KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Pelegrín-García, D, Rychtáriková, M., Glorieux, C. (2017), **Single simulated reflection audibility thresholds for oral sounds in untrained sighted people** *Acta Acustica united with Acustica* 103(3), pp.492-505
Archived version	The archived file is not the final published version of the article.
Published version	https://doi.org/10.3813/AAA.919078
Journal homepage	The definitive publisher-authenticated version is available online at **http://www.ingentaconnect.com/content/dav/aaua**
Author contact	david.pelegringarcia@kuleuven.be *+ 32 (0)16 374454*
Abstract	Human echolocation is an auditory phenomenon, and as such, inherits the benefits and limitations of the auditory system. In this work, we study the detection thresholds for single synthetic reflections in 12 echolocation-naïve, sighted participants, using external and self-generated oral clicks and hissing sounds, and link the results to previous psychoacoustic findings. Simulated obstacle distances ranged between 0.5 and 16 m, equivalent to reflection delays between 3 and 94 ms relative to the direct sound. Participants had to indicate which out of three intervals (3AFC procedure) added a single reflection to the direct sound, and a one-up two-down rule adjusted the reflected-to-direct level difference (RDLD) applied to the simulated reflection. The thresholds decreased with increasing simulated obstacle distance when using oral clicks, but not when using hissing sounds. Detection with hissing sounds is closely related to coloration and loudness detection. For oral clicks, limitations are imposed by forward masking: louder and shorter clicks resulted in

	lower thresholds. The fine acuity with oral clicks observed at long distances suggests that, in situations with low background noise and few competing reflections such as open air spaces, it may be possible to detect reflections from large distant objects like walls or buildings.
Copyright	© (2017) S. Hirzel Verlag/European Acoustics Association. Readers must contact the publisher for reprint or permission to use the material in any form
IR	ftp://ftp.esat.kuleuven.be/pub/sista/pelegrin/17-11.pdf

(article begins on next page)

Single simulated reflection audibility thresholds for oral sounds in untrained sighted people

David Pelegrín-García[1,2], Monika Rychtáriková[3,4], Christ Glorieux[2]

[1)] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, 3001 Leuven, Belgium. david.pelegringarcia@kuleuven.be

[2)] KU Leuven, Department of Physics and Astronomy, Division Soft Matter and Biophysics, 3001 Leuven, Belgium

[3)] KU Leuven, Faculty of Architecture, Research Department of Architecture, Hoogstraat 51, 9000, Gent, Belgium

[4)] STU Bratislava, Stavebná fakulta, Katedra KPS, Radlinského 11, 813 68 Bratislava 15, Slovakia

Summary

Human echolocation is an auditory phenomenon, and as such, inherits the benefits and limitations of the auditory system. In this work, we study the detection thresholds for single synthetic reflections in 12 echolocation-naïve, sighted participants, using external and self-generated oral clicks and hissing sounds, and link the results to previous psychoacoustic findings. Simulated obstacle distances ranged between 0.5 and 16 m, equivalent to reflection delays between 3 and 94 ms relative to the direct sound. Participants had to indicate which out of three intervals (3AFC procedure) added a single reflection to the direct sound, and a one-up two-down rule adjusted the reflected-to-direct level difference (RDLD) applied to the simulated reflection. The thresholds decreased with increasing simulated obstacle distance when using oral clicks, but not when using hissing sounds. Detection with hissing sounds is closely related to coloration and loudness detection. For oral clicks, limitations are imposed by forward masking: louder and shorter clicks resulted in lower thresholds. The fine acuity with oral clicks observed at long distances suggests that, in situations with low background noise and few competing reflections such as open air spaces, it may be possible to detect reflections from large distant objects like walls or buildings.

Keywords: echolocation, audibility thresholds, psychoacoustics.

1 Introduction

In the context of subjective room acoustics, sound reflections modify the perceived quality of music and affect speech intelligibility. At the same time, they also contain useful information about the distance from source and receiver to the boundaries that generate these reflections. In nature, bats and other animals exploit this information to perceive the environment and navigate themselves. This means of sensing has been called echolocation [1]. Humans, blind or sighted alike, can also learn to use echolocation by either using ambient sounds [2] or by emitting sounds such as oral clicks, cane hits, or by means of portable clicking devices, listening to the echoes and interpreting the resulting patterns [3]. When using ambient sounds, this ability is usually referred to as *passive echolocation*. When the user generates the sounds himself, it is referred as *active echolocation*. Echolocation may bring real life benefits to blind users, through higher mobility in unknown places and access to better paid jobs [4].

In order to understand the effects of a single reflection on ambient sound, much of the knowledge gained in the area of subjective room acoustics also applies in the context of human echolocation.

The audibility of single reflections has been researched in the context of room acoustics, in terms of understanding of how distinct reflections (or echoes) may affect the perceived quality of a music performance, reinforce or disrupt a spoken message. Kuttruff [5], summarizing the reports of extensive experimentation on the topic, stated that the perception of a reflection is not necessarily a conscious experience: it can be detected as an increase in loudness, a coloration of the sound, a change in apparent size of the sound source, a sound image localization displacement, or as a repetition of the original sound as a separated event [6]. In the latter case, the reflection is called an 'echo'. Moreover, the audibility thresholds of a single reflection depend on the signal used, as shorter signals generally lead to lower thresholds [5, 7]. At the same time, the threshold depends on the delay of the reflection relative to the direct sound. For a music signal, the threshold is highest for delays of approximately 20 ms. It decreases for shorter delays, due to detection of coloration, and for longer delays, due to detection of separate events [8]. This trend is also observed for bursts of bandpass filtered (100-5000 Hz) white noise of 200 ms duration [9]. Furthermore, the hearing system is more sensitive to detect changes in reflection level rather than in delay [10].

In active echolocation, signals are typically different from speech, music or synthetic noise. Among experienced echolocators (i.e. people who functionally use echolocation), like Daniel Kish [11], oral sounds are most commonly used. They are potentially more useful for the user than hand claps or cane hits because the alignment between the source (mouth) and receivers (ears) is kept constant. In addition, the user has more control on the intensity and quality of the signal. The oral click is a usual choice among oral sounds. Rojas *et al.* [12] describe the oral click as the sound generated by a quick release of the vacuum produced by the tip of the tongue in the upper end of the hard palate. Some other experiments [13, 14] investigating human echolocation showed that hissing sounds (like a sustained /s/ sound) can also be useful in detecting targets, at least at close distances up to 2.5 m. Schenkman *et al.* [14], using pre-recorded bursts of white noise with durations between 5 and 500 ms, found target detection very difficult for distances of more than 2 m.

In the present study, we focus on reflections from 'on-axis' objects, directly located in front of the user at a negligible elevation. There are different aspects involved in their detection, depending on the sound used. For impulsive signals, like oral clicks, the ability to discriminate the reflected from the direct sound as a separate event, referred to as "echo", depends on the time resolving abilities of the auditory system. It is well known that a loud signal (masker) is able to mask a softer impulsive signal even after the first has stopped playing. This effect, referred to as post-masking or forward masking [15, 16] is associated to neural relaxation processes and its impact decreases with time. In the case of stationary sounds, the reflected sound wave interferes with the direct sound and thus changes several qualities of the sound, e.g. the spaciousness, loudness and coloration [6]. Bilsen and Ritsma [17] called *repetition pitch* the tonal character emerging from the combination of a sound and an additional reflection delayed a few milliseconds. Binaural hearing also plays an important role in the perception of spaciousness [18, p.329], more specifically in the qualities of apparent source width and listener envelopment [19]. It is also related to the off-axis detection of objects that are not located in the frontal direction [20], due to the evaluation of interaural level, phase and time differences. Moreover, binaural hearing also integrates the directional information arriving from multiple reflections through the so-called precedence effect[21]. However, since the focus of the present study is on detection of on-axis reflections, these important binaural effects are not taken into account.

As mentioned earlier, the perception of coloration can be linked to the presence of a single reflection. The Coloration Detection Thresholds (CDTs) predicted and measured by Buchholz [22] increase with increasing delay of the reflection, meaning that detection becomes more difficult with longer delays. CDTs decrease as the bandwidth of the signal increases, although low frequencies contribute more than high frequencies to coloration detection. The use of band-pass filtered noise from 3 kHz to 5 kHz as a test signal (comparable to the spectrum of an /s/ sound) resulted in CDTs between -10 dB reflected-to-direct level difference (RDLD) for a direct-to-reflected delay of 2 ms and -5 dB for a delay of 4 ms. Ando and Alrutz [23] measured CDTs between -12 dB and -15 dB for delays of 5 to 40 ms, with lower values for shorter delays. They used stationary Gaussian noise band-pass filtered at a centre frequency of 4 kHz with a bandwidth of 1.5 kHz. In more realistic scenarios where reflections arrive from different directions than the one of the direct sound, coloration is more difficult to detect with two ears than with one ear, due to the effect of binaural decoloration [9, 24].

Another mechanism for detecting the presence of a reflecting object is the detection of changes in loudness, as a single reflection increases the loudness of the direct sound in isolation. Houtsma *et al.* [25] found the Just Noticeable Difference (JND) for white noise to be approximately 0.6 to 0.7 dB, rather independently of presentation level when this is above 20 dB SPL. These JNDs would correspond to RDLDs between -8.5 dB and -7.5 dB if the change in level was due to an added reflection.

The experiments mentioned above were conducted for external sounds such as recorded samples of music, speech or synthesized noise played back through loudspeakers or headphones, but not for self-generated sounds, which reach the ear through the air and through body conduction. External and self-generated sounds are processed differently, as the stapedius reflex may change the sensitivity of the auditory system to airborne sounds but not to body-conducted ones. Threshold experiments with self-generated sounds were carried out by Rice *et al* [26], who investigated the ability of echolocators to detect circular targets with diameters between 2.5 cm and 38 cm at distances between 0.6 and 2.7 m (corresponding to delays between 3 and 16 ms). Detection thresholds were in the order of -20 dB RDLD [27], independent of target distance. In similar experiments, Kellogg [28] studied the Just Noticeable Differences (JND) in size and distance while echolocating circular targets of diameters between 15 and 30 cm and at distances between 0.3 and 1.2 m. It is worth pointing out that, in the context of human echolocation, shorter signals do not always lead to lower thresholds or higher detection rates (e.g. [3, 13, 14]).

Recent works [29–31] have studied human echolocation abilities by using virtual acoustics systems that replicate the effect of reflections from walls and obstacles on self-generated sounds by means of low-latency convolution. More specifically, delay discrimination

in the frontal direction [29], localization of leading and lagging pairs of reflections [30] and sensitivity to changes in room size [31] were studied. With these systems, the simulated reflection patterns can be characterized with an equivalent impulse response between a point in front of the mouth and the entrance of the left and right ear canals. This response is referred to as Oral-Binaural Room Impulse Response (OBRIR) [32] and can be measured with a dummy head having microphones at the ears and a loudspeaker at the mouth position. In the particular case of a simulated single reflection, given the measured OBRIR (which will contain the actual airborne direct sound and the reflection) one can derive the RDLD to compare the intensity of reflected and direct sound as described in e.g. [27]. From measurements with small circular targets of radius r_d like the ones used by Rice *et al* [26] at a distance r_s between 0.5 and 2.5 m in front of a person, an empirical model relating the RDLD to the solid angle $\Omega \approx \pi(r_d/r_s)^2$ occupied by the target [27] is

$$\mathrm{RDLD} = 15.5 \log_{10} \Omega + 15.0[\mathrm{dB}]. \quad (1)$$

The present work aims at bridging knowledge from psychoacoustics and echolocation literature by measuring the audibility thresholds of sighted untrained persons for a single simulated reflection using external (Experiment I) and self-generated sounds (Experiment II). In this way, we focus on the auditory basis of echolocation. In addition, we aim at understanding the differences in performance when using self-generated or external sounds of different durations. We express the thresholds in terms of acoustic quantities (RDLD) instead of physical characteristics of the reflecting objects (e.g. size, texture) to accommodate the experiments and training of echolocation performed with virtual acoustics systems, for which there are no actual reflecting objects. We study a range of reflection distances up to 16 m, much larger than previous echolocation experiments, limited to approximately 3 m. This is done purposefully for obtaining long delays in the order of 100 ms that link echolocation to subjective room acoustics studies. At the same time, such delays are realistic for the detection of landmarks such as buildings or walls via echolocation in open air spaces with low density of obstacles. The thresholds may also be regarded as a reference performance on which to evaluate that of other groups such as early or late blind experts, the effect of some particular training or the degradation linked to some environmental maskers such as background noise or different reverberant conditions.

2 Method

Two related experiments for determining audibility thresholds for a single reflection in an echolocation-like setting were carried out: in Experiment I, the source signals were pre-recorded (external) oral sounds, whereas in Experiment II the oral sounds were self-generated and the reflections were added in real-time with a virtual acoustics system. These thresholds are expressed in terms of the Reflected-to-Direct Level Difference (RDLD) [27], which compares the strength of direct and reflected sound at the ears of a person. This quantity takes into account the effects of typical signal spectrum, mouth-ear airborne direct sound propagation, mouth directivity and diffuse field head-related transfer function. The depicted echolocation-like setting is idealistic as it contains no sound reflections other than the simulated target reflection nor interfering background noise.

2.1 Participants

Participants in the experiments were sighted persons, between 23 and 48 years old, with normal hearing (HL $<$ 20 dB from 250 Hz to 8 kHz following audiometric screening). They were workers or students at the laboratory and were told about the goals of the experiment beforehand. They did not receive any compensation for the participation. In both experiments, there were 12 participants: one woman and 11 men in Experiment I, and 2 women and 10 men in Experiment II. Four of the participants in Experiment I took part in Experiment II a few months later. Wilcoxon signed-rank tests showed that their RDLD thresholds did not differ significantly from the average RDLD thresholds neither in Experiment I ($z = -1.69; p = 0.10$) nor in Experiment II ($z = 0.088; p = 0.93$). Therefore, there were no observable training effects. 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 Experimental procedure

The experiments were conducted in a computerized and automated way in the 125 m^3 semi-anechoic chamber at KU Leuven with additional absorbing panels lined on the floor to absorb effectively the range of frequencies of oral clicks and hissing sounds. Subjects seated in the center of the room and were instructed on the tasks. No training was given to the participants prior to the start of the tests. However, the experimenter monitored the start of the test. If anomalies were observed, the procedure was explained once again more thoroughly and the experiment was restarted.

For each trial in the experiments, the task of the participants was to identify the single interval out of three containing a reflection, in addition to the direct sound present in all intervals. The interval containing the reflection was always assigned randomly and the choice of the participant was forced to be one of

Table 1: Summary of experimental conditions, characterized by the signal (OC = oral click, /s/ = sustained /s/ sound) and the distance of the simulated wall.

Experiment	Signals	Distances (m)	Number of conditions
Exp. I	OC 37 dB, OC 52 dB, /s/ 63 dB	0.5, 1, 2, 4, 8	15
Exp. II	OC, /s/	1, 2, 4, 8, 16	10

the three intervals. Hence the 3-interval, alternative forced choice (3AFC) paradigm [33] was used. By varying the gain G, the RDLD of the reflection was modified in the next trial following an adaptive one-up two-down rule [34]. One wrong answer increased the RDLD, while two consecutive correct answers decreased it. The steps by which RDLD varied were progressively reduced from 8 to 1 dB, with 4 and 2 dB as intermediate steps. The result of a run was the RDLD threshold, defined as the value to which the answers of the participant converged, and obtained as the average of the last 6 reversal points. Because of the one-up two-down rule used, this value theoretically corresponds to a 70.7% probability of correct responses [34].

In Experiment I, participants seated in front of a computer screen with a GUI showing the 3 alternative choices on buttons that they could click to provide their answer. Participants listened sequentially to the three intervals only once and made their choice. Visual feedback (correct/wrong) was given on the choice.

In Experiment II, participants interacted with the setup through a remote controller, which offered actionable buttons to switch intervals during the 3-AFC procedure and to take a decision about the interval. Differently from Experiment I, participants could freely choose and repeat any of the intervals for as long as desired. Feedback to the participant (on the signal to produce, the selected interval and the chosen answer) was exclusively acoustic.

A summary of the experimental conditions presented to the participants is shown in Table 1. For each participant in Experiment I, the different conditions were tested sequentially: all distances were tested in increasing order with an oral click, and then again with a sustained /s/. In Experiment II, all conditions were fully randomized. The distance of 0.5 m could not be tested in Experiment II because the latency of the experimental setup was higher than the delay of that reflection. Instead of 0.5 m, an additional distance of 16 m was tested.

2.3 Stimuli

The stimuli were generated by convolution of computed OBRIRs and two source signals. The OBRIRs were derived for a perfectly reflecting wall with infinite extension in its physical dimensions located at different distances. This theoretical object was chosen because it provides reflections with the same spectrum as the incident sound. The source signals were an oral click and an /s/ sound. Experiments I and II differed from each other in three main aspects. First, in the way stimuli were generated: sounds were pre-recorded in Experiment I, whereas they were self-generated by participants themselves in Experiment II. Second, in the way convolution was performed: offline in Experiment I and real-time in Experiment II. And third, in the apparatus required to process the sound and perform the tests, as we shall explain later.

2.3.1 OBRIR calculation

The OBRIRs were calculated with the Fast Multipole Boundary Element Method (FM-BEM), using the FastBEM software package [35]. A three-dimensional model of a human head, taken from the OpenHear database [36], was used. Its geometry was simplified to contain 20,000 vertices and 39,996 triangular elements, with an average node separation of 4.4 mm. The mouth source was modelled as a monopole at the center of the lips. The reflecting infinite wall was modeled by defining a symmetry plane at the wall position. The pressures generated at the mouth reference point (MRP), 3 cm in front of the center of the mouth, and at the entrance of the blocked ear canals were calculated frequency by frequency, from 40 Hz to 12 kHz in steps of 40 Hz. With this sampling of the frequency domain, it was possible to reconstruct 25 ms of the OBRIR by inverse Fourier transform of the frequency response at a sampling frequency of 24 kHz. The resulting OBRIRs contained reflections from objects up to a distance of 2 m. Time windowing was further used to remove aliases. The window applied contained ones between 0 and 20 ms and half a cycle of a raised cosine decaying from 1 to 0 between 20 and 25 ms. In this way, the OBRIRs for the anechoic case $h_{ane}(t)$, and in the presence of reflecting planes at 0.5 m $h_{0.5}(t)$, 1 m $h_1(t)$, and 2 m $h_2(t)$ in front of the mouth, were calculated. The OBRIRs for reflection only, $h_d^r(t)$, were calculated by subtraction:

$$h_d^r(t) = h_d(t) - h_{ane}(t), \tag{2}$$

where d corresponds to any of the distances of 0.5, 1 or 2 m. Moreover, reflections from 4, 8 and 16 m were generated by applying a delay and a scaling factor to the reflection at 2 m using a far-field approximation:

$$h_d^r(t) \approx \frac{2}{d} h_2^r \left(t - \frac{2(d-2)}{c} \right), \tag{3}$$

with a speed of sound $c = 343$ m/s. As an approximation, no air absorption was considered. The OBRIRs were then up-sampled to 48 kHz. The OBRIRs of direct and reflected sound paths, together with their

frequency response analysed in 1/3rd octave bands are shown in Figure 1 (a) and (b). Note that the frequency responses of the reflected sound paths for the shortest distances up to 2 m were almost identical, with only minor fluctuations in the lowest frequency bands.

2.3.2 Source signals

Experiment I An oral click and a sustained long /s/ were generated by a person knowledgeable in echolocation and were recorded at a sampling frequency of 48 kHz in an anechoic chamber using an omnidirectional microphone Behringer ECM8000 at 50 cm in front of his mouth. These sounds, shown in Figure 2, were used as source material for all participants in Experiment I, and were then convolved with the OBRIRs previously defined.

The oral click was adjusted to produce a sound exposure level (SEL) of 37 dB or 52 dB (corresponding to a peak SPL of either 70 dB or 85 dB) at the entrance of the closed ear canal after applying the anechoic OBRIR, when reproduced through headphones. These levels correspond to soft and loud clicks following the subjective assessment of the authors of the study. Each click had a duration T_{click} of 3.1 ms, a frequency of maximum magnitude $F_{\max}$ of 1.8 kHz and a bandwidth B of 2.6 kHz, according to the definitions in section 2.4. The /s/ sound had $F_{\max} = 4.5$ kHz and $B = 6.5$ kHz, and was adjusted to produce an equivalent SPL of 63 dB, approximately equivalent to a normal vocal effort [37]. We chose only one level because a change in the presentation level of a broadband stationary sound does not influence coloration detection [38] nor loudness JND [25]. Here, we implicitly assumed that the detection of reflections with the /s/ signal was done on coloration and loudness changes. Attack and decay features were not relevant because only steady state of the /s/ sound was recorded and played back.

Experiment II Participants wore a Sennheiser MK2-P miniature microphone with the capsule at the mouth reference point. It captured the oral sounds generated by the participants themselves at the levels they found most suitable.

2.3.3 Stimulus generation

Reference and test stimuli, or signals, were generated. The reference stimulus contained only the direct sound, whereas the test stimulus contained both the direct and the reflected sound. Experiments I and II differed in the equipment used and the processing applied to the source signals.

Experiment I A computer running Matlab R2006b, and PsyLab 2.5 [39] was used for this experiment. It delivered the sound through Sennheiser HD 650 headphones connected to an audio interface RME Fireface UCX.

The test signals were generated according to the top branch of the diagram in Figure 3. The source signals were convolved with an OBRIR resulting from the combination of the direct component $h_{ane}(t)$ and a scaled reflected component $10^{G/20}h_d^r(t)$. The gain G was controlled by the one-up two-down procedure to obtain the desired RDLD. The reference signal, in the lower branch of the same figure, was always obtained by convolution of the recorded source signal with $h_{ane}(t)$. The three intervals had a length of 1 s (the oral clicks were zero-padded to this length) and were presented sequentially, with a pause of 200 ms between them. In the case of the /s/ sound, the signals had a smooth fade-in and fade-out of 100 ms each.

Experiment II The setup of Figure 4 was used to generate stimuli in Experiment II. The oral sounds of the participant were picked up with the microphone and sent through a sound interface RME Fireface UCX to a laptop running OS X in order to add a simulated reflection. As participants generated the sound themselves, the direct sound component was always naturally present. A custom program written in Max 6.1 convolved the oral sounds with the OBRIRs of the reflections only, $h_d^r(t)$, and adjusted the output gain. This program made use of an already existing low-latency, non-uniformly partitioned convolution implementation [40] as the core of the convolution engine. The convolved output was sent to a mixer and then routed to open headphones Sennheiser HD650. Though these headphones were open, they introduced high frequency attenuation. For this reason, an equalizer DBX1231 was introduced in the playback chain. Its purpose was to filter and amplify the direct oral sound generated by a user, so that the spectrum of this sound measured at the ears with or without headphones would be identical. The latency introduced by the equalizer was negligible, which according to Pörschmann is ideal for own-voice auralization [41]. For the added reflection, the total input-output latency measured between the direct sound and the output at the ears, using a Dirac delta as the OBRIR, was 3.5 ms. This means that reflections from objects closer than 0.6 m to the user could not be simulated. The initial taps of the computed OBRIRs were trimmed to account for the latency.

A Matlab program (shown as *Matlab experimental control* in Figure 4) controlled the experimental flow and communicated with the custom Max program via OSC protocol [42]. This program also adjusted the OBRIR to be used in the convolution (i.e. the distance of the obstacle) and the output gain. For the test intervals, the output gain adjusted the reflection level to the desired RDLD, and for the reference intervals, the output gain was set to $-\infty$ dB to mute

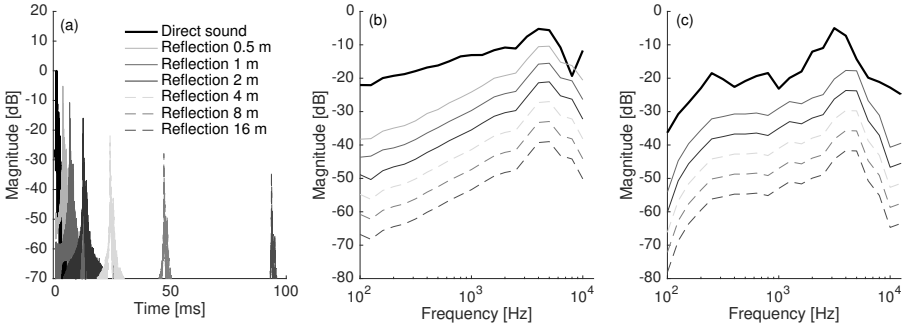

Figure 1: (a) Calculated Oral-Binaural Room Impulse Responses (left ear only) corresponding to the direct sound path and the reflected sound path from an infinite reflecting wall at different distances, (b) calculated 1/3rd octave spectra and (c) 1/3rd octave spectra measured with the setup for Experiment II.

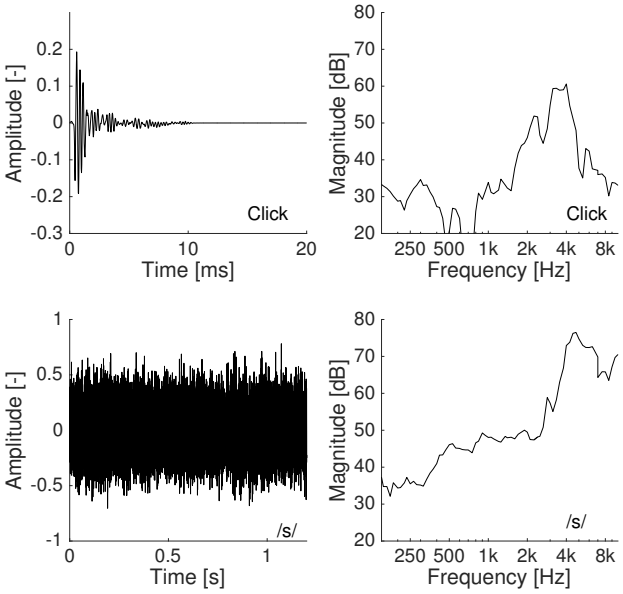

Figure 2: Waveform (left column) and 1/12th octave spectra (right column) of the recorded oral click (top row) and the /s/ sound (bottom row) used as source signals in Experiment I.

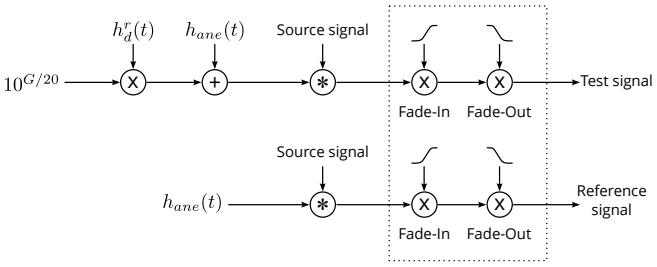

Figure 3: Diagram for the generation of test and reference signals in Experiment I. The value of G was controlled by the one-up two-down procedure and the elements within the dotted box were only used for the /s/ sounds.

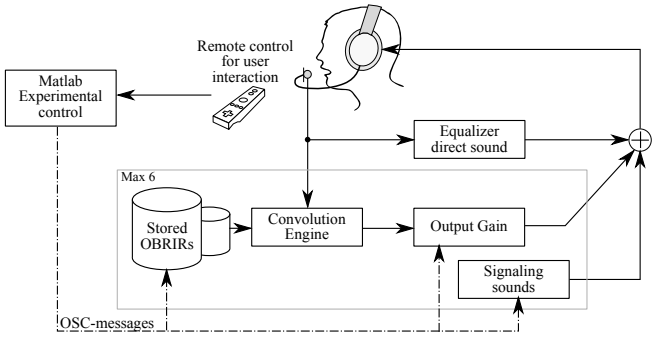

Figure 4: Setup for determination of reflection thresholds for self-generated oral sounds in Experiment II

Table 2: Overall RDLDs of the calculated OBRIRs for a 100% reflecting infinite wall at different distances.

Distance, m	0.5	1	2	4	8	16
RDLD, click-weighted, dB	-6.8	-12.1	-17.8	-23.7	-29.6	-35.6
RDLD, /s/-weighted, dB	-4.5	-9.6	-15.1	-21.0	-27.0	-33.0

the reflection. In addition the control program also triggered signaling sounds to guide the experiment.

2.3.4 Reference RDLD values

The overall RDLD of the OBRIRs in Experiment I (with the implicit assumption of $G = 0$ dB) is shown in Table 2. For calculating the single number RDLD value according to [27], the weighting spectra of a pre-recorded oral click and of a sustained /s/, in 1/3rd octave bands, were used (see section 2.3.2 for details on these signals). The observed variations of 2 to 3 dB in RDLD by using the two spectral weighting functions are within the uncertainty found in [27].

In Experiment II, the actual OBRIRs and RDLDs varied across participants due to individual differences in the mouth-ears path and the microphone/headphone placement on the head. Nevertheless, the actual OBRIRs were estimated through acoustic measurements according to [32], after mount-

ing the setup on a dummy head HEAD Acoustics HMS II.3, which is equipped with microphones at its ears. In addition, a loudspeaker JBL On Tour Micro II was mounted at its mouth. During the measurements, the simulated OBRIRs were set to $h_d^r(t), d = 1, 2 \dots 16$ in a sequential manner. The gain in the electroacoustic chain was kept constant across OBRIRs. The spectra of the direct and reflected components in the measured OBRIRs are displayed in Figure 1(c). Note that these spectra are different from the calculated ones (see Figure 1(b)) because of different reasons. On the one hand, the direct sound was actually generated by the measurement loudspeaker in combination with the EQ device. On the other hand, the reflected sound depended on the response and placement of the setup and measurement microphones and on the response of the headphones when placed on the ears.

The resulting RDLD values for the reflections in Experiment II were 2.2 dB higher than the RDLD values of the calculated OBRIRs shown in Table 2. All RDLD values were derived using the weighting spectra of the oral click and the /s/ sound used as source signals in Experiment I (shown in Figure 2).

Note that the RDLD values of Table 2 are just a reference, because the experiments increased or decreased the strength of the reflections (and thus the RDLD) until the audibility threshold for the reflection was obtained. A measured threshold, expressed in terms of RDLD, higher than the corresponding reference RDLD value of Table 2 would roughly mean that the participant would not be able to detect the infinite wall in a real echolocation task.

2.4 Analysis of emission properties

The temporal and spectral properties of clicks or hissing sounds have an impact on the psychoacoustic mechanisms of masking and coloration. Therefore, these properties may affect human echolocation performance and partially explain inter-subject variability of the results. For this reason, and to assess the repeatability of the signals, participants in Experiment II were asked to generate 10 oral clicks and 10 /s/ utterances representative of their style at the end of the experiment. The sounds were recorded with the headworn microphone located at the MRP (see Figure 4) and several parameters were derived:

- the sound exposure level (SEL or L_E) of the click and the equivalent SPL (L_{eq}) of the /s/ sound, as an indication of loudness.
- the duration of the emission (T_{click} for the click and $T_{/\text{s}/}$ for the /s/ sound), related to the period during which forward masking is active. For the click, the duration was calculated as the time interval where the amplitude of its envelope was higher than 1/4 of the peak value $A_{\max}$. This amount of decay was chosen to penalize "double clicks". For the /s/ sound, the duration was calculated as the time interval where the amplitude of the envelope was higher than 10% of its envelope rms value A_{rms}. This experimentally motivated choice accounted for large irregularities in the emissions of some participants.
- the attack time for the /s/ sound (T_{att}), defined as the time it took for the amplitude of the signal envelope to pass from 10% to 50% of the envelope rms value at the start of the sound.
- the decay time for the /s/ sound (T_{dec}), defined as the time it took for the amplitude of the signal to pass from 50% to 10% of the envelope rms value at the end of the sound. Both the attack and decay times were chosen as indicators of the transient character of the emissions, which could influence the detectability of reflections.
- the bandwidth (B), calculated as the absolute difference of the frequencies where the envelope of the 1/12th octave spectrum decayed to -12 dB respect to its maximum at frequency F_{max}. Note that this bandwidth was calculated from a spectral analysis (in energy per fractional octave band), which resembles closer the human auditory processing, in opposition with 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.

3 Results

The measured RDLD thresholds are shown in Table 3 and Figure 5(a) as a function of the distance to the simulated obstacles.

3.1 Experiment I

As shown in Figure 5(a), the RDLD thresholds for oral clicks in Experiment I decreased with increasing distance. Beyond 1 m, the RDLD thresholds decayed at about 6 dB per doubling of the distance to the obstacle. On the other hand, the RDLD thresholds for the /s/ sound at a level of 63 dB barely depended on distance, and they were higher than those obtained with the oral clicks at and beyond 1 m distance. The inter-participant standard deviation on the RDLD thresholds, also reported in Table 3, was approximately 3 dB (ranging between 2.0 dB and 4.2 dB depending on the condition).

A two-way Analysis of Variance (ANOVA) performed on the RDLD thresholds for the oral clicks using level and distance as the two factors indicates significant effects of both level ($F(1, 114) = 13.2; p < 0.001$) and distance ($F(4, 114) = 109; p < 0.001$) but

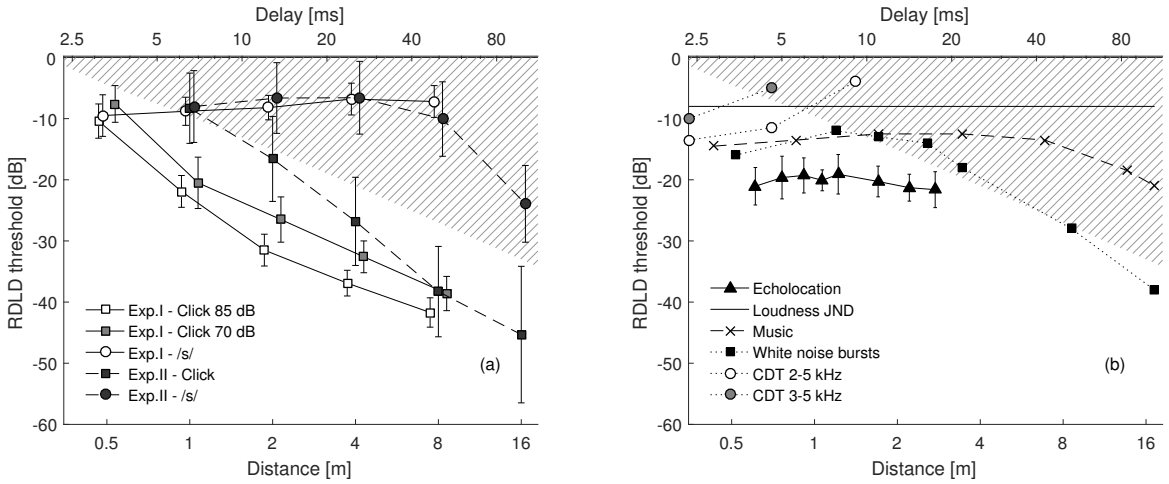

Figure 5: (a) Single reflection audibility RDLD thresholds for oral sounds as obtained in Experiments I and II for untrained sighted people, with error bars indicating ±1 standard deviations, and (b) other reference values derived from the literature: echolocation detection thresholds for a circular target [26, 27], loudness-based JND [25], single reflection thresholds for a music signal [8], diotic reflection masked thresholds for bursts of bandpass filtered (100-5000 Hz) white noise [9], CDTs [22] for bandpass noise from 2 to 5 kHz and from 3 to 5 kHz. Hatched areas indicate reflections stronger than provided by flat 100% reflecting infinite walls, considering the RDLD values of Table 2 for the oral click.

Table 3: Mean (and standard deviation) RDLD audibility thresholds for single reflections with external sounds (Experiment I) and self-generated sounds (Experiment II). Units are dB.

		Simulated obstacle distance					
Exp.	Signal	0.5 m	1 m	2 m	4 m	8 m	16 m
I	Oral click, $L_E = 37$ dB	-7.6 (3.0)	-20.5 (4.2)	-26.5 (3.7)	-32.6 (2.6)	-38.6 (2.8)	
	Oral click, $L_E = 52$ dB	-10.4 (2.8)	-21.9 (2.6)	-31.5 (2.6)	-36.9 (2.1)	-41.7 (2.4)	
	/s/, $L_{\mathrm{eq}} = 63$ dB	-9.5 (3.4)	-8.8 (2.3)	-8.2 (2.0)	-6.8 (2.6)	-7.3 (2.6)	
II	Oral click		-8.3 (5.7)	-16.6 (7.0)	-26.8 (7.2)	-38.3 (7.4)	-45.3 (11.2)
	/s/		-8.0 (5.9)	-6.6 (5.8)	-6.6 (5.9)	-10.1 (6.1)	-23.9 (6.3)

no significant interaction ($F(4, 110) = 0.26; p = 0.90$). Thus, the RDLD thresholds for the clicks at a sound exposure level of 52 dB were on average 3.3 dB lower than at 37 dB.

A one-way ANOVA on the RDLD thresholds for the /s/ sound with distance as the only factor, however, does not indicate a significant effect of distance ($F(4, 50) = 2.01; p = 0.11$).

3.2 Experiment II

The measured RDLD thresholds for self-generated oral clicks in Experiment II (see Figure 5(a)) significantly decreased with distance ($F(4, 50) = 40.5; p < 0.001$), in a similar way as in Experiment I. For the self-generated /s/ sounds, distance had a significant effect on the RDLD thresholds as revealed from an ANOVA test ($F(4, 50) = 16.5; p < 0.001$). Tukey-Kramer post-hoc tests indicate no significant differences in RDLD thresholds among distances from 1 m to 8 m, but only a significant difference between these conditions and the furthest distance of 16 m.

The inter-participant standard deviations for Experiment II (see Table 3) were much higher (approximately twice) than those for Experiment I, due to a higher degree of variability in stimuli generation among participants. Note that participants in Experiment II were required to generate sounds by themselves, whereas the same pre-recorded sounds were used by all participants in Experiment I, which was a pure listening test.

To illustrate the large variance of self-generated stimuli within and across participants, example waveforms and corresponding spectra of oral clicks for participants S3, S4 and S9 are shown in Figure 6. Example waveforms and spectra for /s/ sounds are shown in Figure 7 for participants S8, S10 and S12. Some basic statistical properties (mean, standard deviation, maximum and minimum) of the acoustical parameters described in section 2.4, linked to the oral clicks and the /s/ sounds, are shown in Table 4 and 5.

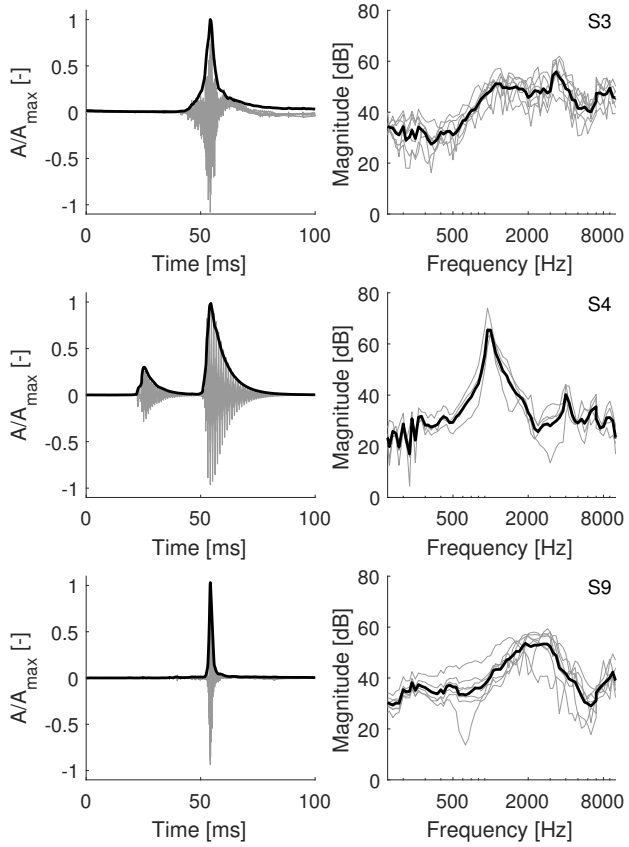

Figure 6: Waveform (left) and spectra (right) of oral clicks used by participants S3, S4 and S9 in Experiment II, recorded at the mouth reference point. Individual waveforms and spectra are shown in gray, while bold black lines display average waveform envelopes and average spectra.

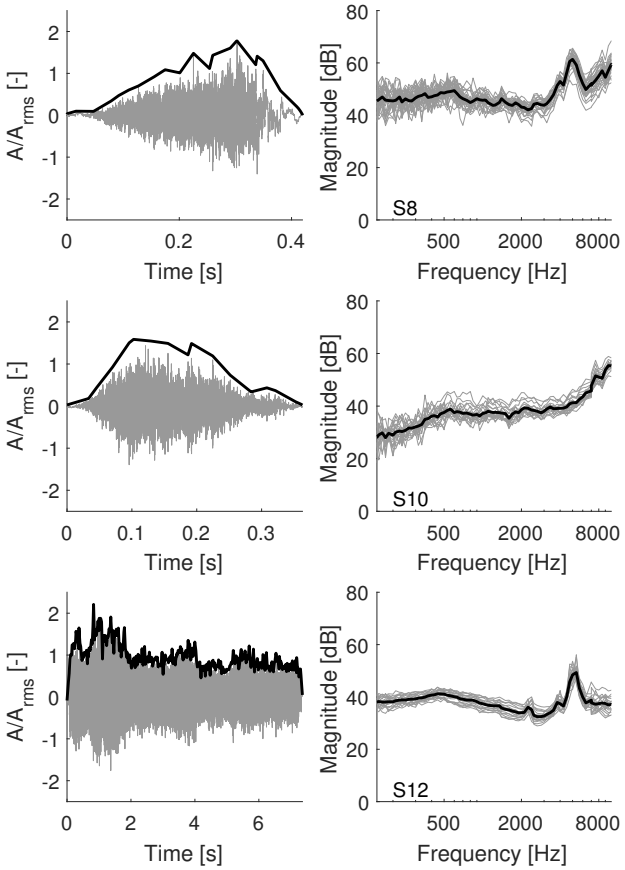

Figure 7: Waveform (left) and spectra (right) of /s/ sounds used by participants S8, S10 and S12 in Experiment II, recorded at the mouth reference point. Individual waveforms and spectra are shown in gray, while bold black lines display average waveform envelopes and average spectra.

A regression analysis was performed to get more insight into the causes for the large variation of RDLD thresholds among participants in Experiment II. We define $\widetilde{\text{RDLD}}$ as the difference between a particular RDLD threshold and the average RDLD threshold for that distance and that signal. Positive or negative $\widetilde{\text{RDLD}}$ values are interpreted as thresholds higher or lower than average. Simple linear regression models predicting $\widetilde{\text{RDLD}}$ from one signal parameter at a time were built.

For self-generated clicks, the only significant model was obtained with L_E as a predictor, shown in detail in Figure 8. It reports a decrease of 0.37 dB in $\widetilde{\text{RDLD}}$ for each additional dB in the click, though only 14.5% of the variance is explained. Thus, there is a slight trend that louder clicks help to have lower RDLD thresholds, as happens in Experiment I. In Experiment I, the average difference in RDLD threshold between the two click levels of 37 and 52 dB was 3.3 dB, corresponding to a decrease of 0.22 dB in RDLD threshold per additional dB of the stimuli. No significant effects were observed with other click parameters.

For the /s/ conditions in Experiment II, significant models (with $p < 0.05$) were found with predictors L_{eq} and T_{dec}. These models are shown in detail in Figure 9. As happens for oral clicks, $\widetilde{\text{RDLD}}$ significantly decreased with increasing level of the /s/ sound, at a rate of -0.32 dB/dB. Longer decay times resulted in higher thresholds, at a rate of 0.1 dB per each ms of increase in T_{dec}.

4 Discussion

In the following sections, we discuss the underlying psychoacoustic mechanisms and other factors that have an impact on the basic human echolocation task of detecting a single reflection in the absence of background noise or additional reverberation. We show several literature results on detection thresholds for a single reflection with different stimuli and conditions [8, 9, 22, 25–27] along with our results in Figure 5 to supplement the discussion. These literature results, despite corresponding to different research areas, reflect underlying psychoacoustic mechanisms such as

Table 4: Mean, standard deviation (SD), maximum and minimum values of the sound exposure level (L_E), duration T_{click}, peak frequency F_{max} and bandwidth B for self-generated oral clicks in Experiment II.

	L_E, dB	T_{click}, ms	F_{max}, kHz	B, kHz
Mean	48.0	12.0	2.9	3.8
SD	7.8	12.8	1.5	2.2
Maximum	59.0	38.9	7.3	8.4
Minimum	32.8	2.2	1.0	0.4

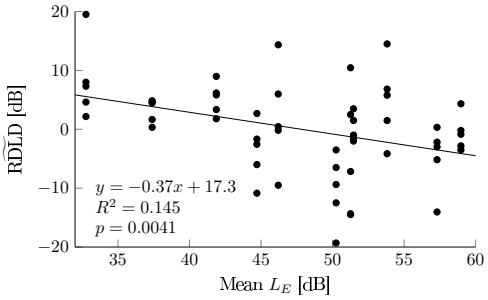

Figure 8: RDLD threshold deviations from per-distance averaged RDLD thresholds ($\widetilde{\text{RDLD}}$) versus oral click L_E in Experiment II. A linear regression model is shown with its equation, coefficient of determination R^2 and p-value.

post-masking, coloration and loudness that are also present in our experiments. From the knowledge of psychoacoustic limiting factors, and from the current results, we point out favorable clicking strategies that enhance the audibility of reflections.

4.1 Role of post-masking

The RDLD thresholds for oral clicks are characteristic of masked thresholds obtained in a forward masking or post-masking paradigm (e.g. [43]), where the masker is the direct sound, i.e. the oral click arriving directly from the mouth to the ears. For this impulsive signal, the direct sound typically does not overlap in time with the reflected sound, which arrives several milliseconds later. In post-masking, longer maskers result in higher masked thresholds. Moreover, shorter maskers with louder levels lead to faster decays of the masked thresholds [43]. The latter effect is observed in Figure 8 for Experiment II and in Figure 5(a) for Experiment I, where louder oral clicks (maskers) had lower RDLD thresholds. In Experiment I, there was an average reduction in measured RDLD thresholds of 3.3 dB between the results for oral clicks with $L_E = 37$ dB and the results for oral clicks with $L_E = 52$ dB. In Experiment II, there was a decrease of 0.4 dB in the average RDLD threshold for each additional dB in the oral click generated by the participant. The oral clicks had L_E ranging from 33 to 59 dB, with an average of 48 dB.

The difference of more than 12 dB in RDLD threshold between Experiments I and II when using the oral click at a distance of 1 m is remarkable. The most likely cause to explain this difference is the variation in oral click duration. Whereas the oral click used as source signal in Experiment I had a duration of 3.1 ms, recorded oral clicks in Experiment II lasted an average of 12 ms (as an indication, the delay of a reflection from 1 m is about 6 ms). This means that there was a significant overlap between direct and reflected sound in Experiment II. As a consequence, simultaneous masking may have been more important in Experiment II than in Experiment I. Thus, the RDLD threshold decay associated to post-masking in Experiment II was shifted towards longer delays, i.e. the decay started at 0.5 m in Experiment I and at 1 m in Experiment II.

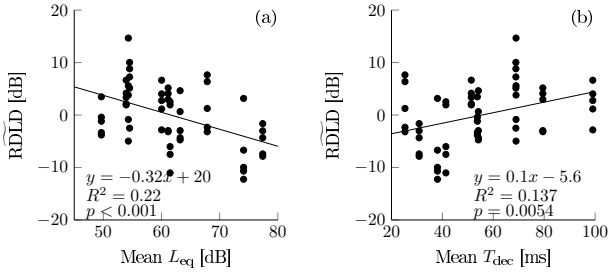

Figure 9: RDLD threshold deviations from per-distance averaged RDLD thresholds ($\widetilde{\text{RDLD}}$) versus (a) the average L_{eq}, and (b) the average decay time T_{dec}, using self-generated /s/ sounds in Experiment II. Linear regression models are shown with their equations, coefficient of determination R^2 and p-value.

These effects also explain the decrease of RDLD thresholds with /s/ sounds at 8 and 16 m in Experiment II. Some participants produced /s/ sounds with decay times T_{dec} as short as 25 ms, which were not simultaneously masking the offset of the reflections at distances of 8 and 16 m (corresponding to delays longer than 40 ms). A similar effect is shown in the reflection masked thresholds for bursts of bandpass filtered white noise [9] in Figure 5(b). Because of the longer signal duration, the RDLD threshold curve for the /s/ sound between 8 and 16 m corresponds to a delay-shifted version of the curves for the oral clicks. The regression model of Figure 9(b) supports this explanation, as it shows RDLD thresholds lower than average with shorter T_{dec}.

4.2 Role of coloration and loudness

In the case of the /s/ sound, the direct and the reflected sound greatly overlap in time. For this reason, one would expect detection based on loudness and coloration variations, as pointed out by Schenkman and Nilsson [44]. When loudness cues are mainly used, the reflected and the direct sound played together

Table 5: Mean, standard deviation (SD), maximum and minimum values of the equivalent SPL L_{eq}, duration $T_{/s/}$, attack time T_{att}, decay time T_{dec}, peak frequency F_{max} and bandwidth B for self-generated /s/ sounds in Experiment II.

	L_{eq},dB	$T_{/s/}$, s	T_{att}, ms	T_{dec}, ms	F_{max}, kHz	B, kHz
Mean	61.2	1.5	54.1	57.1	8.6	7.5
SD	8.4	2.0	15.0	21.8	1.9	2.2
Maximum	77.4	7.2	68.0	99.0	11.3	12.2
Minimum	49.7	0.3	23.0	25.0	5.2	4.7

must be approximately one JND louder than the direct sound in isolation. Assuming a JND is 0.6 to 0.7 dB (similar to that of white noise [25]), and that direct and reflected sound do not sum coherently, the RDLD threshold based on loudness detection would become -8 dB, as shown in Figure 5(b). The measured RDLD thresholds are in this order of magnitude (between -8.8 and -6.6 dB for distances between 1 and 4 m).

At the shortest distances, the direct sound and the reflection interact creating coloration or pitch sensation [17]. Buchholz [22] stated that high frequency components contribute less than low frequency components to detection of coloration. He reported Coloration Detection Thresholds (CDT) of band-pass filtered noise from 3 kHz to 5 kHz between -10 dB for a delay of 2 ms and -5 dB for a delay of 4 ms (see Figure 5(b)). In fact, the /s/ sound of Experiment I had very low energy at frequencies below 3 kHz. There are indications that coloration cues were also used at the shortest distance, as seen from the non-significant decrease in RDLD threshold for the /s/ sound at 0.5 m. In sum, the measured RDLD thresholds for the /s/ sounds at distances below 8 m could be qualitatively obtained by taking the minimum of the CDT and loudness JND curves at each distance.

In Figure 5(b), it is shown that the CDT thresholds measured by Buchholz [22] got lower with increasing bandwidth towards lower frequencies. In fact, the large inter-participant variations in bandwidth for the /s/ sound (SD = 2.2 kHz, see Table 5) might have contributed to facilitate or complicate the coloration detection in these sounds. As a result, the measured RDLD thresholds for /s/ sounds in Experiment II had larger standard deviation than those in Experiment I.

The fact that RDLD thresholds for oral clicks were not lower than for /s/ sounds at the shortest distances indicates that coloration and loudness cues were also used for reflection detection when using oral clicks. Indeed, after the experiments, some participants reported that their detection strategy with clicks was based on a change of tonal character when they could not hear the reflection as a separated event.

It is worth pointing out the qualitative similarity of the detection thresholds obtained for music signals [8] (see Figure 5(b)) and those measured for /s/ sounds when the transient characteristics of the signals had an effect. In [8], thresholds had a maximum at reflection delays of around 20 ms, with better thresholds at shorter delays, due to a more favorable coloration detection, and at longer delays, due to the lower masked thresholds existing after signal decays.

4.3 External vs self-generated sounds

The RDLD thresholds for self-generated clicks (Experiment II) were higher than for external clicks (Experiment I) at short distances of 2 m and below. As indicated in section 4.1, differences can be ascribed to the larger time overlap and masking between direct and reflected sound in Experiment II, because clicks lasted longer in the latter (average of 12 ms) than in Experiment I (3.1 ms).

The contribution of bone conduction (BC) is differing between Experiments I and II. However, the impact of BC for voiceless sounds such as the /s/ and the oral click is rather limited. Above 3 kHz, the BC component for /s/ is about 5 dB below direct airborne sound. For sounds which are more similar to the oral click such as /k/ or /t/, the BC is about 10 dB lower at high frequencies than direct airborne sound [45]. The lack of differences in RDLD threshold across experiments when using the /s/ sound (except for the 16 m distance condition), as seen in Figure 5(a), supports the hypothesis that BC had little influence on the direct sound.

For the loudest clicks, the stapedius reflex might have been activated. However, we can safely assume that the reflex did not influence the RDLD thresholds obtained with clicks at 8 m and below, because it has a latency in the order of 100 ms [46]. This latency corresponds to a reflection distance of 17 m. The activation of the reflex for some subjects may have contributed to the higher standard deviation of the RDLD threshold at 16 m. Moreover, as the reflex only acts on signals louder than about 70 or 80 dB at particular frequencies above hearing threshold [46], we can reasonably assume that it did not influence the RDLD thresholds for /s/ sounds.

4.4 Physical interpretation

The hatched areas in Figure 5 indicate reflections stronger than those provided by infinite flat 100% reflecting walls. The lower bounds of the hatched areas are the lowest values indicated in Table 2 with the

spectral weighting of an oral click. RDLD thresholds laying under this area mean that, on average, participants would not be able to detect a reflection from such a wall in free field, or that detection would only be possible if the wall amplified the sound. RDLD thresholds below the hatched area mean that participants would be able to detect the wall. The difference G between the RDLD thresholds and the hatch lower bound can be linked to the minimum absorption coefficient α of the wall that can be detected, through the relationship $G = 10\log(1-\alpha)$. As a side note, diffraction effects of finite size elements and concave geometries of reflectors can increase the values of Table 2 at some frequencies, provided that the mouth and the ears are perfectly aligned with the axis of the reflector. In this case, the lower bound of the hatched areas in Figure 5 could vary by several dB.

As seen in in Figure 5(b), the RDLDs corresponding to loudness-based JND and CDTs are under the hatched area for distances longer than 1 m. This means that, for long distances (beyond approximately 2 m), the combined loudness and coloration information provided by the reflection of an infinite wall is not easily detectable; detection of a wall reflection is mostly based on detection of separate events (i.e. echoes) as shown in the results for the clicks in Figure 5(a).

In the context of room acoustics, by using music signals [8], reflections from walls beyond 2 m cannot be detected (Figure 5(b)). This is contradicting the knowledge that early reflections are important in the design of performance spaces (e.g. theaters and concert halls). In our analysis, we consider only a first reflection in free field. In realistic spaces, the combination of a wall with other surfaces would increase the strength of the reflected sound by several dB, compared to the reflected sound produced by the wall in free field.

Using the relation between RDLD and solid angle Ω as given in Eq. (1), and the relation between Ω, the distance r_s and the radius r_d of a circular rigid target $\Omega \approx \pi(r_d/r_s)^2$, some rough estimations of the smallest detectable object size can be given from the RDLD threshold measurements. These results should be considered with care, because Eq. (1) has been derived for distances below 2.5 m and for an accurate alignment between the mouth, the ears and the center of the reflector. RDLD thresholds independent of obstacle distance, as for /s/ sounds, correspond to detection of obstacles with constant solid angle. Using /s/ sounds, participants would have been able to detect circular solid targets of 20 cm diameter at 1 m, 44 cm diameter at 2 m, 88 cm diameter at 4 m and of 141 cm diameter at 8 m. In Experiment I, the finest average acuity was found for the loudest clicks at 1 m distance, such that RDLD thresholds correspond to reflections from disks of 7.2 cm diameter. In the case of Experiment II with oral clicks, the RDLD thresholds decrease with increasing obstacle distance, and correspond approximately to the reflection from a solid disk of 20 cm diameter at distances between 1 and 16 m. However, there are several factors in real life that limit the detection of such small objects: first, as distance increases, alignment to the object is less likely and off-axis reflections are much fainter. Second, as RDLD diminishes for smaller objects and longer distances because of the inverse square law (i.e. the theoretical decrease of SPL by 6 dB when doubling distance, as shown in Table 2), background noise is more likely to mask reflections. Third, in real environments there will likely be reflections much stronger than those from the particular target. Moreover, the previous results describe the auditory ability to detect reflections, which is just prerequisite to the much more complex conscious ability of interpreting these reflections to acquire perceptual knowledge of the environment, as done by experienced echolocators. The study of the latter ability is beyond the scope of the present research. From the experiments, we can affirm that subjects had enough auditory sensitivity to detect reflections from objects of e.g. 20 cm diameter at 16 m located on-axis with very low background noise and without any additional reverberation, but we cannot conclude that subjects would be able to perceive these objects in real environments. At the same time, the present results indicate a high likelihood that large distant objects in outdoor settings like walls or buildings may be detectable by means of echolocation, as long as the reflection is audible and not masked by background noise.

4.5 Role of source signal

There was a clear difference in performance depending on whether a participant used impulsive (e.g. clicks) or stationary (e.g. /s/) sounds. Some previous studies aimed at pointing out "good" echolocation signals [12, 47] based on prior ideas that loud signals with short duration, good spectral content and other features are desirable. The results of the current study point out that louder as well as shorter click signals result in lower RDLD thresholds. Thus, louder clicks allow hearing reflections from further obstacles by improving the SNR and keeping the reflection level above the hearing threshold or the background noise level. Moreover, louder and shorter clicks benefit from a faster decay of forward masking and with lower masked thresholds. Nevertheless, detection is still possible with softer clicks, which are less intrusive signals in noise-sensitive situations as indoor environments like libraries. Echolocation users may adjust the loudness of their emission as required in each situation.

For improving detection at short distances, signals should have an extended bandwidth towards lower frequencies to increase the probability of coloration

detection (see section 4.2).

4.6 Role of expertise

The adjustment of signal production to optimize detection is one of the key aspects differentiating skilled and novice echolocators. At the same time, auditory processing of the generated sounds is a crucial aspect. Auditory training seems important in order to achieve a good consistent echolocation performance [30, 48, 49]. In the current study, the focus was on untrained sighted participants. Thus, the measured RDLD thresholds can be seen as a baseline detection performance for this particular group of participants before starting any echolocation training. Further research should look into the thresholds of a group of experienced blind echolocators and into the improvement that can be achieved through sustained training.

Rice *et al*'s [26, 27] thresholds obtained from blind experienced echolocators using the signals of their choice (e.g. clicks, hissing sounds, words or syllables) are shown in Figure 5(b). They were rather flat with distance and, at distances below 1 m, they were about 10 dB lower than those obtained in Experiment II with untrained sighted participants. In the case of [26], the solid angle covered by the obstacles was roughly $5.1{\cdot}10^{-3}$ sr, which corresponds to a solid disk of 4.8 cm diameter at 0.6 m distance, or 24 cm diameter at 3 m distance. It could be hypothesized that four effects were taking place in the flat characteristic of the RDLD thresholds with distance: first, participants in [26] could freely choose the signal to use and therefore used a "best signal" strategy, e.g. hissing sounds with extended bandwidth at short distances and loud short clicks at longer distances. Second, participants in [26] were blind and more experienced in echolocation, which resulted in higher acuity [3]. Third, thresholds at longer distances increased due to the presence of additional reflections from non-anechoic boundaries, which could have made the task more difficult. However, several studies [14, 49] found better performance in normal rooms than in anechoic rooms. A last hypothesis for the flat response is that, as distance increased, the misalignment between the echolocator and the target increased, which would result into a much weaker reflection at the place of the emission. In that case, the effective RDLD of the misaligned targets would be lower than the on-axis RDLD plotted in Figure 5(b), measured in [27], and the thresholds for participants in [26] would actually decrease with distance, as in the current experiments. Note that this last hypothesis, in case of being accurate, points out at the difficulty of controlling, reproducing and extracting psychoacoustic information from echolocation experiments with small targets in real settings.

Although the present study found quite remarkable detection ability for untrained subjects, no indication is given about the ability to interpret reflections in real life. Expertise in echolocation, beyond the ability to generate signals and detect reflections, is achieved through training in such way that users are able to interpret reflection patterns and effectively perceive the environment. Some studies show that sighted subjects can improve echolocation skills such as distance perception [49] and size discrimination [50] through training.

4.7 Shortcomings

In Experiment I, conditions were not randomized across participants. If there were short-term training effects, thresholds for longer distances would be lower than for shorter ones, compared to Experiment II. However, RDLD thresholds for /s/ sounds in Experiment I did not improve with distance nor were any better than in Experiment II, where the conditions were randomized. Thus, there were no noticeable training effects during the time of the experiments. Future research should look into the improvement of RDLD thresholds over a number of training sessions.

It is worth mentioning that the strength of the reflections, relative to that of the direct sound, was characterized with the RDLD. In [27], we point out that there are a number of sources of variability in RDLD that change across participants: the propagation of direct sound from the mouth to the ears, the directivity of the radiated sound, the filtering of the head to incident sound, and the spectral balance of the emitted sound. All these factors lead to variations in actual RDLD across individuals, with a standard deviation of approximately 3 dB. In this study, we have not focused on individual RDLD values. However, a more detailed study could look into the role of actual individual RDLD values on performance and give personalized advice on how to tune echolocation signals to maximize the detection probability of typical reflections.

Nonetheless, the definition of RDLD [27] as a weighted measure in which the spectral regions with more power determine the overall value has its limitations. The auditory system analyzes and integrates changes within frequency bands across time [51]. Thus, it could be possible to use appropriate auditory models to predict more accurately the audibility of reflections and to identify suitable signals for echolocation, adapted to each scenario and each individual. However, the straightforward RDLD measure does allow to get a valuable initial insight in the matter of audibility of echoes, which can inspire investigation strategies by more complex models.

It is important to point out that the detection results of the present study are thresholds obtained through comparisons of intervals with and without reflections. In realistic scenarios, comparisons with direct-sound only are not feasible and therefore sub-

jects have to judge relative variations in reflection patterns following position or orientation changes. Further work should look into how the present results apply in real life scenarios that include finite size objects, room reverberation and background noise.

5 Final remarks

The detection thresholds of single reflections using echolocation signals with negligible background noise were determined by means of psychophysical experiments. The thresholds, expressed in terms of reflected-to-direct level differences (RDLD), were assessed for oral clicks and sustained /s/ sounds in two situations: when these signals were externally generated, and when participants produced the sounds themselves. RDLD thresholds with self-generated sounds were not better than with externally generated sounds. Moreover, due to large differences in the characteristics of the self-generated stimuli, the standard deviation for the RDLD thresholds was double as large as for the thresholds obtained with the externally generated stimuli. Participants achieved more accurate detection using oral clicks rather than stationary sounds at distances beyond 1 m.

For stationary sounds, detection was based on a combination of coloration and loudness cues. For reflection distances shorter than 2 m, coloration plays a role in detection and thus, broadband echolocation signals with enough low frequency content are preferable. For longer distances, thresholds were constant and consistent with expectations based on loudness JND, with the exception of the longest distances. Beyond 8 m, a sudden stop of the signal allowed participants to detect the reflection offset and obtain lower RDLD thresholds.

For oral clicks, the main limitation lies in forward masking, where the masker signal is the self-generated direct sound. For reflections from distances further than 1 m, the RDLD threshold decays 6 dB per doubling of the distance. For this reason, it is important to produce clicks as short as possible. In addition, louder clicks contribute to decreasing the RDLD thresholds due to a faster decay of forward masking, and thus to detecting smaller objects. Previous threshold experiments in echolocation had reported that the minimum detectable object size had a constant solid angle with object distance. On the contrary, the present experiments yielded thresholds that roughly correspond to a constant object size for distances between 1 and 16 m, namely a solid disk of 20 cm diameter perfectly aligned with the mouth and the ears of the echolocator. However, it is quite unlikely to observe such fine acuity in real environments due to the misalignment with the object and the presence of background noise and competing reflections. At the same time, the present results indicate a high likelihood that large distant objects in outdoor settings like walls or buildings may be detectable by means of echolocation, as long as the reflection is audible and not masked by background noise.

The gap between the measured detection thresholds with self-generated signals at distances below 2 m and those of previous detection experiments, where blind people performed better, is most likely due to the effects of experience and signal used. Therefore, untrained people may improve their echolocation detection performance through sustained practice of auditory and signal production skills.

Acknowledgements

This research work was carried out at the Laboratory of Acoustics, Department of Physics and Astronomy, of KU Leuven, in the frame of the Postdoctoral Fellowship of the Research Foundation Flanders (FWO-Vlaanderen) with grant no. 1280413N. We thank two anonymous reviewers for their extensive and constructive criticism that helped improving the quality of this article.

References

[1] N. Fletcher, "Animal Bioacoustics," in *Springer Handbook of Acoustics* (T. D. Rossing, ed.), ch. 19, pp. 785–804, New York, NY: Springer, 2007.

[2] D. Ashmead, R. S. Wall, S. B. Eaton, K. A. Ebinger, M.-M. Snook-Hill, D. A. 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.

[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, 2014.

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

[5] H. Kuttruff, "Subjective Room Acoustics," in *Room Acoustics*, pp. 204–250, London and New York: Taylor & Francis and Spon Press, 5th ed., 2009.

[6] M. Barron, "The subjective effects of first reflections in concert halls–the need for lateral reflections," *Journal of Sound and Vibration*, vol. 15, pp. 475–494, 1971.

[7] W. Burgtorf and H. Oehlschlägel, "Untersuchungen über die richtungsabhängige Wahrnehmbarkeit verzögerter Schallsignale (Studies on the

directional perceptibility of delayed sound signals)," *Acustica*, vol. 14, pp. 254–266, 1964.

[8] P. Schubert, "Die Wahrnehmbarkeit von Rückwürfen bei Musik (On the perception of reflections with music)," *Z. Hochfrequenztechn. u. Elektroakust.*, vol. 78, pp. 230–245, 1969.

[9] J. M. Buchholz, "Characterizing the monaural and binaural processes underlying reflection masking.," *Hearing research*, vol. 232, no. 1-2, pp. 52–66, 2007.

[10] W. Reichardt and W. Schmidt, "Die Wahrnehmbarkeit der Veranderung von Schallfeldparametern bei der Darbietung von Musik (The Detectability of Changes in Sound Field Parameters for Music)," *Acustica*, vol. 18, pp. 274–282, 1967.

[11] D. Kish, "Human echolocation: How to "see" like a bat.," *New Scientist*, vol. 202, no. 2703, pp. 31–33, 2009.

[12] 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, 2009.

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

[14] B. N. Schenkman and M. E. Nilsson, "Human echolocation: Blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object.," *Perception*, vol. 39, pp. 483–501, 2010.

[15] A. J. Oxenham, "Forward masking: Adaptation or integration?," *The Journal of the Acoustical Society of America*, vol. 109, pp. 732 – 741, 2001.

[16] B. C. J. Moore, "Frequency Selectivity, Masking and the Critical Band," in *An Introduction to the Psychology of Hearing*, pp. 65–126, Bingley, UK: Emerald, 2004.

[17] F. Bilsen and R. Ritsma, "Repetition pitch and its implication for hearing theory," *Acustica*, vol. 22, pp. 63–73, 1969.

[18] J. Blauert, *Spatial Hearing*. Cambridge, MA: The MIT Press, revised ed ed., 1997.

[19] J. S. Bradley, "Review of objective room acoustics measures and future needs," *Applied Acoustics*, vol. 72, no. 10, pp. 713–720, 2011.

[20] D. Rowan, T. Papadopoulos, D. Edwards, and R. Allen, "Use of binaural and monaural cues to identify the lateral position of a virtual object using echoes," *Hearing Research*, vol. 323, pp. 32–39, 2015.

[21] R. Y. Litovsky, H. S. Colburn, W. a. Yost, and S. J. Guzman, "The precedence effect," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.

[22] J. M. Buchholz, "A quantitative analysis of spectral mechanisms involved in auditory detection of coloration by a single wall reflection.," *Hearing research*, vol. 277, pp. 192–203, 2011.

[23] Y. Ando and H. Alrutz, "Perception of coloration in sound fields in relation to the autocorrelation function," *The Journal of the Acoustical Society of America*, vol. 71, pp. 616–618, 1982.

[24] A. M. Salomons, *Coloration and binaural decoloration of sound due to reflections*. Phd dissertation, TU Delft, 1995.

[25] A. J. M. Houtsma, "Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching," *The Journal of the Acoustical Society of America*, vol. 68, pp. 807–813, 1980.

[26] C. E. Rice, S. H. Feinstein, and R. J. Schusterman, "Echo-Detection Ability of the Blind: Size and Distance Factors.," *Journal of experimental psychology*, vol. 70, no. 3, pp. 246–255, 1965.

[27] D. Pelegrín-García, M. Rychtáriková, and C. Glorieux, "Audibility Thresholds of a Sound Reflection in a Classical Human Echolocation Experiment," *Acta Acustica united with Acustica*, vol. 102, no. 3, pp. 530–539, 2016.

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

[29] 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, no. 5, pp. 673–682, 2012.

[30] L. Wallmeier, N. Gessele, and L. Wiegrebe, "Echolocation versus echo suppression in humans," *Proceedings of the Royal Society of Biology*, vol. 280, p. 20131428, 2013.

[31] S. Schörnich, L. Wallmeier, N. Gessele, A. Nagy, M. Schranner, D. Kish, and L. Wiegrebe, "Psychophysics of Human Echolocation," in *Basic Aspects of Hearing* (B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, and H. E. Gockel, eds.), vol. 787 of *Advances in Experimental Medicine and Biology*, pp. 311–319, New York, NY: Springer New York, 2013.

[32] D. Cabrera, H. Sato, W. Martens, and D. Lee, "Binaural measurement and simulation of the room acoustical response from a person's mouth

to their ears," *Acoustics Australia*, vol. 37, no. 3, pp. 98–103, 2009.

[33] T. Wickens, "Forced-choice procedures," in *Elementary Signal Detection Theory*, pp. 93–110, Oxford: Oxford University Press, 2002.

[34] H. Levitt, "Transformed up-down methods in psychoacoustics.," *The Journal of the Acoustical Society of America*, vol. 49, pp. 467–477, 1971.

[35] Advanced CAE Research, "FastBEM Acoustics 3.0 User Guide," tech. rep., Cincinnati, Ohio, US, 2012.

[36] R. R. Paulsen, J. A. Baerentzen, and R. Larsen, "Markov random field surface reconstruction.," *IEEE transactions on visualization and computer graphics*, vol. 16, pp. 636–646, 2010.

[37] I. R. Cushing, F. F. Li, T. J. Cox, K. Worrall, and T. Jackson, "Vocal effort levels in anechoic conditions," *Applied Acoustics*, vol. 72, no. 9, pp. 695–701, 2011.

[38] F. Bilsen and R. Ritsma, "Some parameters influencing the perceptibility of pitch," *The Journal of the Acoustical Society of America*, vol. 47, pp. 469–475, 1970.

[39] M. Hansen, "PsyLab Version 2.5," tech. rep., Oldenburg, Germany, 2013.

[40] 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.

[41] C. Pörschmann, "One's own voice in auditory virtual environments," *Acta Acustica united with Acustica*, vol. 87, pp. 378–388, 2001.

[42] M. Wright, "Open Sound Control: an enabling technology for musical networking," *Organised Sound*, vol. 10, p. 193, 2005.

[43] H. Fastl and E. Zwicker, "Masking," in *Psychoacoustics. Facts and models*, pp. 61–110, Berlin Heidelberg: Springer-Verlag, 3rd ed., 2007.

[44] B. N. Schenkman and M. E. Nilsson, "Human echolocation: pitch versus loudness information.," *Perception*, vol. 40, no. 7, pp. 840–852, 2011.

[45] S. Reinfeldt, P. Ostli, B. Håkansson, S. Stenfelt, and P. Östli, "Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction," *The Journal of the Acoustical Society of America*, vol. 128, pp. 751–762, 2010.

[46] S. A. Gelfand, *Hearing : an introduction to psychological and physiological acoustics*. London : Informa Healthcare, 5th ed ed., 2010.

[47] J. A. M. Rojas, J. A. Hermosilla, R. S. Montero, and P. L. L. Espí, "Physical Analysis of Several Organic Signals for Human Echolocation: Hand and Finger Produced Pulses," *Acta Acustica united with Acustica*, vol. 96, pp. 1069–1077, 2010.

[48] O. A. Ramos and C. Arias, "Human echolocation: the ECOTEST System," *Applied Acoustics*, vol. 51, no. 4, pp. 439–445, 1997.

[49] A. Tonelli, L. Brayda, and M. Gori, "Depth echolocation learnt by novice sighted people," *PLoS ONE*, vol. 11, pp. 1–14, 06 2016.

[50] S. Teng and D. Whitney, "The Acuity of Echolocation: Spatial Resolution in Sighted Persons Compared to the Performance of an Expert Who Is Blind," *Journal of Visual Impairment & Blindness*, vol. 105, no. 1, pp. 20–32, 2011.

[51] J. J. Hant and A. Alwan, "A psychoacoustic-masking model to predict the perception of speech-like stimuli in noise," *Speech Communication*, vol. 40, pp. 291–313, 2003.