Human echolocation: localizing reflections of self-generated oral sounds in laboratory

David Pelegrín-García[1], Monika Rychtáriková[1,2], Christ Glorieux[1]
[1] Laboratory of Acoustics, Division Soft Matter & Biophysics, Dept. Physics & Astronomy, KU Leuven, 3001 Heverlee, Belgium.
[2] STU Bratislava, Faculty of Civil Engineering, Dept. of Building Structures, Radlinskeho 11, Bratislava, 813 68, Slovak Republic.

Summary
The active sensing and perception of the environment by auditory means is typically known as echolocation. Through the emission of oral sounds and the interpretation of the reflections in relation to the direct sound, blind people can acquire spatial knowledge about their surroundings and improve their mobility in unknown spaces. While this technique is becoming more common in Orientation & Mobility training, it has not yet become a mainstream practice. This paper aims, on one hand, at presenting this modality of perception and its underlying sensory mechanisms and, on the other hand, at showing the results of a laboratory experience at the Laboratory of Acoustics at KULeuven, in which we investigate the ability of echolocation-naïve sighted subjects to use echolocation for aligning themselves toward virtual silent targets generated through an acoustic virtual reality system. It is shown that all subjects were able to complete the tasks, although detection of targets at closer distances entailed more difficulty than at further distances. Significant individual training effects were observed and should be accounted for in future similar tests.

PACS no. 43.66.Qp, 43.66.Lj

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. Some blind people have learnt to echolocate [3], i.e. to detect and localise obstacles and environmental features based on the reflections they produce in response to self-generated sounds, typically oral clicks [4], or even to ambient noise [5]. 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, ILD, ITD 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. Localisation of reflections is particularly precise due to a partial inhibition of the precedence effect during echolocation [15]. This technique represents an active perception mode [16], meaning that the perception of auditory space integrates the auditory sensation with the vestibular and proprioceptive feedback [17] and thus head movements are crucial [18] for effective mobility and detailed tasks like shape perception [14].

The primary visual cortex, used for processing of visual information in sighted people, is dedicated to processing of echoes in some early blind echolocators [19], which may result in higher sensitivity to echo cues [20] and source localization [21] than in sighted people. 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 [22].

Acoustic Virtual Reality (AVR) systems which account for head orientation are regarded useful for the acquisition of auditory space maps [23], for evoking sensations arising in echolocation (e.g. [10, 17]) and for the conduction of psychoacoustic tests (e.g. [15]). Sighted subjects are able to learn basic echolocation tasks using an AVR system [10, 11]. Therefore, these systems show a potential to explore effective learning strategies in echolocation and gain further knowledge about its psychophysical mechanisms.

The present paper introduces a pilot study, which made use of the AVR system developed at our labo-

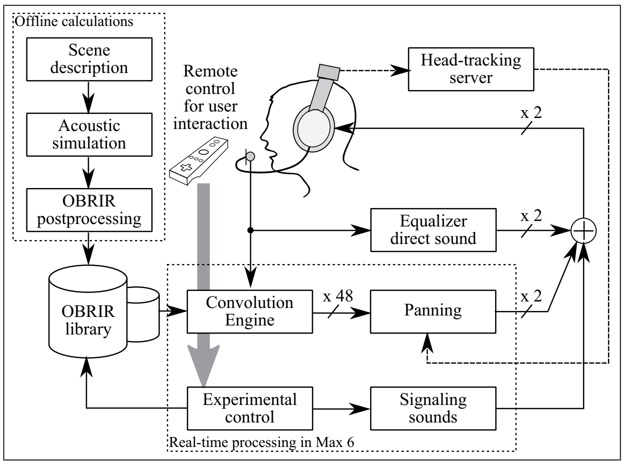

Figure 1. Block diagram of the experimental apparatus.

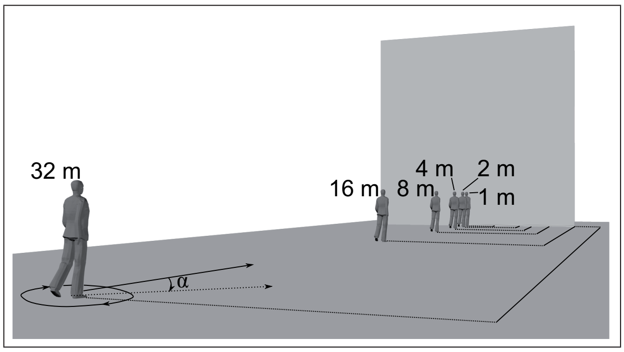

Figure 2. Schematic representation of the conditions simulated in the experiment, where subjects were virtually placed at different distances in front of a large concrete wall and on top of a concrete floor. In each of the 6 distances, subjects were free to rotate around.

ratory [24], in which the ability of echolocation-naïve subjects to find virtual silent targets—a hard reflecting wall at different distances—around them, by using self-generated oral clicks, was evaluated.

2. Method

Using the AVR system developed in our laboratory [24], five echolocation-naïve sighted subjects, aged 25 to 39 years old, had to find a virtual wall located at six different distances between 1 m and 32 m and at a random orientation, with the only aid of self-generated oral clicks.

2.1. Apparatus

The AVR system, described in detail in [24], is schematically represented in Figure 1. The system recreated a scene, which had been described and simulated beforehand and which was contained in an oral-binaural room impulse response (OBRIR), a function that characterises the propagation of sound between the mouth and the ears of a receiver. After the calculation, the direct sound was removed from these OBRIRs, as well as the first 3.5 ms in order to compensate for the latency of the system. As the user could freely rotate the head in the horizontal plane, there were 24 OBRIRs at each point, corresponding to orientations at each 15°. The OBRIRs were stored in a library which the real-time module accessed.

Such a real-time module was implemented in the software Max from Cycling'74. The oral sounds generated by a user were picked up with a microphone and split into two paths. One was sent to an equaliser to compensate for the attenuation introduced by the open-headphones on the direct sound at the ears, mainly at high frequencies above 1 kHz. The other path was sent to a convolution engine based on a zero-latency, non-uniform partitioned convolution, implemented in the HISSTools [25]. This convolution engine performed 48 simultaneous convolutions (between the input and each of the 24 2-channel OBRIRs). With the help of a head-tracking device, the orientation of the user was determined. It was used to calculate the output signal by panning the outputs of the two convolution pairs that had OBRIRs at angles closest to the user orientation.

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

The output of the OBRIRs (which contain the effect of the acoustic reflections of the environment), mixed with the compensation for the direct sound attenuation the and signalling sounds, was played back through open headphones. The sampling rate of the AVR system was 96 kHz.

2.2. Conditions

There were six different conditions, corresponding to the reflections that would occur on the oral sounds of a person with a concrete (thus reflective) wall of dimensions 10 m × 10 m at distances of 1, 2, 4, 8, 16 and 32 m on a concrete floor at his/her own ears. A schematic representation of the simulated scenarios/distances is shown in Figure 2.

The reflections of the wall and the floor were simulated with the room acoustics simulation software CATT-Acoustic™v9.0c. A receiver was placed at the middle point in between the ears, and a source simulating the mouth (and thus with the average directivity pattern of human voice) was placed 0.1 m in front of the receiver and pointed away from it. 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 separated calculations, the wall was located at each of the six different distance conditions (1, 2, 4, 8, 16 and 32 m) from the receiver. At each distance, simulations were performed for 24 orientations of source/receiver, always rotating the source around the fixed receiver, at intervals of 15°.

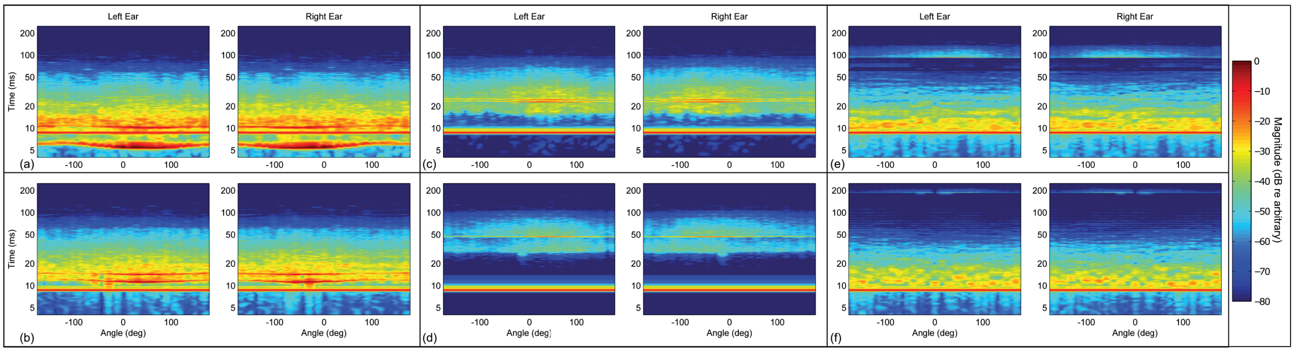

Figure 3. Representation of energy level (in colour) vs time (in logarithmic scale) vs angle at the left and right ears for each of the simulated conditions: wall at (a) 1 m, (b) 2 m, (c) 4 m, (d) 8 m, (e) 16 m and (f) 32 m.

The wall and the floor had an absorption coefficient of 0.01 at 125 Hz monotonously increasing to 0.05 at 4 kHz. These surfaces had a default scattering 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 162.000 rays were used, and the length of the impulse response was set to 0.5 s . Diffraction was not active. 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*) was used.

In post-processing, the OBRIRs had the direct sound and the first 3.5 ms removed and the sampling rate was increased from 44.1 kHz to 96 kHz.

An energy-time representation of the OBRIRs is shown in Figure 3 as a function of the horizontal angular rotation with respect to the wall normal. For each of the six conditions, there are two graphs corresponding to the left and right ears. Whereas ITDs are difficult to observe due to the scale of the graph (up to 400 ms vs hundreds of μs in the case of ITDs), ILDs are more remarkable. For example, focusing on the wall reflection at 4 m distance (and a delay of about 25 ms) in Figure 3(c), for user rotations of 45° towards the right (positive angles), the left ear receives more intense energy 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 is always visible at a delay of approximately 8 ms, independently of orientation and wall distance.

2.3. Experimental procedure

Test sessions were preceded by an explanation of the task to the subjects and a training trial at each of the distance conditions, so that subjects could get acquainted with producing clicks and listening to the reflections. It was explained that, in far conditions, reflections were perceived as separate events (or echoes) and in close conditions, they were perceived as coloration or as angle-dependent ILDs and therefore subjects should attain to loudness or coloration cues to find the obstacle.

In each trial, the starting angle $\alpha = \alpha_0$ of the subject with respect to the wall (see Figure 2) was randomised. The task of the subject was to align him/herself with the wall, thus to find $\alpha = 0$, by only using oral clicks. Once aligned, the subject had to press a button in the controller.

After each trial, feedback was given according to the accuracy of the user. E.g. messages saying *very good* were played back for deviations within 15°, *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. Distance conditions were randomised.

The orientation and sounds produced by the user were logged at each trial, making it possible to determine the accuracy of the answer, the time required to answer, the number of clicks and the total angular displacement, which were used as outcome variables of the experiment.

3. Results

The performance of the subjects in the task was evaluated in terms of four outcome variables, namely the angular deviation from the right angle, the time required to give an answer, the number of clicks generated and the total angular displacement. The independent variables regarded as having a potential effect on the outcomes were the subjects themselves, the distance condition, the experimental order, the initial angle on the trial. Interactions between subject and distance condition and between subject and experimental order were considered too. A series of ANOVA (ANalysis Of VAriance) models on each outcome variable were built using all the independent variables and interactions simultaneously. The resulting p-values are summarised in Table I.

It can be seen that the subject and the distance condition had the most significant effects on the outcome variables, except on the angular deviation. A

Table I. Summary of p-values obtained in ANOVA models for each of the measured outcomes, based on the subject, the distance condition (regarded as a factor), the initial angle, the experimental order and interactions between the subject, distance condition and order. Bold face ($p \leq 0.01$) and italics ($0.01 < p \leq 0.05$) are used to remark significance.

	Angular deviation	N Clicks	Time	Total angle
Subject	0.33	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-16}}$	$\mathbf{< 10^{-4}}$
Distance	0.87	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-10}}$
Initial angle	0.25	0.25	0.39	0.32
Order	0.59	*0.012*	0.66	0.14
Subject* distance	0.91	*0.027*	$\mathbf{< 10^{-4}}$	0.30
Subject* order	0.097	*0.039*	**0.010**	*0.027*

summary of the outcome variables is shown in the box plots of Figure 4, grouped by subject on the left column or by distance condition in the right column.

The average angular deviation was less than $\pm 10^o$ at any condition or for any subject (and the grand average was approximately 0.6^o). This means that all subjects were able to complete successfully the task, independently of the difficulty it posed, and that the system worked as intended for this particular task.

Difficulty should rather be judged on the spread of the angular deviation or the time required to answer. It can be seen that angular deviations at distances of 2 or 4 m had a larger spread than the angular deviations at 8 or 16 m. Moreover, the time required to answer at distances of 8 m or further was generally lower than the time required to answer at distances of 4 m or shorter. This is also visible in the total angular displacement, highly correlated with the time required to answer.

By looking at the time required to answer in Figure 4(e), there are significant differences among subjects which may be attributed to individual skill level. Subjects sB and sE may have the highest skill level, subjects SC and SD a medium skill level, and subject SA a low skill level.

A significant interaction between subject and distance condition in the time required to answer (as seen in Table I) indicated that the relative difficulty of distance conditions differed among subjects. This is explicitly displayed in Figure 5, where it can be seen that e.g. the 32 m condition was much easier than the 1 m condition for subject sA but not for subject sD.

Given the significant interaction between subject and experimental order in the time required to answer, training effects seem to be relevant. Figure 6 shows the time required to give an answer for each

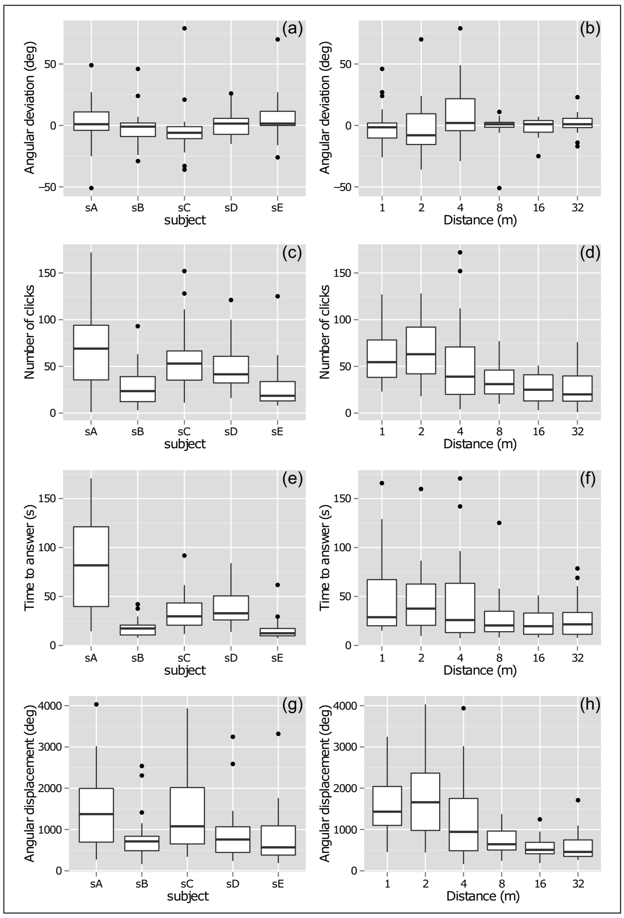

Figure 4. Summary box plots of the outcome variables in the experiments as a function of subject (averaged across conditions, on the left column) or as a function of conditions (averaged across subjects, on the right column): angular deviation (a)-(b), number of clicks (c)-(d), time required to answer (e)-(f) and total angular displacement (g)-(h).

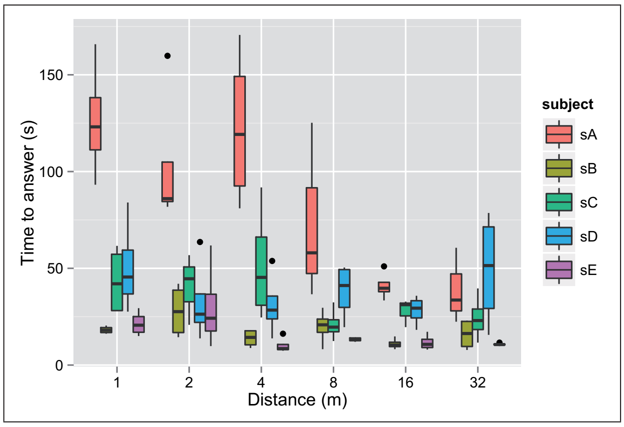

Figure 5. Time required to answer as a function of the distance to the virtual wall, grouped by subjects.

trial and each subject, with linear trends grouped by subject. There was a decrease of the response time with presentation order for the least skilled subject (sA), indicating a familiarisation with the task. In the

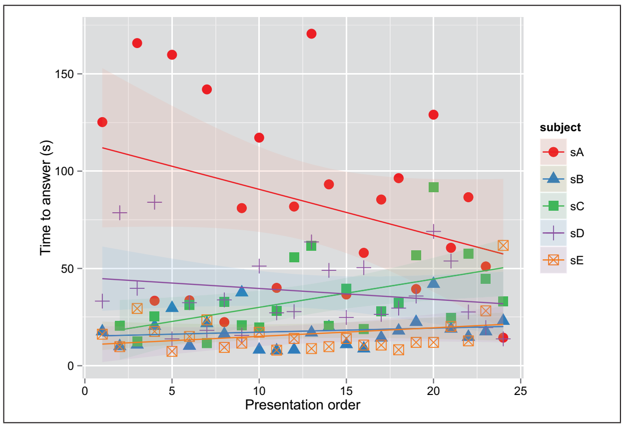

Figure 6. Time required to answer as a function of presentation order and regression models for each subject separately.

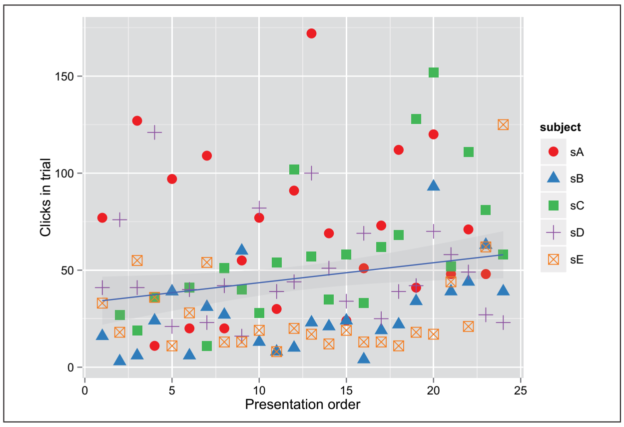

Figure 7. Number of clicks as a function of presentation order and overall linear regression model.

most skilled subjects, training effects were not so relevant.

Another significant main effect on Table I is that of the presentation order on the number of clicks generated ($p = 0.012$). This increase might be due to an increased facility of generating clicks as the experiment runs; i.e. an increased clicking rate resulting from training.

4. Discussion

By using the AVR system, all subjects—sighted and without previous experience in echolocation—were able to successfully echolocate a virtual wall at different distances and orient themselves towards it, i.e. they were able to complete the experimental task with a low average angular deviation. At the same time, large individual differences in expertise/skill level across subjects were observed. The skill level was assumed to be inversely linked to the angular deviation from the correct angle and to the time required to complete the task.

It was observed that closer conditions (virtual wall at 4 m or less) were generally more difficult than further conditions (virtual wall at 8 m or more) for our subjects. For such long distances, the reflected clicks were generally perceived as separate events, thus as external sound sources, which normal hearing people are good at localising. However, in the case of closer distances, the reflection fused with the direct sound, and this probably posed a difficulty to sighted subjects who are not used to localise sounds in the presence of a correlated masking sound coming directly from the mouth, maybe due to the prevalence of the precedence effect. However, correct localisation of nearby obstacles is crucial for echolocation, and untrained subjects may develop this skill through training, as some of the trends in our results support.

The differences in skill level led to a significant training effect, i.e. a reduction of the time required to complete the task (as shown in Figure 6) for the least skilful subject. In view of this effect, it is necessary either to extend the amount of training trials until a stable performance is reached, or not to give feedback during the experiment. Since feedback is a main factor in learning [26], neglecting feedback would slow down the observed training effects.

Despite the limitations of the AVR system, which made use of OBRIRs spatially sampled at 15°, directional cues were preserved thanks to panning techniques—similar to the placement of a virtual source in between the loudspeakers in a standard stereo setup. Whereas this technique has provided good experimental results in the proposed simple scenarios, it remains to be tested whether more complex scenarios with multiple reflections would also be fairly recreated in the AVR system. An open question to answer in future research is whether the echolocation knowledge acquired by subjects using the AVR system offers an advantage in real-world tasks. Further experiments will account for training effects, increase the statistical power with more subjects, study the performance in a similar real-life task and compare subjects' performance to that of an expert echolocator.

5. Conclusions

The main conclusions of the work are the following:

- Large differences in skill level were found among subjects. Nevertheless, all of them were able to use echolocation to detect the angular location of a wall.
- Training effects differed among subjects, therefore more training sessions are required in order to reach a stable performance in the task, independent of presentation order.
- Reflections with longer delays, coming from far walls, were easier to localise than reflections with shorter delays, coming from nearby walls.

All in all, the results reinforce the idea that Auditory Virtual Reality systems provide helpful means to study a number of tasks related to human echolocation.

Acknowledgement

DPG is grateful to FWO-V for supporting this research by financing his postdoctoral grant no. 1280413N.

References

[1] B. Blesser and L.-R. Salter, *Spaces Speak, Are You Listening? Experiencing Aural Architecture.* Cambridge, MA: The MIT Press, 2007.

[2] J. Lewald, "Exceptional ability of blind humans to hear sound motion: implications for the emergence of auditory space.," *Neuropsychologia*, vol. 51, pp. 181–6, Jan. 2013.

[3] A. J. Kolarik, S. Cirstea, S. Pardhan, and B. C. J. Moore, "A summary of research investigating echolocation abilities of blind and sighted humans.," *Hearing research*, vol. 310C, pp. 60–68, Apr. 2014.

[4] J. A. M. Rojas, J. A. Hermosilla, R. S. Montero, and P. L. L. Espí, "Physical Analysis of Several Organic Signals for Human Echolocation: Oral Vacuum Pulses," *Acta Acustica united with Acustica*, vol. 95, pp. 325–330, Mar. 2009.

[5] D. Ashmead, R. S. Wall, S. Eaton, K. A. Ebinger, M.-M. Snook-Hill, D. Guth, and X. Yang, "Echolocation reconsidered: Using spatial variations in the ambient sound field to guide locomotion," *Journal of Visual Impairment and Blindness*, vol. 92, no. 9, pp. 615–632, 1998.

[6] S. Hayes, "Facial vision of the sense of obstacles," in *Perkins Publications, 12*, Waterwon, MA: Perkins Institution and Massachussets School for the Blind, 1935.

[7] M. Supa, M. Cotzin, and K. M. Dallenbach, "Facial Vision: The Perception of Obstacles by the Blind," *The American Journal of Psychology*, vol. 57, no. 2, pp. 133–183, 1944.

[8] T. Papadopoulos, D. Edwards, D. Rowan, and R. Allen, "Identification of auditory cues utilized in human echolocation. Objective measurement results," *Biomedical Signal Processing And Control*, vol. 6, pp. 280–290, July 2011.

[9] D. Rowan, T. Papadopoulos, D. Edwards, H. Holmes, A. Hollingdale, L. Evans, and R. Allen, "Identification of the lateral position of a virtual object based on echoes by humans.," *Hearing research*, vol. 300, pp. 56–65, June 2013.

[10] L. Wallmeier and L. Wiegrebe, "Ranging in human sonar: effects of additional early reflections and exploratory head movements.," *PloS one*, vol. 9, p. e115363, Jan. 2014.

[11] S. Schörnich, A. Nagy, and L. Wiegrebe, "Discovering your inner bat: echo-acoustic target ranging in humans.," *Journal of the Association for Research in Otolaryngology : JARO*, vol. 13, pp. 673–682, Oct. 2012.

[12] C. E. Rice, "Human echo perception," *Science*, vol. 155, no. 763, pp. 656–664, 1967.

[13] W. N. Kellogg, "Sonar system of the blind," *Science*, vol. 137, no. 3528, pp. 399–404, 1962.

[14] J. L. Milne, M. a. Goodale, and L. Thaler, "The role of head movements in the discrimination of 2-D shape by blind echolocation experts.," *Attention, perception & psychophysics*, vol. 76, pp. 1828–37, May 2014.

[15] L. Wallmeier, N. Gessele, and L. Wiegrebe, "Echolocation versus echo suppression in humans," *Proceedings of the Royal Society of Biology*, vol. 280, no. August, p. 20131428, 2013.

[16] C. Arias, F. Bermejo, M. Hüg, N. Venturelli, D. Rabinovich, and A. H. Ortiz Skarp, "Echolocation: An Action-Perception Phenomenon," *New Zealand Acoustics*, vol. 25, no. 2, pp. 20–27, 2012.

[17] L. Wallmeier and L. Wiegrebe, "Self-motion facilitates echo-acoustic orientation in humans," *Royal Society Open Science*, vol. 1, p. 140185, 2014.

[18] L. D. Rosenblum, M. Gordon, and L. Jarquin, "Echolocating Distance by Moving and Stationary Listeners," *Ecological Psychology*, vol. 12, pp. 181–206, July 2000.

[19] L. Thaler, S. R. Arnott, and M. a. Goodale, "Neural correlates of natural human echolocation in early and late blind echolocation experts.," *PloS one*, vol. 6, p. e20162, Jan. 2011.

[20] A. Dufour, O. Després, and V. Candas, "Enhanced sensitivity to echo cues in blind subjects.," *Experimental Brain Research*, vol. 165, pp. 515–9, Sept. 2005.

[21] N. Lessard, M. Paré, F. Lepore, and M. Lassonde, "Early-blind human subjects localize sound sources better than sighted subjects.," *Nature*, vol. 395, no. 6699, pp. 278–280, 1998.

[22] L. Thaler, "Echolocation may have real-life advantages for blind people: an analysis of survey data.," *Frontiers in physiology*, vol. 4, p. 98, Jan. 2013.

[23] L. Picinali, A. Afonso, M. Denis, and B. F. Katz, "Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge," *International Journal of Human-Computer Studies*, vol. 72, pp. 393–407, 2014.

[24] D. Pelegrín-García, M. Rychtáriková, C. Glorieux, and B. F. G. Katz, "Interactive auralization of self-generated oral sounds in virtual acoustic environments for research in human echolocation," in *Proceedings of Forum Acusticum 2014*, (Krakow, Poland), 2014.

[25] A. Harker and P. A. Tremblay, "The HISSTools Impulse Response Toolbox: Convolution for the Masses," in *ICMC 2012: Non-cochlear Sound*, pp. 148–155, The International Computer Music Association, 2012.

[26] R. Sweetow and C. V. Palmer, "Efficacy of individual auditory training in adults: a systematic review of the evidence.," *Journal of the American Academy of Audiology*, vol. 16, pp. 494–504, 2005.