kHz following audiometric screening) and without previous experience in echolocation. They participated on a voluntary basis and did not receive any compensation for the participation. Informed consent was given by all participants and ethics approval was granted for this research by the Medical Ethics Committee at UZ KU Leuven (number B322201317883).

2.2. Stimuli

Stimuli were designed to replicate a simple outdoor echolocation exercise in which participants had to identify the direction of a large wall (e.g. a building) and turn towards it. Virtual flat concrete walls (of dimensions 10 m × 10 m) at distances d of 1, 2, 4, 8, 16 and 32 m on a concrete floor (see Fig. 1) of 100 m × 100 m were simulated through streaming convolution of the OBRIRs characterizing these scenarios with the oral clicks produced by the participants themselves. The following sections give more insight on the computation of the OBRIRs and on the actual stimuli generation via streaming convolution.

2.2.1. OBRIR calculation

The reflections of the wall and the floor were simulated with the room acoustics simulation software CATT-Acoustic^{TM}v9.0c. A binaural receiver was placed at the middle point in between the ears, and a source simulating the mouth with the average frequency-dependent directivity pattern of the human voice [35] was placed 0.1 m in front of the receiver and pointed away from it. Such a pattern, on a frequency-band basis, is similar to the one produced by echolocation clicks [36]. The receiver was always pointing towards the source. Both source and receiver were placed at a height of 1.5 m from the floor. In separate calculations, the wall was placed at each of the six different distance conditions (d = 1, 2, 4, 8, 16 and 32 m) from the receiver. These distances were chosen to represent a large range of distances, as in a previous study [17], and were logarithmically spaced because the reflected level decays approximately 6 dB per doubling of distance. At each distance d, simulations were performed for 24 reference orientations θ_0 of source/receiver, always rotating the source position around the fixed receiver, at intervals $\Delta\theta$ of 15°.

The wall and the floor had an absorption coefficient of 0.01 at

[1] Note that the RDLD is unambiguously defined for a single reflection. In the case of multiple reflections, there could be an ambiguous interpretation. On the one hand, the RDLD could be referred to a single reflection of interest by filtering out all other competing reflections. On the other hand, the RDLD could take into account all existing reflections to calculate the strength of the total reflected sound. In this study, we use the latter interpretation.

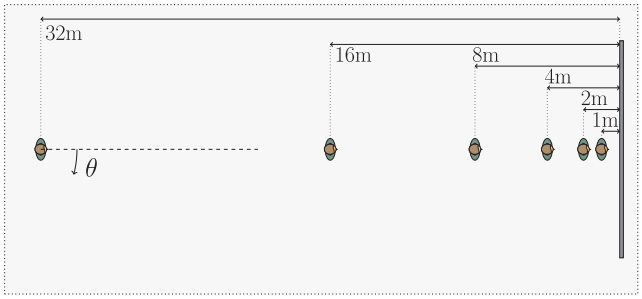


Fig. 1. Simulated concrete wall at distances of 1, 2, 4, 8, 16 and 32 m in front of a participant, who had to align him/herself (dotted arrow) to the direction of the wall (solid arrow), thus minimizing the angular deviation θ.

125 Hz monotonously increasing to 0.05 at 4 kHz. These surfaces had a default scattering coefficient of 10% at all frequencies. The OBRIRs were determined by simulation using algorithm number 2 in TUCT (CATT-Acoustic's calculation engine). A total of 1 million rays were used, and the length of the impulse response was set to 0.5 s. Diffraction was not incorporated in the simulation (disabled option). For binaural output, the HRTF dataset measured at RWTH ITA Aachen with a sampling rate of 44.1 kHz (file *ITA_1_plain_44.dat* included in CATT-Acoustic) was used.

An energy-time representation of the OBRIRs is shown in Fig. 2 as a function of the horizontal angular rotation with respect to the wall normal (defining $\theta = 0$). For each of the six conditions, there are two graphs corresponding to the left and right ears. For the shortest distance in Fig. 2, the reflection visibly does not form a straight line with orientation and its time of arrival is different in both ears; the difference in arrival times at the two ears for a given angle is thus the interaural time difference. Due to the logarithmic scale used in the graph, the effect of head orientation on the interaural time differences cannot be noticed for further wall distances. In any case, the head orientation dependence of interaural level differences is more prominently visible. For example, focusing on the wall reflection at 4 m distance (corresponding to a delay of about 25 ms) in Fig. 2, for user rotations of 45° towards the right (positive angles), the left ear receives more intense sound than the right ear. The opposite happens for rotations of 45° towards the left (negative angles), when the right ear receives more energy than the left ear (because the right ear is closer to the wall and the left ear becomes shadowed by the head). The effect of the floor reflection, whose energy is 17.8 dB lower than that of the direct sound, is always visible at a delay of approximately 9 ms with respect to the direct sound, independently of orientation and wall distance. Although this additional reflection could complicate the interpretation of the results, we decided to keep it, in order to be more representative of real echolocation tasks, where the floor reflection is most commonly present.

In post-processing, the OBRIRs had the direct sound and the first 4.0 ms removed to account for the AVR latency between the microphone excitation and the processed response at the headphones. The resulting OBRIRs at each distance were labeled $\mathbf{h}_d(\theta_0) = [\mathbf{h}_{d,\mathrm{L}}(\theta_0)^T\ \mathbf{h}_{d,\mathrm{R}}(\theta_0)^T]^T$, d = 1 m,2 m,...,32 m, θ_0 = 0°,15°,...,345° and they constituted the "OBRIR library". Each OBRIR had L samples in each channel (left and right); i.e. for the left channel $\mathbf{h}_{d,\mathrm{L}}(\theta_0) = [h^{(0)}_{d,\mathrm{L}}(\theta_0), h^{(1)}_{d,\mathrm{L}}(\theta_0), \ldots h^{(L-1)}_{d,\mathrm{L}}(\theta_0)]^T$ and similarly for the right channel.

One way to characterize the strength of a reflection or group of reflections is by using the Reflected-to-Direct Level Difference (RDLD) [29], which takes into account the spectral characteristics of the reflection, the sound pressure level of a typical oral click $L_{p,\mathrm{click},f}$, and the inverted equal loudness curve of 40 phon according to ISO 226:2003 [37] $T_{W,f}$. On the f-th 1/3rd octave frequency band ($f = 1...N$), RDLD_f is calculated as the difference between the energy level of the reflected sound $L_{\mathrm{me,refl},f}$ and that of the direct sound $L_{\mathrm{me,dir},f}$. Then

$$\mathrm{RDLD} = L_{\mathrm{me,refl}} - L_{\mathrm{me,dir}} = 10\log_{10}\left(\frac{\sum_{f=1}^{N} 10^{\frac{L_{p,\mathrm{click},f}+\mathrm{RDLD}_f+T_{W,f}}{10}}}{\sum_{f=1}^{N} 10^{\frac{L_{p,\mathrm{click},f}+T_{W,f}}{10}}}\right). \quad (1)$$

The resulting distance-dependent RDLD values of the OBRIRs with the source aligned towards the wall (thus identical for left and right channels), postprocessed to exclude the first order reflection from the floor, are shown in Fig. 3. These values were calculated in an approximate way, assuming that the direct sound was the one provided by CATT Acoustic (which does not take into account diffraction effect around the head in the propagation from the mouth to the ears), and then applying an offset to all calculated values so that the RDLD for the wall at 1 m would be identical to that calculated at an infinite 100% reflecting wall [17] plus 3 dB to account for the second order reflection at the floor. For the sake of comparison, the RDLD detection thresholds for a single reflection using self-generated oral clicks [17] are also shown in Fig. 3.

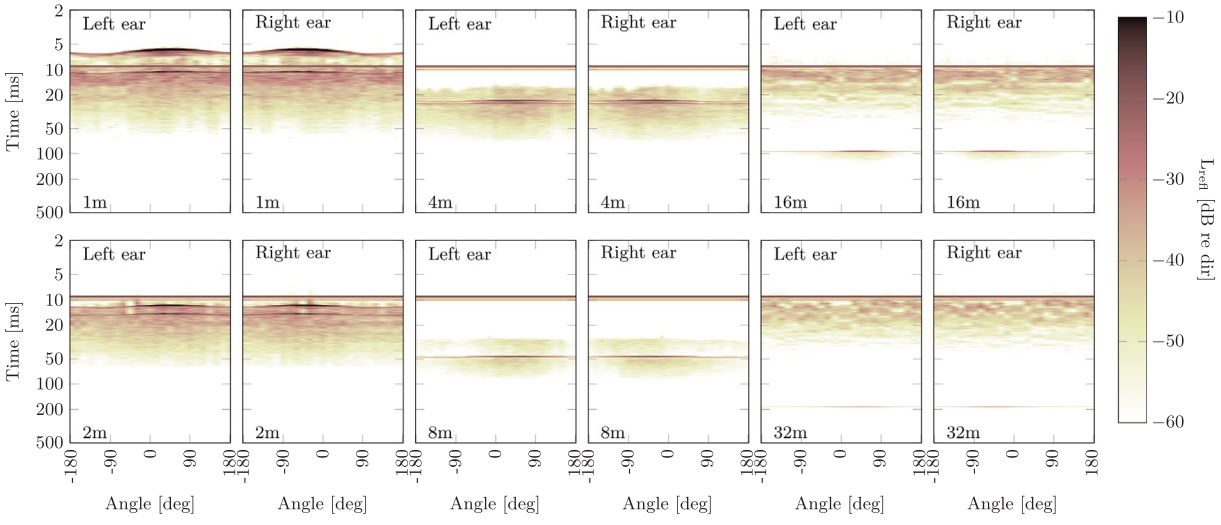


Fig. 2. Energy level of the reflected sound (relative to the direct sound) vs angle and time (in logarithmic scale) at the left and right ears for each of the simulated wall distances of 1 m, 2 m, 4 m, 8 m, 16 m and 32 m.

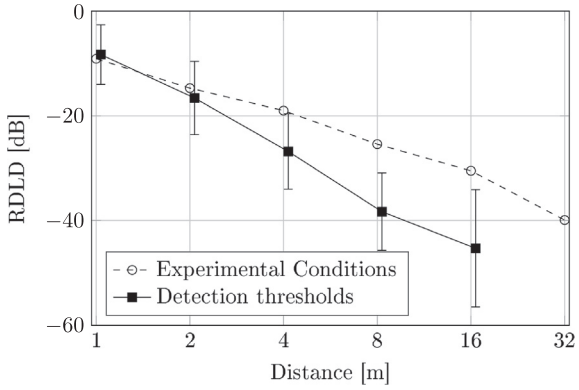


Fig. 3. RDLD values of the wall reflections in the experiment (at $\theta = 0$) and average RDLD thresholds for which a group of untrained sighted people were able to detect a single reflection (70.7% point in the psychometric curve) [17]. Whiskers indicate standard deviations of the RDLD thresholds across subjects. RDLD values lower than the RDLD thresholds would mean that reflections would not be audible for a large percentage of the population.

2.2.2. Stimuli generation

The AVR system is schematically represented in Fig. 4. Parts of this system have been described in an earlier article [17]. The system re-created a scene (at one of the distances in Fig. 1) by loading a set of 24 OBRIRs, corresponding to all orientations around the person with a resolution of 15° at a single distance, from the OBRIR library into the real-time module.

Such a real-time module was implemented in MAX software from Cycling'74. The oral sounds generated by a user were picked up with a microphone and split into two replicas. Before being presented to the ears via open headphones, one replica was sent to an equalizer, in order to compensate for the attenuation introduced by the headphones on the direct sound at the ears, mainly at frequencies above 1 kHz. It was important to restore the natural direct sound, existing without headphones, because in echolocation, this natural direct sound influences the detection ability for the reflections and their interpretation in terms of coloration, loudness and delay. The other replica was acquired with an RME Fireface UCX audio interface and sent to a convolution engine based on a low-latency, non-uniform partitioned convolution, implemented in HISSTools [38]. This convolution engine performed 48 simultaneous convolutions (between the input and each of the 24 2-channel OBRIRs). A head-tracking device (Yost Labs 3-Space sensor) mounted on the headphones provided the head orientation of the user in the horizontal plane with a maximum latency of 20 ms and the resolution was set in runtime to 1° (although the device had a spatial resolution finer than 0.08°). The head orientation was used to calculate the output signal by linearly panning the outputs of the two convolution pairs that had OBRIRs at angles closest to the user orientation. Users felt that the system was responsive in updating the location of the virtual wall and did not report any artifacts related to head-tracking latency.

The user had a remote control in order to interact with the experimental control program (by pressing a button to start or to indicate the response). Signaling sounds containing instructions or feedback on user actions were also played back.

The output of the OBRIRs, mixed with the direct sound, compensated for high frequency attenuation, and with signaling sounds, was played back through open headphones. The sampling rate of the AVR system was 44.1 kHz.

The overall gain applied to the output of the OBRIRs depended on the microphone and headphone sensitivities and on a number of gain parameters in the audio interface and through different software layers. For this reason, a calibration procedure was required. The OBRIR of a large reflecting panel (2 m × 2 m) at a distance of 2 m from the center of the mouth was measured in an anechoic chamber with a dummy head with microphones at its ears and a loudspeaker at its mouth. This dummy head had been used in previous RDLD measurements [29] and yielded results comparable to those obtained with a Head and Torso Simulator B&K type 4128. In the following, the energy of the reflected path in this measurement, averaged for the left and right channels, is noted as $e_{r,p}$. The AVR system (with an arbitrary gain g at the convolution output) was set on the dummy head, the panel was removed from the anechoic chamber and $\mathbf{h}_2(\theta_0 = 0)$ was simulated. A second OBRIR measurement was performed, leading to an energy of the reflected path $e_{r,v}$, averaged between left and right channels. With these measurements, the gain of the convolution output in the AVR system for

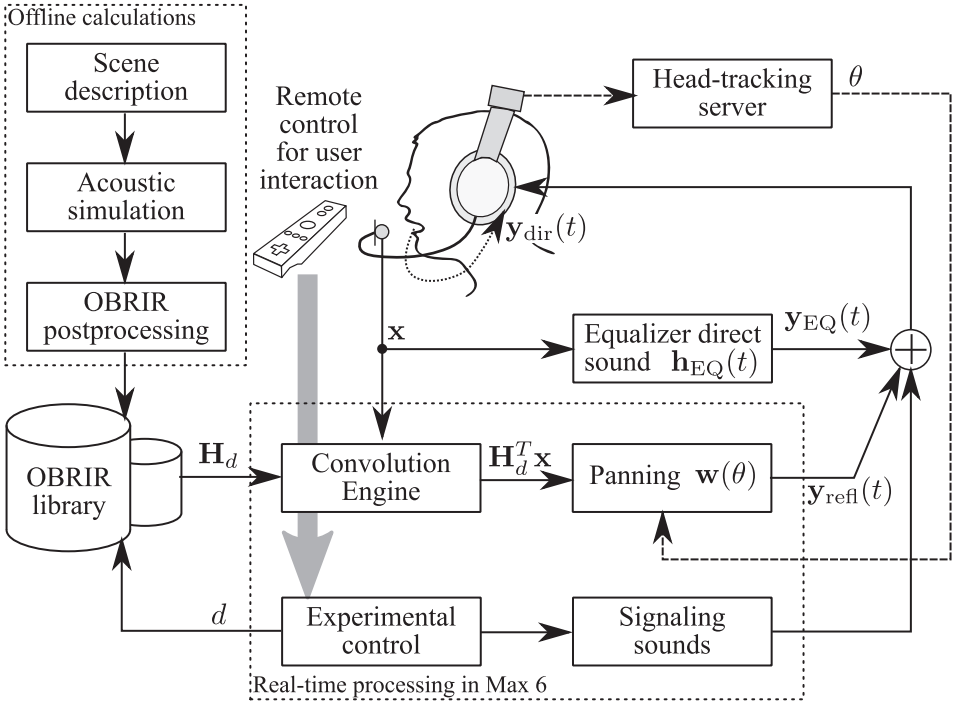


Fig. 4. Block diagram of the experimental apparatus.

the left and right channels was updated to $g' = g\sqrt{e_{r,p}/e_{r,v}}$. The sensitivity of the left and right headphone transducers did not differ by more than 0.1 dB in the frequency range of interest (200 Hz–16 kHz).

An equivalent mathematical description of the AVR system, based on a time-domain convolution, is given in Appendix A.

2.3. Experimental procedure

Before the start of the experiments, participants were instructed about the task they had to perform. They were informed that, in each trial, a virtual wall would be simulated at a random direction and distance around them (see Fig. 1), that this virtual wall would reflect the sound produced by them, and that their task was to find the direction of the wall by only using oral click sounds, align themselves towards this virtual wall by turning their head and/or body and press a button on a controller to report the direction. In case they would not be able to determine the presence of a wall after trying for a reasonable amount of time, they were asked to press a button at any direction. There was no explicit time limit. Note that this task implicitly required subjects to detect the wall in order to be able to localize it. This is not trivial, given that reflection levels were near threshold as shown in Fig. 3.

No training trials were performed. However, the start of the experiments was monitored and if the task was not clear, instructions were clarified and the experiment was restarted. No specific training was given on how to produce the click signal, so as to respect the default clicking strategies of the participants and to be able to determine the impact of different click properties on performance. For convenience, the orientation of the user θ is expressed relative to the actual orientation of the virtual wall (located at $\theta = 0$), whereas the random initial orientation is noted θ_0.

The experiments took place in a 300 m^3 semi-anechoic room of floor dimensions 14 m × 8 m with additional absorbing panels lined on the floor to absorb effectively the range of frequencies of oral clicks. The A-weighted equivalent background noise level was measured below 20 dB. Participants stood in the middle of the room, on a carpet of 0.5 m × 0.5 m, where they had to remain during the experiment. For the sake of keeping participants at ease while iteratively rotating themselves and avoid tangling up the equipment wires, participants were not blindfolded.

After each trial, acoustic feedback was given according to the accuracy of the user in order to provide an engaging user experience. E.g. messages saying *very good* were played back for deviations within 15° to the right or to the left from the true orientation, *you can do it better* up to 45°, or *you were far away from the right angle* for further deviations.

Each distance condition was repeated 4 times, leading to a total of 24 trials.

The orientation and sounds produced by the participants were logged at each trial, making it possible to determine the accuracy of the answered angle, the response latency and the number of clicks, which were used as outcome variables of the experiment.

2.4. Emission properties

The source signal properties play a role in human echolocation performance [17]. For this reason, the oral clicks in the recordings were extracted and analyzed. In a first stage, potential oral clicks were identified by looking at the peaks that exceeded an arbitrary low threshold and were separated by at least 100 ms. In a second stage, all potential oral clicks were visualized and wrongly identified samples (corresponding to speech, noise or artifacts from the microphone) were discarded. The remaining oral clicks were analyzed to derive the following parameters:

- the sound exposure level, L_E, as an indication of intensity and loudness, which in turn affects forward masking dynamics [39].
- the duration of the emission, T_{click}, related to the period during which forward masking is active [39]. It was calculated as the time interval where the amplitude of the envelope was higher than 10 dB below the peak value.
- the peak frequency F_{peak} at which the spectrum reaches its maximum value. This value is reported in other echolocation studies (e.g. [11,40,41]).
- the bandwidth B, defined here as the ratio of the frequencies (in octaves) where the envelope of the 1/12th octave spectrum decayed to −10 dB respect to its maximum value. Note that this bandwidth was calculated from a spectral analysis (in energy per fractional octave band), which resembles the human auditory processing, as opposed to the usual bandwidth derived from the Fourier spectrum (expressing energy per Hz). A broader bandwidth contains more information that the auditory system can use to extract reflection cues.
- the lower frequency F_{min} used for the calculation of B. Extended low frequency content was reported to enhance coloration detection [16].

2.5. Statistical analysis

Statistical analysis has been carried out on the response variables *angle* (also referred to as *angular deviation*) θ, *response latency* (or time to answer) t_A and *number of clicks* N_C, considering the *distance* condition d as the main independent variable. Responses obtained with $N_C = 0$ were treated as outliers and discarded from the data set. Due to the strong correlation between t_A and N_C ($\log N_C = 0.92 \log t_A + 0.18$, Pearson $r = 0.80$), N_C is omitted from the results.

We made the general assumption that in some trials, participants could detect the reflection, whereas in other trials, they could not. Note that the experimental procedure did not allow us to know with certainty whether a participant detected a wall or not on each individual trial, and thus we did not attempt to report whether detection happened or not on a trial basis. However, considering the ensemble of responses under this assumption, the angular data contained a mixture of guess answers uniformly distributed across all possible angles, obtained when participants did not detect the reflection, answers located around the true location, obtained when participants detected the reflection, or even answers located at a mirrored location behind the participant in case of front-back mistakes. Given this kind of data, it is particularly suitable to use the model for azimuth localization proposed in [42], where the probability of responses is modeled with a probability density function f_Θ:

$$f_\Theta(\theta;\mu,\kappa,p_1,p_2,p_{\text{unif}}) = p_1 f_{1,\Theta}(\theta;\mu,\kappa) + p_2 f_{2,\Theta}(\theta;\mu,\kappa) + \frac{p_{\text{unif}}}{2\pi}, \tag{2}$$

where $f_{1,\Theta}(\theta;\mu,\kappa)$ is the probability density function of the correctly detected angles towards the direction of the wall, $f_{2,\Theta}(\theta;\mu,\kappa)$ is the probability density function of responses where a front-back mistake was done and $1/(2\pi)$ is a uniform (guess) distribution. The parameters p_1, p_2, p_{unif} are in the interval [0,1] and $p_1 + p_2 + p_{\text{unif}} = 1$.

In the model proposed in [42], $f_{1,\Theta}(\theta;\mu,\kappa)$ is a von Mises distribution [43], which among the circular distributions is the one having properties similar to the normal distribution in the linear domain. The von Mises distribution has a probability density function

$$f_{1,\Theta}(\theta;\mu,\kappa) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)}, \tag{3}$$

where κ is the *concentration parameter*, μ the mean of the distribution in radians, and $I_0(\kappa)$ the modified Bessel function of the first kind of order zero [44]. Its standard deviation is $\sqrt{1-I_1(\hat{\kappa})/I_0(\hat{\kappa})}$ and for large κ, the distribution becomes concentrated towards the mean. The probability density function for front-back mistakes is

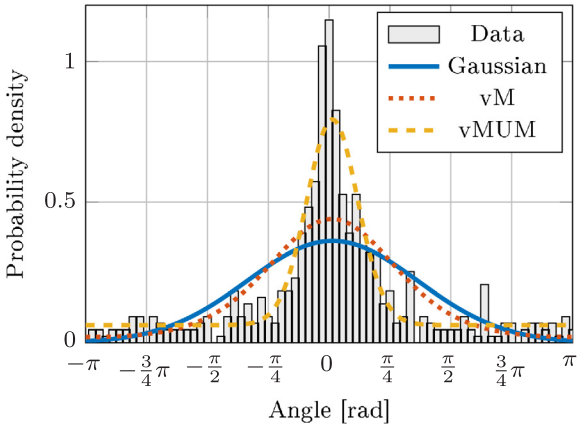


Fig. 5. Histogram of the measured angular deviation, pooled at all distances, and fitted Gaussian, von Mises (vM) and von Mises and Uniform Mixture (vMUM) models.

$$f_{2,\Theta}(\theta;\mu,\kappa) = \frac{e^{-\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)}. \quad (4)$$

In the present experiments, the source was directive and emitted most of its energy in the same direction as the binaural receiver. Thus, it was less likely to observe front-back confusions than in localization of external sound sources, so we typically obtained $p_2 \approx 0$. In view of this, we refer to $p_1 = 1-p_{\text{unif}}$ as the *probability of detection*, to p_{unif} as the *guess probability*, and we relate μ and κ to the *bias* and the *accuracy* of the answer, respectively.

Fig. 5 shows how different statistical models fit the 500 points of angle data, pooled for all participants and conditions using maximum likelihood estimators. Unsurprisingly, the normal distribution (with two parameters – mean and standard deviation) fits the data poorly. The Akaike Information Criterion (AIC), an estimator of how well a statistical model fits the data after taking into account the number of degrees of freedom, is 1520 (the lower the better). Also the von Mises model, which has two parameters (μ and κ), fails to discriminate between random errors and correct decisions. This results in a significantly larger dispersion around the mean (AIC: 1432). On the other hand, the von Mises and Uniform Mixture (vMUM) model (2), which has 4 parameters (p_1, p_2, μ and κ), is capable of capturing both components (AIC: 1300).

After obtaining the parameters of the vMUM model (2) from the observations via maximum likelihood estimation, parametric statistical tests for equality in means and concentration factors [42] were applied to assess differences in angle between groups. In order to assess the differences between p_1 or p_{unif} for different data groups, two-proportions z-tests were used.

Given the relatively innovative approach to analyze angular data following the vMUM model, we provide an alternative analysis based on a repeated measures ANOVA of the absolute unsigned angle to study differences across distance conditions. Note that this method could not provide an estimate of detection rates, as guess trials were just another source of variability in the observations, together with bias and non-accurate localization. Repeated measures ANOVA was also applied to study the response latency variable with distance as the only within-subjects variable and no between-subject factors.

3. Results

The dependent variables *angle* and *response latency* were strongly dependent on the participant, as shown in Fig. 6. Some participants, like S5 and S20 were remarkably accurate in angle and spent very little time to complete the task, whereas other participants, like S11 and S15, showed a remarkably higher spread in their angle and response latency data.

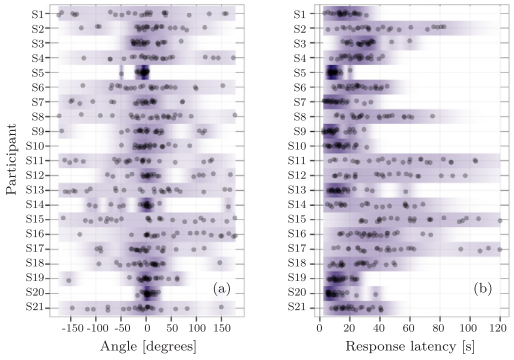


Fig. 6. Individual distribution of the response variables (a) angle and (b) response latency. Individual data pooled across distance conditions are shown as circles and darker areas indicate higher statistical frequency of the responses.

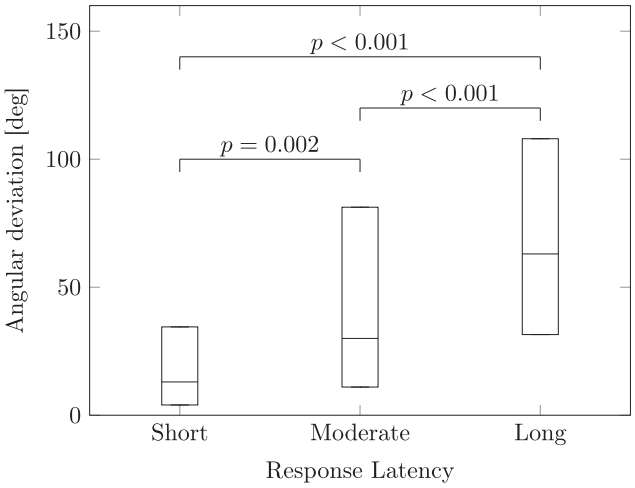


Fig. 7. Absolute angular deviation for short ($t_A < 20$ s), moderate (20 s $< t_A <$ 60 s) and long response latencies ($t_A > 60$ s). The bars indicate the range between the first and third quartiles, and the horizontal lines inside them indicate median values. Significance levels between pairs of data according to Wilcoxon Signed-Ranks tests are displayed above the boxes.

When participants answered quickly, they tended to respond more accurately, as shown in Fig. 7. The absolute angular deviation for moderate latencies between 20 and 60 s was significantly higher than for latencies shorter than 20 s (Wilcoxon Signed-Ranks score $Z = 3.1$, $p = 0.002$), but significantly lower than for latencies longer than 60 s ($Z = 6.2$, $p < 0.001$).

The parameters for the vMUM model (2) fitted to all the angular data (as shown in Fig. 5) were $\hat{\mu} = 2.4°$, $\hat{\kappa} = 9.3$, $\hat{p}_1 = 0.61$, $\hat{p}_2 = 0$, $\hat{p}_{\text{unif}} = 0.39$. The small value of $\hat{p}_2$ indicates that no front/back confusions were present in the data. Using a frequentist interpretation, participants detected a reflection in 61% of the trials, whereas in the remaining 39%, participants did not detect any reflection. The mean direction in the data $\hat{\mu} = 2.4°$ is not significantly different from zero ($p = 0.077$). Thus, on average, participants aligned themselves towards the virtual wall without any noticeable bias.

During the experiments, participants generated very different clicks, as shown in Fig. 8. For example, the click of participant S3 was much shorter and less intense (Fig. 8(a)) than the one of S21 (Fig. 8(c)), which was a 'double click'. The click of S21 had also a lower peak frequency (Fig. 8(d)) than that of S3 (Fig. 8(b)).

The individual values and approximate distribution of the emission properties L_E, T_{click}, F_{peak}, B and F_{min} are shown in Fig. 9 and the summary statistics (average, standard deviation, minimum and maximum values) on Table 1.

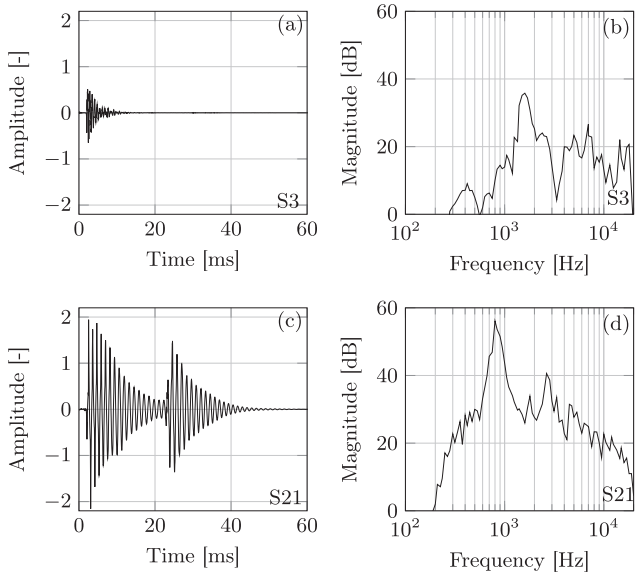


Fig. 8. Example clicks produced by participants S3 and S21 in time domain ((a) and (c)) and corresponding spectra ((b) and (d)).

3.1. Effect of distance

Fig. 10(a) shows the average angle $\hat{\mu}$ when the wall was detected, along with its standard deviation $\sqrt{1-I_1(\hat{\kappa})/I_0(\hat{\kappa})}$ as a function of distance. Figs. 10(b) and (c) show the detection rate $\hat{p}_1$ and the response latency t_A, respectively. The bias, the standard deviation of the angular responses and the response latency generally decreased with increasing distance beyond 2 m, whereas the detection rate had its minimum at 2 m. The distance of 2 m therefore yielded the worst results.

The repeated measures ANOVA on absolute angular deviation confirmed the previous results. Distance had an overall significant effect ($F(5,120) = 8.51$, $p < 0.001$) and post hoc t-tests with Bonferroni correction revealed a significantly higher absolute angular deviation at 2 m than at 16 m (mean: 33.3°, std. err.: 6.9°, $p < 0.001$), at 2 m than at 32 m (mean: 36.7°, std. err.: 8.3°, $p = 0.003$), at 4 m than at 16 m (mean: 20.9°, std. err.: 5.6°, $p = 0.015$), and also at 4 m than at 32 m (mean: 24.3°, std. err.: 6.6°, $p = 0.017$).

For the purpose of comparing distance conditions with the vMUM model, the distances of 1 m is referred to as *short* distance, 2 and 4 m are grouped into *middle* distances, 8 m is a *long* distance and 16 and 32 m are *very long* distances. The differences in detection rate at short ($\hat{p}_1 = 0.78$, $N_{obs} = 83$) and medium distances ($\hat{p}_1 = 0.46$, $N_{obs} = 168$) were statistically significant ($z = -4.8$;$p < 0.001$), and between very long ($\hat{p}_1 = 0.77$, $N_{obs} = 165$) and medium distances ($z = -5.8$;$p < 0.001$), indicating that more target detections were performed at short and very long distances. In addition, the difference in concentration parameters at medium ($\hat{\kappa} = 4.2$, $N_{obs} = 168$) and very long distances ($\hat{\kappa} = 21.4$, $N_{obs} = 165$) was also statistically significant ($p < 0.001$). This result indicates a more accurate localization and lower angular dispersion at very long distances.

Table 1
Mean, standard deviation, minimum and maximum values of sound exposure level L_E, duration T_{click}, peak frequency F_{peak}, bandwidth B and minimum frequency F_{min} of the oral clicks.

	L_E [dB]	T_{click} [ms]	F_{peak} [kHz]	B [octave]	F_{min} [kHz]
Average	50.9	11.5	3.3	1.3	2.1
Std. dev.	6.2	10.3	2.1	0.4	1.2
Min.	35.0	1.9	0.8	0.5	0.6
Max.	64.8	43.1	10.8	2.6	8.5

The repeated measures ANOVA on response latency revealed a significant effect of distance ($F(5,120) = 12.0$, $p < 0.001$) and pointed a significantly higher latency at 1 m than at 16 m (mean: 12.4 s, std. err.: 3.0 s, $p = 0.005$), at 1 m than at 32 m (mean: 11.8 s, std. err.: 2.8 s, $p = 0.004$), at 2 m than at 16 m (mean: 16.1 s, std. err.: 3.8 s, $p = 0.004$), at 2 m than at 32 m (mean: 15.7 s, std. err.: 3.8 s, $p = 0.006$), at 4 m than at 16 m (mean: 14.0 s, std. err.: 3.8 s, $p = 0.019$), and also at 4 m than at 32 m (mean: 13.6 s, std. err.: 3.6 s, $p = 0.015$).

3.2. Effect of signal parameters

The effect of signal parameters on the main outcome variables was studied by grouping the observations according to low or high L_E ($\lessgtr$ 50 dB), short or long T_{click} ($\lessgtr$ 8 ms), low or high B ($\lessgtr$ 1.2 octaves), low or high F_{peak} ($\lessgtr$ 2.8 kHz) and low or high F_{min} ($\lessgtr$ 1.5 kHz). The threshold of 1.5 kHz for defining low or high F_{min} was chosen as the high frequency limit for using interaural time difference as a cue to localize pure tones [45] whereas the other thresholds were set a posteriori to separate observations in groups of similar size.

Following the groups described above, data for the measured angle was further split according to the distance group (short/middle/long/very long distances) and vMUM models were fitted. The most remarkable effects of signal parameters were those shown in Fig. 11. At middle distances (see Fig. 11(a)), a short T_{click} led to a significantly higher detection rate $\hat{p}_1$ (0.52 vs 0.33 for long T_{click}, $z = 2.5$;$p = 0.01$). At very long distances (see Fig. 11(b)), a lower F_{min} resulted in a significantly higher $\hat{\kappa}$ (50.3 vs 13.8 for high F_{min}, $p < 0.001$). Also at very long

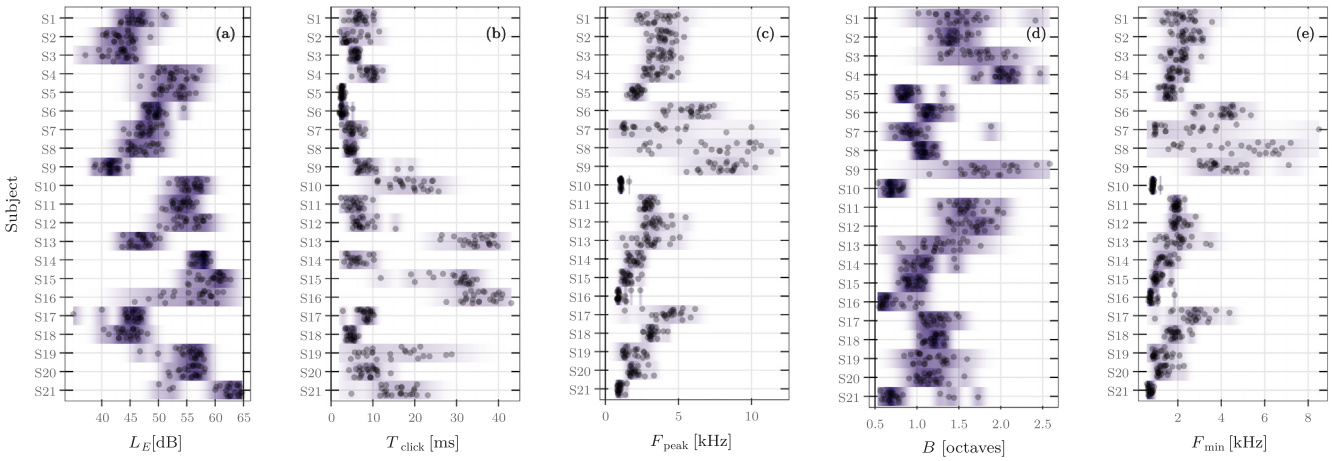


Fig. 9. Individual distribution of (a) sound exposure level L_E, (b) duration T_{click}, (c) peak frequency F_{peak}, (d) bandwidth B and (e) minimum frequency F_{min} of the emissions produced by the participants. Each point corresponds to the parameter averaged across all clicks in one trial.

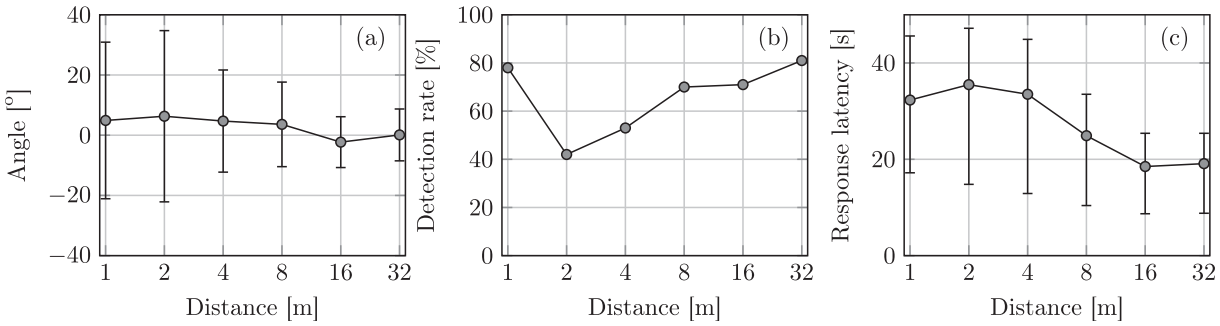


Fig. 10. (a) Mean angle $\hat{\mu}$ with error bars showing ± 1 standard deviation $\sqrt{1-I_1(\hat{\kappa})/I_0(\hat{\kappa})}$ of the detected signals, (b) detection rate $\hat{p}_1$ and (c) time required to answer with error bars showing the 25–75% interquartile range, all of them as a function of distance to the simulated wall.

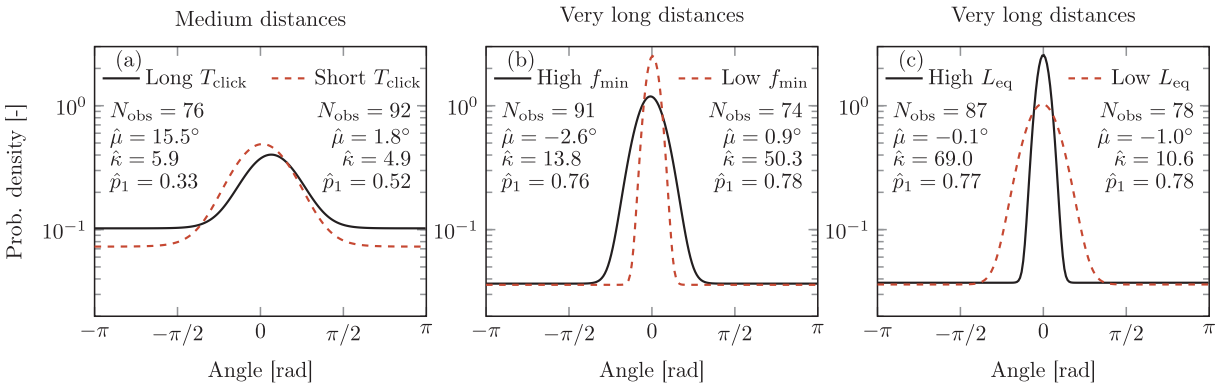


Fig. 11. Influence of the signal properties on the fitted vMUM models describing the angular deviation response: (a) long/short duration T_{click} for middle distance conditions, (b) low/high F_{min} for very long distance conditions and (c) low/high intensity L_E for very long distance conditions.

distances (see Fig. 11(c)), a high L_E led to a significantly higher $\hat{\kappa}$ (69.0 vs 10.6 for low L_E, $p < 0.001$). Other combinations of distance and signal parameters did not yield significant results. No remarkable effects were observed at all for either B nor F_{peak}.

ANOVA tests did not find any significant effect of the different signal parameters on the response latency.

3.3. Training effects

There were no main effects of presentation order ($F(2,394) = 0.49$, $p = 0.61$) in the response latency, when pooling results in three groups containing the 8 first trials (A), the 8 middle ones (B) and the 8 last ones (C). In the 8 first trials, the vMUM parameters for the angle were ($\hat{\mu}_A = 1.7°$, $\hat{\kappa}_A = 7.9$, $\hat{p}_{1,A} = 0.59$) with $N_{obs,A} = 167$; for the 8 middle trials they were ($\hat{\mu}_B = 3.7°$, $\hat{\kappa}_B = 9.7$, $\hat{p}_{1,B} = 0.71$) with $N_{obs,B} = 168$ and for the last 8 trials, they were ($\hat{\mu}_C = 1.1°$, $\hat{\kappa}_C = 10.3$, $\hat{p}_{1,C} = 0.54$) with $N_{obs,C} = 165$. In the latter case, the fit reveals a cluster of answers towards the rear direction, but with $p_2 = 0.01$ indicating some possible front/back confusion. Since we did not find any substantial evidence of front/back confusion in other conditions and the amount of observations falling into this distribution was very low, we do not further elaborate on this result. No significant differences were found among $\hat{\mu}_A$,$\hat{\mu}_B$ and $\hat{\mu}_C$ ($p = 0.55$ between A and B, $p = 0.87$ between A and C and $p = 0.43$ between B and C). A significant difference was found between $\hat{p}_{1,A}$ and $\hat{p}_{1,B}$ ($z = -2.3$; $p = 0.02$) and between $\hat{p}_{1,B}$ and $\hat{p}_{1,C}$ ($z = -3.4$; $p < 0.001$). Although the detection rate was lower for the last samples, the accuracy $\hat{\kappa}_C = 10.3$ was slightly higher than $\hat{\kappa}_B = 3.7$. However this difference was not statistically significant ($p = 0.88$).

4. Discussion

By using the AVR system, participants—sighted and without previous experience in echolocation—were able to successfully echolocate a virtual wall at different distances (1, 2, 4, 8, 16 and 32 m) and orient themselves correctly towards it in approximately 61% of the trials. This finding agrees with a previous study that found that self-motion allowed participants to align themselves accurately to a target direction by means of echolocation using oral clicks [22]. At the same time, large individual differences in expertise/skill level across untrained participants were observed, although vMUM models were not explicitly fitted to each individual, due to the low number of trials per participant (24). The skill level was assumed to be inversely linked to the deviation from the correct angle and to the response latency. Thus, it was assumed that a shorter response latency was associated with easier conditions.

4.1. Dependence on distance

The detection rate $\hat{p}_1$ was highly dependent on the distance of the simulated wall (Fig. 10(b)). It had a minimum at 2 m ($\hat{p}_{1,unif,\,2m} = 0.42$) and increased towards lower $\hat{p}_{1,unif,\,1m} = 0.78$ and higher distances $\hat{p}_{1,unif,\,32m} = 0.81$. The same trends were observed in Fig. 10 for the angle bias $\hat{\mu}$, the standard deviation of the angle and the response latency, indicating that the distance of 2 m was the most difficult one to answer. The detection performance of a single reflection is known to be in general highly dependent on the stimulus and the distance [17] but, for bandpass filtered white noise bursts, worst reflection detection thresholds were found [46] at moderate delays (around 10 ms, reflection distance of 1.7 m), improving towards shorter or longer delays. In the case of shorter delays (including the distance condition of 1 m), it was reported that there is an improvement in coloration detection [46], whereas at longer delays, forward masking is less relevant [39] and

reflection offsets become audible. Furthermore, with low detection rates $\hat{p}_1 = 1-\hat{p}_{unif}$, it is reasonable to assume that localization of the wall, when detected, was performed at near-(masked) threshold level, especially at short distances. Localization at near-(absolute) threshold level is less accurate than at more moderate levels [34]. Fig. 3 relates the RDLD values of the experimental conditions (i.e. how loud actually was the reflected sound in relation to the direct sound during the experiment) to the RDLD thresholds of a group of untrained sighted participants [17] (i.e. how loud the reflection had to be in relation to the direct sound in order to be detected, averaged across participants). In this figure, it is more clearly seen that detection of reflections at short distances occurred near threshold, resulting in low values of $\hat{p}_1$, and that the gap between the actual RDLD in the experiment and the RDLD threshold increased with distance, resulting in higher detection rates $\hat{p}_1$ at further distances, as shown in Fig. 10(b).

In the light of the precedence effect [18], finding a virtual wall is a lag localization experiment, where the lead is the direct sound (which can be assumed to be localized inside the head, without relevant interaural differences) and the lag is the reflected sound from the wall direction. Moreover, the floor reflection acts as an additional lag without interaural differences and occurs after the wall reflection (for the 1 m distance condition) or before it (beyond 2 m). It is known that lag localization becomes worse and the likelihood of fusion increases at short delays [47]. For this reason, we may hypothesize that the worse performance at 2 m is linked to the partial masking of the wall reflection by the floor reflection, an increased fusion likelihood and a higher discrimination suppression [47] in the lag.

In view of this, the improvement in performance at long distances beyond 2 m can be related to the decreasing effect of forward masking after the offset of the direct click signal (average duration $T_{click} = 13.3$ ms, comparable to the reflection delay from a wall at 2.3 m) which results in a lowering of the masked thresholds with increasing delays (or reflection distances), in a higher probability of detection, in lower fusion rates and in lower lag discrimination suppression.

Note that, similarly to previous studies [9,33,48], detection rates decreased dramatically when the distance of the obstacle increased from 1 to 2 m distance (e.g. in [9], echolocation performance decreased to chance for an object distance of 1.8 m). For very far distances of 8 m and further, our results indicate that a large obstacle like a single wall can be localized very accurately via echolocation. It is likely that previous studies focused on echolocation for detection and avoidance of rather small obstacles in indoor environments. However, due to latency constraints in our experimental apparatus, we could not test distances shorter than 1 m (e.g. 0.5 m) which could support the evidence that echolocating an object becomes easier with decreasing distance under 2 m. Whereas detection of reflections at short ranges is useful for obstacle avoidance, detection of reflections at long ranges can be useful to identify environmental landmarks. It is likely that the limiting requirement for using echolocation at long distances is that the level of the reflection is above the background noise and the absolute threshold of hearing. An increase in click L_E would generate a more intense reflection but at the same time, it would trigger the stapedius reflex [49], which would in turn reduce the auditory sensitivity to a reflection. Note that, since the latency of the stapedius reflex is about 100 ms [49], it would only affect the audibility of reflections from objects further than 16 m. Therefore, if a typical click has an L_E of 50 dB (Table 1) and the RDLD of a flat large wall at 100 m is roughly −50 dB (extrapolating from Fig. 3), its reflection would have an L_E of 0 dB. Such a reflection would probably not be audible by most adults even without any background noise, which is unlikely to happen in an outdoor environment.

4.2. Impact of emission properties

The virtual walls at long distances were localized more accurately when participants used clicks with $F_{min} < 1.5$ kHz. This result suggests that participants could access more information about the interaural time differences (which in the case of pure tones, are only available below approximately 1.5 kHz [45]) in addition to cues provided by interaural level differences. By having access to more spatial cues, localization becomes more precise. Other studies, however, found extraction of interaural time differences from the lag less robust than extraction of interaural level differences [50].

The fact that trials using clicks with higher L_E led to an increased accuracy with high $\hat{\kappa}$, especially at long distances, means that the reflection was perceived as a separate event well above threshold and that localization was more accurate [34] than on the barely audible reflection in the cases with lower L_E. This finding agrees with the results of a previous study [17], in which we found that louder clicks led to lower detection thresholds of a single reflection.

The use of short $T_{click} < 8$ ms at distances of 2 and 4 m resulted in a higher detection rate $\hat{p}_1$ (0.52 vs 0.33 with $T_{click} > 8$ ms). There were long clicks, as those shown in Fig. 8(c) with duration comparable to reflection delays. Due to the influence of simultaneous and forward masking [39,17], wall reflections were less likely to be detected by the participants.

It is remarkable that the bandwidth of oral clicks did not have a systematic effect on the results, since it is commonly reported that emissions with high frequency content [31,4,9] are generally beneficial for echolocation. This should not be a surprise, since the current study focused on the localization of a large wall that reflected the sound back to the emitter with similar efficiency in a broad frequency range. Other more subtle echolocation tasks, such as size, texture, and shape discrimination and off-axis horizontal and vertical localization may benefit from high-frequency cues [9]. For this reason, we believe that oral clicks with a broad frequency content are beneficial for echolocation.

Other stationary or hissing sounds used in echolocation [12] may be useful to improve detection and localization of obstacles at short distances, especially below 2 m. However, reflections from large surfaces further than 2 m are generally not audible by using hissing sounds because they fall below detection thresholds [17]. Thus, this study focused exclusively on the use of impulsive sounds such as oral clicks.

4.3. Training effects

During the experiments, qualitative feedback was given to the participants on the accuracy of their responses. Feedback is a main factor in learning [51] that could have led to observable training effects, but nevertheless these were not observed in the present experiments, as reported in Section 3.3. This is unsurprising, since in the current study, participants finished the tests in about 20 min, whereas participants in [11] reached a stable echolocation performance after receiving extensive amount of training for 4–12 weeks.

4.4. Impact of blindness

The primary visual cortex, used for processing of visual information in sighted people, was found to be dedicated to processing of echoes in some early blind echolocators [52], which may result in higher sensitivity to echo cues [19] and source localization [53] than in sighted people. In addition, blind people are more sensitive to interaural level differences, specially in lag detection [50]. We can hypothesize that blind echolocators are able to direct their spatial attention to the reflection interaural time and level differences [19], which would further reduce fusion [54] and allow extracting more reliable localization information. In this case, RDLD detection thresholds for some blind echolocators might be lower than those shown in Fig. 3 and localization might become more accurate.

For the sake of keeping the sighted participants at ease while iteratively rotating themselves to find best alignment towards the wall, they were not blindfolded. While it is unlikely, the visual bias could have introduced an increased variability in the responses.

4.5. Challenges

The AVR system made use of OBRIRs spatially sampled at 15° using panning techniques to generate intermediate angles. While, in principle, this could be an important source of bias due to the distortion of binaural cues, previous research has shown that localization in the horizontal plane using linearly interpolated HRTFs with a resolution of 15° is not degraded with respect to the localization with precise HRTFs [55]. As explained in Appendix A, linear interpolation of HRTFs and panning are equivalent operations in a linear context. Whereas the resolution of 15° with panning in the studied simple scenarios was perceptually acceptable, it remains to be tested whether more complex scenarios with multiple reflections would also be fairly recreated in the AVR system.

In addition, non-individualized HRTFs were used. These are believed to deliver lower localization accuracy in the median plane [56] compared to individualized HRTF data and introduce front/back reversals [57,42]. However, head tracking-controlled sound reproduction improves localization accuracy and reduces front/back reversals [58]. Furthermore, the current experiments restricted localization to the horizontal plane and front/back reversals were not existing because the directivity of the oral click was included in the calculation of the OBRIRs. In this case, when subjects had the wall at their back, the energy returned was very low because there was very little energy radiated towards the back direction at frequencies above 1 kHz.

An open question to answer in future research is whether the echolocation knowledge acquired by participants using the AVR system offers an advantage in real-world tasks. If this was proven, there would be a door open to the systematic exploration and use of optimal individualized training strategies for echolocation with the aid of AVR systems, in combination with real-life training.

5. Conclusions

In an experiment to localize reflections from a virtual wall at distances between 1 and 32 m using only self-generated oral clicks, sighted untrained participants were able to detect the wall in 61% of the trials, however with large differences across individuals and distances. The distance of 2 m was found to be the most difficult condition (detected in 42% of the trials), because the reflection fused with the direct sound and participants were not able to use coloration cues, as it was the case for the distance of 1 m (detected in 78% of the trials), nor detect the reflection as a separated event, as happened with the conditions of 16 and 32 m, which were the easiest ones (detected in 77% of the trials and with much higher accuracy than at 1 m). These results suggest that echolocation can be used effectively not only to avoid obstacles at short distances but also to identify large distant environmental landmarks such as buildings.

The use of shorter and louder clicks led to an increased detection rate, due to a more limited action of forward masking, and to a more accurate localization ability on a reflection level well above threshold. The use of clicks with increased energy at low frequencies allowed for the more effective use of interaural time differences to improve the accuracy of reflection localization at long distances and it led to an increased likelihood of coloration detection at short distances.

All in all, the results of the study suggest that shorter and louder clicks with lower frequency content should be preferred to localize large objects like walls at a large range of distances. In addition to signal production aspects, efforts in echolocation training should be especially directed towards the practice of the most difficult conditions for distances around 2 m.

Acknowledgements

This research work was carried out at the Laboratory of Acoustics, Department of Physics and Astronomy, KU Leuven and at the ESAT Laboratory of KU Leuven, in the frame of the Postdoctoral Fellowship of the Research Foundation Flanders (FWO-Vlaanderen) with Grant No. 1280413 N. This work was also supported by the FP7-PEOPLE Marie Curie Initial Training Network Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS), funded by the European Commission under Grant Agreement No. 316969.

Appendix A. Equivalent mathematical description of the AVR system

This appendix gives an equivalent mathematical description of the Acoustic Virtual Reality System of Fig. 4 based on a time-domain convolution.

The microphone signal $x(t)$, with t the continuous time variable, was digitized and buffered in the length-L vector $\mathbf{x}(n) = [x(n),x(n-1),...,x(n-L+1)]^T$. The discrete time index $n = 1,2...$ is related to the continuous time variable via the sampling rate $f_s = 44.1$ kHz as $t = n/f_s$.

The direct sound at the ears was attenuated at high frequencies due to the presence of the headphones. An equalizer was applied to the microphone signal and played back through the headphones in order to restore those frequencies attenuated while causing minimum latency (in the order of hundreds of μs).

One distance condition d was characterized with the matrix $\mathbf{H}_d = [\mathbf{H}_{d,\mathrm{L}}|\mathbf{H}_{d,\mathrm{R}}]$ containing the OBRIRs at 24 directions, i.e.

$$\mathbf{H}_{d,\mathrm{L}} = [\mathbf{h}_{d,\mathrm{L}}(\theta_0 = 0°)\,\mathbf{h}_{d,\mathrm{L}}(\theta_0 = 15°)\cdots\mathbf{h}_{d,\mathrm{L}}(\theta_0 = 345°)]. \tag{5}$$

The convolution engine delivered the convolution between the input signal $\mathbf{x}(n)$ and each of the OBRIRs for each direction, i.e. $\mathbf{H}_d^T\mathbf{x}(n)$. The angle θ of the user was acquired and used to weight the outputs of the convolution engine through a panning operation to deliver the simulated reflection $\mathbf{y}_{\mathrm{refl}}(n) = [y_{\mathrm{refl,L}}(n),y_{\mathrm{refl,R}}(n)]^T$:

$$\mathbf{y}_{\mathrm{refl}}(n) = [\mathbf{w}(\theta)|\mathbf{w}(\theta)]\left\{\begin{bmatrix}\mathbf{H}_{d,\mathrm{L}}^T\\ \hline \mathbf{H}_{d,\mathrm{R}}^T\end{bmatrix}\mathbf{x}(n)\right\}. \tag{6}$$

The weighting function $\mathbf{w}(\theta)$ is defined as

$$\mathbf{w}(\theta) = [w(\theta,\theta_0 = 0°),w(\theta,\theta_0 = 15°),...,w(\theta,\theta_0 = 345°)]^T \tag{7}$$

with

$$w(\theta,\theta_0) = \max\left[\frac{15°-|\theta-\theta_0|}{15°},0\right], \tag{8}$$

provided that the difference between the current angle θ and the angle parameter θ_0 is expressed within the range θ_0.

Linear interpolation of OBRIRs and panning are equivalent in a linear system, and their difference lies in the order the terms of Eq. (6) are calculated. While the panning operation is performed in that equation, linear interpolation corresponds to the pre-computation of the equivalent

OBRIR before multiplication with the source signal, i.e.

$$\mathbf{y}_{\mathrm{refl}}(n) = \left\{ [\mathbf{w}(\theta)|\mathbf{w}(\theta)] \begin{bmatrix} \mathbf{H}_{d,\mathrm{L}}^{T} \\ \hline \mathbf{H}_{d,\mathrm{R}}^{T} \end{bmatrix} \right\} \mathbf{x}(n). \tag{9}$$

Finally, $\mathbf{y}_{\mathrm{refl}}(t)$ was reproduced through headphones and added to the direct sound component $\mathbf{y}_{\mathrm{dir}}(t) + \mathbf{y}_{\mathrm{EQ}}(t)$.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.apacoust.2018.04.018.

References

[1] Blesser B, Salter L-R. Spaces speak, are you listening? Experiencing aural architecture. Cambridge, MA: The MIT Press; 2007.
[2] Lewald J. Exceptional ability of blind humans to hear sound motion: implications for the emergence of auditory space. Neuropsychologia 2013;51(1):181–6.
[3] Kolarik AJ, Cirstea S, Pardhan S, Moore BCJ. A summary of research investigating echolocation abilities of blind and sighted humans. Hear Res 2014;310C:60–8.
[4] Rojas JAM, Hermosilla JA, Montero RS, Espí PLL. Physical analysis of several organic signals for human echolocation: oral vacuum pulses. Acta Acust United with Acust 2009;95(2):325–30.
[5] Ashmead D, Wall RS, Eaton S, Ebinger KA, Snook-Hill M-M, Guth D, et al. Echolocation reconsidered: using spatial variations in the ambient sound field to guide locomotion. J Visual Impair Blind 1998;92(9):615–32.
[6] Hayes S. Facial vision of the sense of obstacles. In: Perkins Publications, 12, Perkins Institution and Massachussets School for the Blind, Waterwon, MA; 1935.
[7] Supa M, Cotzin M, Dallenbach KM. Facial vision: the perception of obstacles by the blind. Am J Psychol 1944;57(2):133–83.
[8] Papadopoulos T, Edwards D, Rowan D, Allen R. Identification of auditory cues utilized in human echolocation. Objective measurement results. Biomed Signal Process Control 2011;6(3):280–90.
[9] Rowan D, Papadopoulos T, Edwards D, Holmes H, Hollingdale A, Evans L, et al. Identification of the lateral position of a virtual object based on echoes by humans. Hear Res 2013;300:56–65.
[10] Wallmeier L, Wiegrebe L. Ranging in human sonar: effects of additional early reflections and exploratory head movements. PloS One 2014;9(12):e115363.
[11] Schörnich S, Nagy A, Wiegrebe L. Discovering your inner bat: echo-acoustic target ranging in humans. J Assoc Res Otolaryngol: JARO 2012;13(5):673–82.
[12] Rice CE. Human echo perception. Science 1967;155(763):656–64.
[13] Kellogg WN. Sonar system of the blind. Science 1962;137(3528):399–404.
[14] Milne JL, Goodale Ma, Thaler L. The role of head movements in the discrimination of 2-D shape by blind echolocation experts. Percept Psychophys 2014;76(6):1828–37.
[15] Salomons AM. Coloration and binaural decoloration of sound due to reflections. PhD dissertation, TU Delft; 1995.
[16] Buchholz JM. A quantitative analysis of spectral mechanisms involved in auditory detection of coloration by a single wall reflection. Hear Res 2011;277:192–203.
[17] Pelegrín-García D, Rychtáriková M, Glorieux C. Single simulated reflection audibility thresholds for oral sounds in untrained sighted people. Acta Acust United with Acust.
[18] Litovsky RY, Colburn HS, Yost Wa, Guzman SJ. The precedence effect. J Acoust Soc Am 1999;106(4):1633–54.
[19] Dufour A, Després O, Candas V. Enhanced sensitivity to echo cues in blind subjects. Exp Brain Res 2005;165(4):515–9.
[20] Wallmeier L, Gessele N, Wiegrebe L. Echolocation versus echo suppression in humans. Proc Roy Soc Biol 2013;280:20131428.
[21] Arias C, Bermejo F, Hüg M, Venturelli N, Rabinovich D, Ortiz Skarp AH. Echolocation: an action-perception phenomenon. NZ Acoust 2012;25(2):20–7.
[22] Wallmeier L, Wiegrebe L. Self-motion facilitates echo-acoustic orientation in humans. Roy Soc Open Sci 2014;1:140185.
[23] Rosenblum LD, Gordon M, Jarquin L. Echolocating distance by moving and stationary listeners. Ecol Psychol 2000;12(3):181–206.
[24] Vercillo T, Milne JL, Gori M, Goodale MA. Enhanced auditory spatial localization in blind echolocators. Neuropsychologia 2015;67:35–40.
[25] Thaler L. Echolocation may have real-life advantages for blind people: an analysis of survey data. Front Physiol 2013;4:98.
[26] Thaler L, Goodale MA. Echolocation in humans: an overview. Wiley Interdiscip Rev: Cognit Sci 2016;7:382–93.
[27] Picinali L, Afonso A, Denis M, Katz BF. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. Int J Human-Comput Stud 2014;72:393–407.
[28] Cabrera D, Sato H, Martens W, Lee D. Binaural measurement and simulation of the room acoustical response from a person's mouth to their ears. Acoust Australia 2009;37(3):98–103.
[29] Pelegrín-García D, Rychtáriková M, Glorieux C. Audibility thresholds of a sound reflection in a classical human echolocation experiment. Acta Acust United with Acust 2016;102(3):530–9.
[30] Kolarik AJ, Scarfe AC, Moore BCJ, Pardhan S. An assessment of auditory-guided locomotion in an obstacle circumvention task. Exp Brain Res 2016;234:1725–35.
[31] Cotzin M, Dallenbach KM. Facial vision: the role of pitch and loudness in the perception of obstacles by the blind. Am J Psychol 1950;63:485–515.
[32] Ammons CH, Worchel P, Dallenbach KM. Facial vision: the perception of obstacles out of doors by blindfolded and blindfolded-deafened subjects. Am J Psychol 1953;66:519–53.
[33] Schenkman BN, Nilsson ME. Human echolocation: blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object. Perception 2010;39:483–501.
[34] Sabin AT, Macpherson EA, Middlebrooks JC. Human sound localization at near-threshold levels. Hear Res 2005;199(1-2):124–34.
[35] Chu W, Warnock A. Detailed directivity of sound fields around human talkers. Institute for Research in Construction, National Research Council Canada. Tech. Rep, Canada; 2002.
[36] de Vos R, Hornikx M. Acoustic properties of tongue clicks used for human echolocation. Acta Acust United with Acust 2017;103(6):1106–15.
[37] International Standard Organisation. Acoustics – Normal equal-loudness-level contours (ISO 226:2003); 2003.
[38] Harker A, Tremblay PA. The HISSTools Impulse Response Toolbox: Convolution for the Masses. In: ICMC 2012: non-cochlear sound, The International Computer Music Association; 2012. p. 148–55.
[39] Fastl H, Zwicker E. Masking. Psychoacoustics. Facts and models. 3rd ed.Berlin Heidelberg: Springer-Verlag; 2007. p. 61–110.
[40] Thaler L, Castillo-Serrano J. People's ability to detect objects using click-based echolocation: a direct comparison between mouth-clicks and clicks made by a loudspeaker. PloS One 2016;11(5):e0154868.
[41] Thaler L, Reich GM, Zhang X, Wang D, Smith GE, Tao Z, et al. Mouth-clicks used by blind expert human echolocators – signal description and model based signal synthesis. PLOS Comput Biol 2017;13(8):e1005670.
[42] De Sena E, Brookes M, Naylor P, van Waterschoot T. Localization experiments with reporting by gaze: statistical framework and case study. J Audio Eng Soc 2017;65:982–96.
[43] Mardia K, Jupp P. Basic concepts and models. Directional Statistics. Chichester, UK: John Wiley & Sons, Ltd; 2009. p. 25–56. Ch. 3.
[44] Abramowitz M, Stegun IA. Bessel functions of integer order, in: Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol. 55, National Bureau of Standards, Washington, D.C.; 1964. p. 355–434 [ch. 9].
[45] Brughera A, Dunai L, Hartmann WM. Human interaural time difference thresholds for sine tones: the high-frequency limit. J Acoust Soc Am 2013;133(5):2839–55.
[46] Buchholz JM. Characterizing the monaural and binaural processes underlying reflection masking. Hear Res 2007;232(1-2):52–66.
[47] Litovsky RY, Shinn-Cunningham BG. Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression. J Acoust Soc Am 2001;109:346–58.
[48] Schenkman BN, Nilsson ME. Human echolocation: pitch versus loudness information. Perception 2011;40:840–52.
[49] Gelfand SA. Hearing: an introduction to psychological and physiological acoustics. 5th ed., London: Informa Healthcare; 2010.
[50] Nilsson ME, Schenkman BN. Blind people are more sensitive than sighted people to binaural sound-location cues, particularly inter-aural level differences. Hear Res 2016;332:223–32.
[51] Sweetow R, Palmer CV. Efficacy of individual auditory training in adults: a systematic review of the evidence. J Am Acad Audiol 2005;16:494–504.
[52] Thaler L, Arnott SR, Goodale Ma. Neural correlates of natural human echolocation in early and late blind echolocation experts. PloS One 2011;6(5):e20162.
[53] Lessard N, Paré M, Lepore F, Lassonde M. Early-blind human subjects localize sound sources better than sighted subjects. Nature 1998;395(6699):278–80.
[54] London S, Bishop CW, Miller LM. Spatial attention modulates the precedence effect. J Exp Psychol. Human Percept Perform 2012;38(6):1371–9.
[55] Wenzel E, Foster S. Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. Proc IEEE Workshop Appl Signal Process Audio Acoust 1993:102–5.
[56] Møller H, Sørensen MF, Jensen CB, Hammershøi D. Binaural technique: do we need individual recordings? J Audio Eng Soc 1996;44(6):451–69.
[57] Rychtáriková M, Chmelík V, Roozen NB, Glorieux C. Front-back localization in simulated rectangular rooms. Appl Acoust 2015;90:143–52.
[58] Begault DR, Wenzel EM, Anderson MR. 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 2001;49:904–16.