Detection, thresholds of human echolocation in static situations for distance, pitch, loudness and sharpness

Bo N. Schenkmana,1,⇑, Vijay Kiran Gidlab,1

a Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Sweden
b Blekinge Institute of Technology, Karlskrona, Sweden

Article info
Article history:
Received 10 April 2019
Received in revised form 10 October 2019
Accepted 12 January 2020

Keywords:
Blind
Echolocation
Psycho-acoustics
Auditory models
Detection thresholds

Abstract
Human echolocation describes how people, often blind, use reflected sounds to obtain information about their ambient world. Using auditory models for three perceptual variables, loudness, pitch and one aspect of timbre, namely sharpness, we determined how these variables can make people detect objects by echolocation. We used acoustic recordings and the resulting perceptual data from a previous study with stationary situations, as input to our analysis. One part of the analysis was on the physical room acoustics of the sounds, i.e. sound pressure level, autocorrelation and spectral centroid. In a second part we used auditory models to analyze echolocation resulting from the perceptual variables loudness, pitch and sharpness. Based on these results, a third part was the calculation of psychophysical thresholds with a non-parametric method for detecting a reflecting object with constant physical size for distance, loudness, pitch and sharpness. Difference thresholds were calculated for the psychophysical variables, since a 2-Alternative-Forced-Choice Paradigm had originally been used. We determined (1) detection thresholds based on repetition pitch, loudness and sharpness varied and their dependency on room acoustics and type of sound stimuli. We found (2) that repetition pitch was useful for detection at shorter distances and was determined from the peaks in the temporal profile of the autocorrelation function, (3) loudness at shorter distances provides echolocation information, (4) at longer distances, timbre aspects, such as sharpness, might be used to detect objects. (5) It is suggested that blind persons may detect objects at lower values for loudness, pitch strength and sharpness and at further distances than sighted persons. We also discuss the auditory model approach. Autocorrelation was assumed as a proper measure for pitch, but we ask whether a mechanism based on strobe integration is a viable possibility.

1. Introduction

Persons with blindness use echolocation to obtain information about their surroundings. A person or a source in the environment emits a sound and the reflection is perceived. Both static and dynamic information are used for this sensory information, i.e. when there is no movement involved respectively when the person, the reflecting object or both are moving. In both cases the person has to perceive if an object or obstacle is in front of him/her. This perceptual decision is determined by a threshold of detection. The threshold may vary because of a number of variables, like the type of emitted sound, e.g. clicks or hisses, the number of repetitions of the sound, the position of the sound source relative to the person, and as mentioned if motion is involved as well as the experience and expertise of echolocation. For a review of human echolocation, see [1–3]. Physical properties may have different effects on psychoacoustic parameters that are used to determine if an object is in front or not. Three psychoacoustic parameters are particularly important as sources for human echolocation, namely pitch in the form of repetition pitch, loudness, and spectral information that is perceived as timbre. Repetition pitch is the perception of a pitch that arises when a sound is repeated with itself after a short interval. We describe how the information provided by pitch, loudness or timbre may result in their respective detection thresholds for echolocation. We limit ourselves to stationary situations, i.e. when neither object nor person is moving. When movement is involved, more potential information may be provided [4,5]. We also determine the threshold for the distance where a person may detect a reflecting object. A number of auditory models were applied to the physical stimuli and we related the results of these models to the perceptual responses of

https://doi.org/10.1016/j.apacoust.2020.107214
0003-682X/© 2020 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Psychoacoustic and neuroimaging methods are very useful for describing the high echolocating ability of the blind and their underlying processes. However, they do not fully reveal the information in the acoustic stimulus that determines echolocation (at least when the source for the information is not known) and how this information is encoded in the auditory system. We wanted to know how this information is represented and processed in the human auditory system. One fruitful way to study the information necessary for human echolocation is by signal analysis on the acoustic physical stimulus, using auditory models which mimic human hearing. Analyzing the acoustic stimulus using these models provide insight into processes for human echolocation. They may also allow testing of hypotheses by comparing models [7].

Louderness, pitch and timbre are three perceptual attributes of an acoustic sound that are relevant for human echolocation. Almost all sounds with some complexity have a pitch, loudness and timbre [8]. The first two characteristics are uni-dimensional, whereas timbre consists of many dimensions.

Louderness is the perceptual attribute of sound intensity and is that attribute of auditory sensation in terms of which sounds can be ordered on a scale from quiet to loud (ASA 1973 [9]). Several models [8] have been proposed to calculate the average loudness that would be perceived by listeners. The excitation pattern is transformed into specific loudness, which involves a compressive non-linearity. The total area calculated for the specific loudness pattern is assumed to be proportional to the overall loudness.

Pitch is “that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale” (ASA 1960 [10]). According to the paradox of the missing fundamental, the pitch evoked by a pure tone remains the same if we add additional tones with frequencies that are integer multiples of that of the original pure tone, i.e. harmonics. It also does not change if we then remove the original pure tone, the fundamental [11]. Psychoacoustic studies also show that pitch exists for sounds which are not periodic. Human echolocation, is an instance of such sounds, namely iterated ripple noise. This sound models some of the human echolocation signals, e.g. [12]. In order to overcome the limitations of the place and time hypothesis two new theories have been proposed, pattern matching [11,13,14], and a theory based on autocorrelation [11,15]. The autocorrelation hypothesis assumes temporal processing in the auditory system. It states that, instead of detecting the peaks at regular intervals, the periodic neural pattern is processed by coincidence detector neurons that calculate the equivalent of an autocorrelation function [11,15]. An alternative to this theory based on autocorrelation is the strobe temporal integration (STI) of Patterson, Allerhand, and Giguere [16]. In accordance with STI the auditory image underlying the perception of pitch is obtained by using triggered, quantized, temporal integration, instead of autocorrelation. The STI works by finding the strobes from the neural activity pattern and integrating them over a certain period, see also Appendix A: “Pitch strength calculations”.

Timbre has been defined as that attribute of an auditory sensation which enables a listener to judge that two non-identical sounds, similarly presented and having the same loudness and pitch, are dissimilar (ANSI 1994 [17]). In our study, we focus on one aspect of timbre, namely sharpness.

These are the three primary sensations associated with a musical tone [18]. Already pioneers of human echolocation, Cotzin and Dallenbach [19] discussed their respective roles for echolocation. Schenkman and Nilsson [20] showed that both pitch and loudness are relevant for humans when echolocating, but that pitch is more important. Roederer [18, p. 156] argued that timbre perception is the first stage in how we perceive tone source recognition. In music this corresponds to how we identify an instrument. For echolocation, this would correspond to how we recognize a reflecting object, but to some extent also its detection. Pitch perception is more directed to object detection. Timbre is not only relevant for human, but also for animal echolocation. Wiegrebe [21] used one dimension of timbre for describing bat echolocation. Timbre is less well understood than pitch or loudness, but it has a role in how both humans and animals echolocate.

A special important case of how we perceive pitch is repetition pitch. Human echolocation signals consist of an original sound along with a reflected or delayed signal. Several studies have been presented to explain the pitch perception of such sounds. Bassett and Eastmond [22] examined the physical variations in the sound field close to a reflecting wall. They reported a perceived pitch caused by the interference of direct and reflected sounds at different distances from the wall; the pitch value being equal to the inverse of the delay. In a similar way, Small and McClellan [23] and Bilsen [24], delayed identical pulses and found that the pitch perceived was equal to the inverse of the delay, naming it time separation pitch, and repetition pitch, respectively. When a sound and the repetition of that sound are listened to, a subjective tone is perceived with a pitch corresponding to the reciprocal value of the delay time [25]. Bilsen and Ritsma [25] explained the repetition pitch phenomenon by the autocorrelation peaks or the spectral peaks. Yost [26] performed experiments using iterated ripple noise stimuli and concluded that autocorrelation was the underlying mechanism that had been used by the listeners to detect repetition pitch.

In this report we have analyzed synthetic sounds that were used to elicit echoes from reflecting objects. In a real-life situation, a blind person would use sounds produced by the person himself/herself [27,28].

Human echolocation may be directed at detection, localization or identification of an object. Our report deals with the detection of objects. For localization of objects, i.e. where in the room an object is located, the precedence effect will be important. Relating to human echolocation, this has been studied by Pelegrin-Garcia et al [29] and by Wallmeier et al [30].

1.1. Aims and hypothesis

The aims of the present study were:

(1) To study different components of the information in the acoustic stimulus, that determines echolocation.
(2) To determine the thresholds for different components of the information in the acoustic stimulus, that are important factors for the detection distance to reflecting objects.
(3) To find out how the acoustic information that determines high echolocation ability is represented in the human auditory system.

More specifically, our hypotheses were:

(1) Detection thresholds for echolocation based on repetition pitch, loudness and one dimension of timbre, sharpness, depend on the room acoustics and type of sound that is used, providing a possibility to detect a reflecting object.
(2) Repetition pitch is more prominent and useful for detection at shorter distances (<200 cm) and is determined from the peaks in the temporal profile of the autocorrelation function, computed from the neural activity pattern.
(3) Detection at shorter distances (<200 cm), based on loudness, provides additional information for listeners, since at longer distances than 200 cm the loudness differences between object and non-object will be <1 dB.
(4) At longer distances (greater than about 300 cm) where pitch and loudness information is absent, timbre aspects such as sharpness information may be used by some listeners to detect objects. Pitch and loudness has been found for short distances (e.g. [20]). Buchholz [31] studied pitch or coloration for early reflections while Pelegrin-Garcia, Rychtáříková and Giorieux [29] found detection at longer distances. This indicates that other attributes than pitch or loudness are used at longer distances.

In previous studies as by Schenkmam and Nilsson [6] and Schenkmam, Nilsson and Grbic [32], no thresholds were calculated for pitch, loudness or any aspect of timbre. We calculated the thresholds in [6], using auditory models. Additionally, when fitting the data we used a local non-parametric method that makes no assumptions of an underlying function, which we see as an advantage, while in the original study, [6], parametric methods were used.

2. Method

Sound travels and undergoes transformations because of the room acoustics. The first step is to understand the information that is received at the human ear in a room. Signal analysis is useful for this purpose, as we then can analyze the characteristics of the sound, which have been transformed due to the room conditions. Such an analysis on human echolocation was conducted by Papadopulos et al. [33]. They performed a physical analysis of binaural cues and concluded that these cues depended on the lateral position, orientation and distance to the echo-generating object. The second step is to analyze how the characteristics of the sound, that contains certain information, are represented in the auditory system. Here auditory models are useful, since the information of interest is transformed in an analogous way to how the auditory system is known to process it. Keeping track of the information from the outer ear to the central nervous system is an important part in describing how listeners perceive sounds and explaining the differences between listeners with different characteristics, e.g. visually handicapped vs sighted persons. This is the methodology which we used for this report. The overview of our methodology is illustrated in Fig. 1. A more detailed description of the methodology is presented below.

2.1. Sound recordings used

In this section we describe briefly how the sound recordings of SN2010 [6] were made. For more detailed descriptions, see the original article. In [6], the binaural sound recordings were conducted in an ordinary conference room and in an anechoic room using an artificial manikin. The object was a reflecting 1.5 mm thick aluminum disk with a diameter of 0.5 m. Recordings were conducted at distances of 0.5, 1, 2, 3, 4, and 5 m between microphones and the reflecting object. In addition, recordings were made with no obstacle in front of the artificial manikin. Durations of the noise signal were 500, 50, and 5 ms; the shortest corresponds perceptually to a click. The electrical signal was a white noise. However, the emitted sound was not perfectly white, because of the non-linear frequency response of the loudspeaker and the system. A loudspeaker generated the sounds, resting on the chest of the artificial manikin. The reverberation time, T60, for the conference room was 0.4 s.

We calculated SPL both as dB and as dBA for the calibration of signal, but the differences were negligible, so only dB was used. The recordings were made with level calibration.

To give a visual understanding of the character of the sounds, spectrogram analysis of the recordings of study [6] in the different rooms for the 5 and 500 ms sounds are presented in Figs. 2 and 3, respectively. The frequency axis is in logarithmic scales to better resemble the auditory spacing in the ear, that will give an improved knowledge on the relevant auditory frequency content of the signals used. Buchholz [31] showed that the frequency range of signals, in particular at low frequencies and longer delays/distances is important for the detection of reflections.

2.2. Room acoustics

One aim of this study was to study information used for human echolocation. The room acoustics will have an effect on how a person perceives the acoustic information in it. We therefore analyzed how room acoustics affects three physical attributes that are useful for human echolocation. The respective results for Sound Pressure Level (SPL), Autocorrelation Function (ACF) and Spectral Centroid (SC) are presented separately in section 4, “Results”.

3. Models

The auditory models we used for pitch, loudness and sharpness, and their individual blocks are shown in Fig. 4. The distance threshold was calculated without an auditory model. We describe below the most important aspects of the models and where we differ from the original implementation. The reference to the original articles are presented in respective sub sections.

3.1. Loudness: binaural loudness, short and long term loudness

The sound pressure level analysis describes the physical intensity of a sound that may affect human echolocation. The perception of loudness is a psychological attribute that depends on intensity, but also on a number of other parameters, like frequency selectivity, bandwidth and duration of the sound. We here present perceptual aspects of loudness for echolocating sounds, using the loudness model of Glasberg and Moore [34]. This loudness model computes the frequency selectivity and compression of the basilar membrane in two stages, (1) by computing the excitation pattern and (2) by the specific loudness of the input signal. Physiologically they are interlinked, and a time domain filter bank which simulates both the selectivity and the compression might be appropriate. We chose this model instead of the auditory image model (AIM) developed by Patterson et al. [16,35] because of the better prediction of the equal loudness contours in ISO 2006 [36]. This model estimates the loudness of steady sounds and of time varying sounds, by accounting for frequency selectivity and compression of the human auditory system. A detailed description of each stage of the loudness models can be found in articles by Glasberg and Moore [34,37]. The loudness model was implemented in PsychoSound3 by Cabrera, Ferguson, and Schubert [38], being a GUI driven Matlab environment for analysis of audio recordings, which can be downloaded from http://www.psysound.org.

Short Term Loudness (STL). The short-term loudness is a lowpass version of the instantaneous loudness. It is determined by averaging the instantaneous loudness using an attack constant, $\alpha = 0.045$, and a decay constant, $\beta = 0.02$ (Eq. (1)). The values of $\alpha$ and $\beta$ were chosen so that the model will give reasonable predictions for variations of loudness with duration and amplitude modulated sounds [36].

$$\text{STL}(n) = \begin{cases} \alpha + \beta \text{STL}_{n-1} & \text{if } \text{IL}(n) \geq \text{STL}(n-1) \\ \alpha \text{STL}_{n-1} + (1 - \alpha) \text{STL}_{n-1} & \text{if } \text{IL}(n) < \text{STL}(n-1) \end{cases}$$

(1)
Fig. 1. Overview of the methodology used for calculating acoustic and psycho-acoustic values. The acoustic recordings and the perceptual measures were used as inputs for the signal and auditory model analysis. The perceptual attributes loudness, pitch and sharpness were quantified and related to the participants’ perceptual measures using a non-parametric psychometric function.

Fig. 2. Spectrograms of the 5 ms sound recordings. Anechoic room with no object and with object at 50 and 200 cm (A, B and C), conference room with no object and with object at 50 and 200 cm (D, E and F), in [6].
The long-term loudness is a low-pass version of the short-term loudness. Its parameter was calculated by averaging the instantaneous loudness using an attack constant, $a_1 = 0.01$ and a decay constant, $r_1 = 0.0005$ (Eq. (2)). The values of $a_1$ and $r_1$ were chosen so that the model should give reasonable predictions for the overall loudness of sounds that are amplitude modulated at low rates [36].

\[
LTL(n) = \begin{cases} 
\alpha_{a1} \times STL(n) + (1 - \alpha_{a1}) \times LTL_{n-1} & \text{if } STL(n) \geq LTL(n-1) \\
\alpha_{r1} \times STL(n) + (1 - \alpha_{r1}) \times LTL_{n-1} & \text{if } STL(n) \leq LTL(n-1)
\end{cases}
\]  

(2)

As noted above, loudness is affected by binaural hearing. To model binaural loudness, a number of psychoacoustic facts have been considered (for details see [36]). Early results suggested that...
the level difference required for equal loudness of monaurally and diotically presented sounds was 10 dB. The subjective loudness of a sound doubles with about every 10 dB increase in physical intensity, and therefore it was assumed in the early loudness model of Glasberg and Moore [34] that loudness sums across ears. However, later results suggested that the level difference required for equal loudness is rather between 5 and 6 dB. Glasberg and Moore therefore presented a new model to account for the lower dB values based on the concept of inhibition.

Glasberg and Moore [34] implemented inhibition for binaural hearing by a gain function. Initially, the specific loudness pattern was smoothed with a Gaussian weighting function and the relative values of the smoothed function at the two ears were used to compute the gain functions of the ears. The gains were then applied to the specific loudness patterns at the two ears. The loudness for each ear was calculated by summing the specific loudness over the center frequencies and the binaural loudness was obtained by summing the loudness values across the two ears [36]. This procedure was used to calculate the binaural loudness values for this report.

Moore and Glasberg [37] assumed that the loudness of a brief sound is determined by the maximum of the short term loudness, while the long term loudness may correspond to the memory for the loudness of an event that can last for several seconds. For a time varying sound (e.g., an amplitude modulated tone), it is appropriate to consider the long time loudness as a function of time to calculate the time varying loudness. However, in this report, as the stimuli presented to the participants were noise bursts and can be considered steady and brief, we follow the assumption of Glasberg and Moore [34] of using the maximum of short time loudness as a measure of the loudness of the recordings.

3.2. Pitch analysis: autocorrelation

The AIM model is a time-domain, functional model of the signal processing in the auditory pathway as the system converts a sound wave into the percept that we experience. This representation is referred to as an auditory image by analogy with the visual image of a scene that we experience in response to optical stimulation. The AIM simplifies the peripheral and the central auditory systems into modules. A more detailed description of each module of AIM can be found at [http://www.acousticscale.org/wiki/index.php/AIM2006_Documentation](http://www.acousticscale.org/wiki/index.php/AIM2006_Documentation). The auditory image model has been implemented in Matlab by Bleeck, Ives, and Patterson [39] and the current version is known as AIM-MAT. AIM-MAT was downloaded from [https://code.soundsoftware.ac.uk/projects/aimmat](https://code.soundsoftware.ac.uk/projects/aimmat). The autocorr module was only present in the 2003 version of AIM and can be downloaded from [http://w3.pdn.cam.ac.uk/grou ps/cnbh/aimmanual/download/downloadframeset.htm](http://w3.pdn.cam.ac.uk/groups/cnbh/aimmanual/download/downloadframeset.htm).

Perceptual research suggests that at least some of the fine grain time interval information is needed to preserve timbre information [40–42]. Auditory models often time average the NAP information, which unfortunately then loses the fine grain information. To prevent this, AIM uses a procedure called Strobe Temporal Integration (STI) as a final stage to represent processing in the central nervous system, which is subdivided into two modules, (i) strobe finding, and (ii) temporal integration.

**Strobe Finding (SF):** A sub module of AIM is used to find the strobes from the NAP. It uses an adaptive strobe threshold to issue a strobe and the time of the strobe is that associated with the peak of the NAP pulse. After the strobe is initiated the threshold initially rises along a parabolic path and then returns to the linear decay to avoid spurious strobes. The duration of the parabola is proportional to the centre frequency of the channel and the height to the height of the strobe. After the parabolic section of the adaptive threshold, its level decreases linearly to zero in 30 ms. An example of how the threshold varies and how the strobes are calculated is shown in Fig. 5.

**Temporal Integration (TI):** The temporal integration is implemented in AIM by a module called stabilized auditory image (SAI). Initially, a temporal integration is initiated when a strobe is detected. If strobes are detected within a 35 ms interval, each strobe initiates a temporal integration process. To preserve the shape of the SAI to that of the NAP, the new strobes are initially weighted high (the weights are also normalized so that the sum of the weights is equal to 1) making the older strobes contribute relatively less to the SAI.

However, we chose not to use the strobe temporal integration as the final stage, but it does not exclude that this might be how pitch information for echolocation is represented in the auditory system. To determine whether it is autocorrelation or strobe temporal integration that best explains repetition pitch perception and possibly also physiological processes involved in the auditory system, further experiments and analysis are needed. We present as an example results obtained from using the strobe temporal integration module for a 500 ms signal in Appendix A: “Pitch strength calculations”.

3.2.1. Autocorrelation function (ACF)

Corresponding physiological processes to autocorrelation are presumed to take place in the central nervous system [15,24,26]. By using the autocorr module of AIM one can implement models of hearing based on such processes. The autocorr module takes the NAP as input and computes the ACF on each center frequency channel of the NAP by using a duration of 70 ms, hop time of 10 ms and a maximum delay of 35 ms.

Repetition pitch is a percept that underlies human echolocation for detecting objects. It is usually experienced as a coloration of the sound, perceived at a frequency equal to the inverse of the delay time between the sound and its reflection [12,25]; see also [22]. In SN2010 the reflecting object was at distances of 50, 100, 200, 300, 400 and 500 cm resulting in delays of 2.9, 5.8, 11.6, 17.4, 23.2 and 29 ms, where the repetition pitches would correspond to 344, 172, 114, 86, 57, 43 and 34 Hz, respectively. However, the actual delays might vary because of factors like the recording set up, speed of sound etc. and therefore the actual repetition pitch would be different. To test the presence of repetition pitch at these
frequencies together with how this information would be represented in the auditory system, we used the PCP, BMM and NAP modules of the AIM, summarized presented above, to analyze the recordings from SN2010 [6].

Repetition pitch can be created by presenting iterated rippled noise stimuli. The peaks in the autocorrelation function of these sounds have been seen as the basis for the perception of repetition pitch [26,35]. Hence, instead of the strobe finding and the temporal integration modules in AIM, we used the autocorr module as the final stage in our analysis to quantify the information for repetition pitch. Analysis by autocorrelation provides a feasible way to quantify repetition pitch, which is needed for explaining and understanding human echolocation.

After generating the ACF with the autocorr module, the AIM has a dual profile development module, which sums up the ACF along both the temporal and the spectral domain. These features are relevant in depicting how temporal and spectral information might be represented and are useful for analyzing repetition pitch. From the recordings, followed by intermediate analysis and the autocorrelation output, ACF(f,t), the time averaged spectrum and frequency weighted over time was calculated, resulting in the relation output, ACF(f,t), the time averaged spectrum and frequency weighted spectrum, SP(Hz), for the temporal and spectral domain, as shown in Eqs. (3) and (4), respectively. These features are relevant for depicting how temporal and spectral information might be represented for analyzing repetition pitch. The results of this analysis are presented in “Results”, subsection 4.2.

\[ TP(t) = \sum_{f=50}^{8000} ACF(f,t) \]  
\[ SP(Hz) = \sum_{f=0}^{3200} ACF(f,t) \]

3.3. Timbre: Sharpness analysis

To address more specifically how human hearing perceives timbre, Fastl and Zwicker [43] computed the weighted centroid of the specific loudness rather than of the Fourier Transform. This psychological measure of timbre is called sharpness and indicates how sound is perceived to vary from dull to sharp. We describe in section 4.4 how the spectral centroid was used as a measure for timbre perception. The spectral centroid was computed on the time varying Fourier Transform. We conducted the sharpness analysis for our recordings using code available from Psysound [38], see Section 4.5. Sharpness varies over time and therefore the median was used as a representative measure for the perceived sharpness. The unit for sharpness is called acum.

There are, to our knowledge, only a few studies on thresholds of sharpness. Pedrielli, Carletti, and Casazza [44] found that their participants had a just noticeable difference for sharpness of 0.04 acum. You and Jeon [45] found in a study on refrigerator noise that their participants had a just noticeable difference for sharpness of 0.08 acum.

We present the values for the corresponding analysis of the perceptual measures, i.e. loudness, pitch and sharpness separately in Section 4, “Results”.

3.4. Threshold values, absolute and difference, based on auditory model analysis

For echolocation we define the absolute threshold as the value of a perceptual attribute (e.g. loudness, pitch or sharpness) in a situation, e.g. a recording with a reflecting object at which a person has a 75 percentage correct responses. The difference threshold is the difference of responses between situations, e.g. recordings with and without object, at which a person has 75 percentage of correct response [46]. The experimental procedure used in the experiments of [6] was 2AFC, and therefore the difference threshold is the relevant measure of the echolocation ability of the persons. The difference threshold was calculated from absolute thresholds, so both are for clarity shown in this article. The procedure for finding the difference thresholds and the corresponding results is presented below.

3.4.1. Non-parametric versus parametric modeling of psychometric function for threshold analysis

Psychometric functions relate perceptual results to physical parameters of a stimulus. Commonly the psychometric function is estimated by parametric fitting, i.e. it is assumed that the underlying relationship can be described by a specific parametric model. The parameters of such a model are then estimated by maximizing the likelihood. However, the most correct parametric model underlying the description of the psychometric function is unknown. Therefore, estimating the psychometric function based on the assumptions of a parametric model may lead to incorrect interpretations [47]. To address this problem, Zychaluk and Foster [47] implemented a non-parametric model to estimate the psychometric function. This psychometric function is modeled locally without assuming a “true” underlying function. Since the true relationship for the variables that determine human echolocation is unknown, we chose the method proposed by Zychaluk and Foster [47] in our analysis of the perceptual data.

A generalized linear model (GLM) is usually used when fitting a psychometric function. It consists of three components: a random component from the exponential family, a systematic component \( \eta \) and a monotonic differentiable link function \( g \), that relates the two. Hence, a psychometric function, \( P(x) \), can be modeled by using Eq. (5). The parameters of the GLM are estimated by maximizing the appropriate likelihood function [47]. The efficiency of the GLM relies on how much the chosen link function, \( g \), approximates the “true” underlying function.

\[ \eta(x) = g(P(x)) \]  

In non-parametric modelling, instead of fitting the link function \( g \), the function \( \eta \) is fitted using a local linear method. For a given point, \( x \), the value \( \eta(u) \) at any point \( u \) in a neighborhood of \( x \) is approximated by Eq. (6) [47].

\[ \eta(u) \approx \eta(u) - (u - x)\eta'(x) \]  
\[ \eta'(x) \]  

where \( \eta'(x) \) is the first derivative of \( \eta \). The actual estimate of the value of \( \eta(x) \) is obtained by fitting this approximation to the data over the prescribed neighborhood of \( x \). Two features are important for this purpose, kernel \( K \) and the bandwidth \( h \). A Gaussian kernel is preferred, as it has unbounded support and is best for widely spaced levels. An optimal bandwidth can be chosen using plugin, bootstrap or cross validation methods [47]. As no method is guaranteed to always work, to find the optimal bandwidth for the analysis we chose a bootstrap method with 30 replications. When the bootstrap method failed to find an optimal bandwidth, cross validation was used to establish the optimal bandwidth.

4. Results

4.1. Sound pressure level (SPL) and loudness analysis

As has been pointed out, e.g. by Rowan et al [48], binaural information may be utilized for echolocation purposes. We therefore calculated the SPL values for both ears. The mean SPL values of...
the 500 ms recordings in study SN2010 [6] are shown in Fig. 6. The values for the 5 ms and 50 ms recordings are for reasons of space not shown here. As mentioned above in section 2.1, for the recordings with no object, two series were conducted in SN2010 [6], each with 10 recordings. As can be seen in Fig. 6, the values are very close to each other.

The SPL values in Fig. 6 left subplot show the effect of room acoustics for level differences, between the ears and between the rooms. The extent to which this information affected the listeners is not obvious, as loudness perceived by the human auditory system cannot be related directly to the SPL, e.g. [8].

The means of the maxima values of Short Term Loudness in sone for the 10 versions for the 5, 50 and 500 ms recordings in the rooms of SN2010 [6] are presented in Fig. 6, right subplot. The loudness values follow the same pattern as the sound pressure level analysis (Fig. 6 left subplot). However, the values in Fig. 6, right subplot, are psychophysical and depict both room acoustics and aspects of human hearing that are important for human echolocation. Relating the psychophysical loudness results to the echolocation of the test persons will be made in section 4.6.2, “Loudness thresholds, absolute and difference, for object detection”.

4.2. Autocorrelation function (ACF) and pitch analysis

The theoretical values for repetition pitch for the recordings of [6] were calculated using Eq. (7). The corresponding values for recordings with objects at distances of 50, 100, 200, 300, 400 and 500 cm would be approximately 344, 172, 86, 57, 43 and 34.4 Hz, assuming sound velocity to be 344 m/s. As the theory based on autocorrelation uses temporal information, repetition pitch perceived at the above frequencies can be explained by the peaks in the ACF at the inverse of the frequencies, i.e. approximately at 2.9, 5.8, 11.6, 17.4, 23.2 and 29 ms, respectively. The autocorrelation analysis was performed using a 32 ms frame, which would cover the required pitch period. A 32 ms hop size was used to analyze the ACF for the next time instants of 64 ms, 96 ms etc. In order to compare the peaks among all the recordings, the ACF was not normalized to the limits $-1$ to $1$.

$$RP = \frac{\text{speed of sound}}{2 \times \text{distance of the object}}$$ (7)

where RP is Repetition Pitch.

It is not evident when longer or shorter duration signals are beneficial for human echolocators. For example, in study [6] the participants performed well with the longer duration signals. For a single short burst the person had only one chance to perceive the signal and its echo. As we consider repetition pitch to be one of the main information sources for human echolocation, the short-term frame analysis will give more insights on how repetition pitch is perceived for both shorter and longer duration signals. This can be visualized from the ACFs in Figs. 7 and 8, where the ACFs of a 5 and a 500 ms sound are shown for a reflecting object at a distance of 100 cm. The 50 ms sound and other distances are for reasons of space not shown here.

We used the module, originally called dual profile, to analyze the temporal and spectral results (cf. Eqs. (3) and (4)). However, the recordings with the object at 300 to 500 cm in study SN2010 [6] do not provide any additional information for the module and were therefore not included in this autocorrelation analysis.

When presenting the figures graphically, we used the first 70 ms time interval of the recordings. The time averaged spectrum and frequency weighted over time are presented for the 5 ms signal durations and the two rooms in SN2010 [6]. A similar analysis for the 50 and 500 ms signals would not add any new information, and hence are not presented. It is important to note that the amplitude scale of the y-axis is different in each subplot of Fig. 9. The investigated attribute is pitch and each subplot with reflecting
object should be compared with the subfigure with No object for each condition. A distinct peak in a subfigure which is absent in the subfigure with No object indicates the potential occurrence of the perception of a pitch. One should remember that the visual impression of a peak in a subfigure with No object may misleadingly indicate an auditory peak, unless one observes the different scales on the different y-axis for the different subfigures. The next section will deal with how to select peaks based on their peak strength.

As mentioned above, the theoretical frequencies of the repetition pitch for recordings with object at 100, 150 and 200 cm were 172, 114 and 86 Hz, respectively. The analysis of the 5 ms recordings show peaks approximately at these frequencies. For example, both the anechoic and the conference room in Fig. 9c had peaks (marked by arrows in the subfigures) approximately at 172 Hz; at 82 Hz in Fig. 9d. One reason that the peaks were not exactly at the theoretical values is probably due to the experimental setup of SN2016 [6] and of the room acoustics. Both the anechoic and conference room in Fig. 9b had no peaks at their corresponding theoretical frequencies, which is due to a wider range in the y-axis scale (cf. Fig. 9 y-axis labels). The spectral profiles on the other hand did not have peaks close to the theoretical frequencies (cf. Fig. 9 dashed lines). There were small spectral differences but these may provide timbre perception but not information for pitch. However, for concluding that the temporal profile (solid line) is a sufficient condition for detecting reflecting objects by repetition pitch, a further analysis is needed which quantifies the peaks in the temporal profile.

To determine the role of temporal information for detecting objects based on repetition pitch, the pitch strength development module of AIM was used. It measures the pitch perceived based on the peak strength. We elaborate this in the next section. The temporal profiles will be shown to have peaks at the theoretical frequencies of repetition pitch which, we believe, explains the perception of repetition pitch and thus also a major cause for detection by echolocation of the reflecting objects in the two studies.

4.3. Pitch strength

The peaks in the temporal profile of the autocorrelation function that we computed with the dual profile module of AIM were distributed without apparent order or meaning. To find a peak that corresponds to a pitch, the AIM model has a pitch strength module which calculates the pitch strength. This determines if a peak is random or not. This module first calculates the local maxima and their corresponding local minima. The ratio of peak height to the peak width of the peak (local maxima) is subtracted from the mean of the peak height between two adjacent local minima to obtain the pitch strength (PS) of a particular peak.

\[ \text{PS} = \frac{\text{PH}}{C_0} - \text{PHLM} \]

where PS is the calculated pitch strength, PH is the height of the peak and PHLM is the mean of the peak height between two adjacent local minima. In Appendix A: “Pitch strength calculations”, we give an example of how the pitch strength was calculated by Eq. (8).

The results of the calculated pitch strength for recordings of study SN2010 [6] are presented in Fig. 10. As can be seen, peaks were misleadingly identified for recordings without an object, which would not have caused a pitch perception. This happens because the pitch strength algorithm identifies all local maxima and minima in a sound and thus also calculates the pitch strength for random peaks (that have local maxima). On the other hand, the pitch strengths of these are very small.

Fig. 10 shows that for the 5 ms and 50 ms duration signals the pitch strength was greater than 1 for the object distances of 50 and
Fig. 9. The time averaged spectrum and frequency weighted over time of a 5 ms signal recorded in the anechoic and the conference room in SN2010 [6]. The solid line is the sum of the ACF along the spectral axis and the dashed line is the sum of the ACF along the time delay axis for 70 ms time interval. The temporal and spectral profiles are scaled to be compared to each other. The ‘x’ axis is changed in the temporal profile by using the inverse relationship of time to frequency, f = 1/t. The subfigures A are for the No object recording and subfigures B, C and D are for the recordings with object at 50, 100 and 200 cm, respectively, for the anechoic and the conference room, respectively.

Fig. 10. Pitch strength (autocorrelation index) for the 10 versions of the recordings in SN2010 [6] with the 5, 50 and 500 ms duration signal, means and standard errors. The pitch strength values are shifted horizontally for clarity. There were two series of recordings with no reflecting object, making two zero points for distance.
100 cm in the anechoic and conference rooms of SN2010 [6]. For the 500 ms duration signal, the strength was greater than 1 at distances of 50 and 100 cm in the anechoic room and at the distances of 50, 100, and 200 cm in the conference room. (The time frames had 35 ms time delay computed from a 70 ms interval NAP signal. Each frame had a hop time of 10 ms.) In SN2010 [6] both rooms had high pitch strengths at a particular frequency that lasted for 14 to 18 time frames of 35 ms intervals, each with a hop time of 10 ms.

The perceptual results of SN2010 [6] showed that the participants were able to detect the objects with a high percentage correct at object distances 50 and 100 cm in the anechoic room and at 50, 100, and 200 cm in the conference room [6]. As presented in the previous paragraph, the pitch strength was greater than 1 at these conditions. Pitch seems to be the most important information that listeners use to detect objects at these distances, see e.g. [20]. Therefore, these results imply that there might be a perceptual threshold approximately equal to 1 (autocorrelation index) for pitch strength in echolocating situations. The peak with such pitch strength must exist for certain time frames for a person to perceive a repetition pitch. This also depends on the room acoustics. Relating the pitch strength results to the performance of the participants in study [6] will be made in section 4.6 on threshold values. Before this we shall analyze the results in terms of a timbre property, namely sharpness, for which the spectral centroid is calculated.

4.4. Spectral centroid (SC)

To compute the spectral centroid, the recordings were analyzed using a 32 ms frame with a 2 ms overlap. The spectral centroid for each frame was computed by Eq. (9). As the spectral centroid for each frame is a time varying function, it is plotted as a function of time. The means of the spectral centroid for the 10 versions at each condition for the 500 ms of the left ear recordings are shown in Fig. 11.

\[
\text{Spectral Centroid} = \frac{\sum (\text{Frequency} \times \text{FFT(frame)})}{\sum (\text{FFT(frame)})}
\]  

(9)

The subfigures in Fig. 11 show that the spectral centroid varies over time and for different recording conditions. These results are based on a purely physical analysis, a Fast Fourier Transform (FFT) analysis of the sounds. However, the spectral analysis performed by the auditory system is more complex than a FFT that was used to compute the spectral centroid. In the next section we consider the role of human hearing by using auditory models to analyze the sharpness attribute of the sounds. In order to compare the results quantitatively, the median of the spectral centroid over time was calculated for each recording condition, see Fig. 12.

4.5. Sharpness analysis

Assuming that 0.04 acum is a threshold value for sharpness, then Fig. 12 shows that the difference in median sharpness in SN2010 [6] was greater than threshold for the object at 0.5 and 1 m when compared to the recordings without the object, cf. the differences in sharpness values in Fig. 12 at 0, 0.5 and 1 m in the anechoic and in the conference room. It is however possible that at shorter distances (say<2 m) repetition pitch and loudness information might be more relevant for providing echolocation information than sharpness information.

In study SN2010 [6] with reflecting object at distances 2, 3, 4 and 5 m for 5 ms (anechoic and conference rooms), 50 ms (anechoic and conference rooms) and 500 ms signal (conference room) durations, the recordings had differences in median sharpness of 0.04 acum when compared to the recordings without an object. On the other hand, in the anechoic room of SN2010 for the 500 ms signal duration [6], the recordings with object at 4 m and 5 m had differences in sharpness approximately greater than 0.04 acum when compared to the recordings without the object (Fig. 12). This is perceptual information that blind persons might use to detect and identify a reflecting object at longer distances than, say at 2 m. We discuss this issue further in the next section.

4.6. Threshold values

4.6.1. Distance thresholds for object detection

We fitted the psychometric function to the mean proportion of correct responses dependent on the distance to the reflecting object. Fig. 13 shows both the non-parametric modeling (local linear fit) and the parametric modeling of the perceptual results for the blind test persons in study SN2010 [6]. To show the advantage of the non-parametric approach, as an example we plot the mean percentage correct as a function of distance for recordings with the 500 ms signals in the anechoic and conference rooms. The link function used for the parametric modeling was the Weibull function. Visual inspection shows that this link function was not appropriate, since the fit does not correspond well with the perceptual results. As mentioned before, if one knows the underlying link function for the psychophysical data, then the parametric fit is a better fit than the local non-linear fit, but for the data we are analyzing we do not know and cannot assume a particular link function that the local linear fit correlates well with the perceptual results. This demonstrates some of the advantages of using non-parametric modeling.

The means of the proportion of correct responses of the test persons were used for the psychometric fitting. If the individual responses had been used the individual thresholds would vary, but the local linear fit would probably still be well correlated with the perceptual results. Therefore, the results in the remaining part of the present section will be based on the psychometric function using local linear fit for the mean proportion correct answers. We used the implementation of the non-parametric model fitting in Matlab by Zychaluk and Foster [47].

Distance perception of an object is not a perceptual attribute that was presented directly to the participants in study SN2010 [6]. Therefore, the distance threshold obtained from the psychometric fit is a derived quantitative threshold. The distance threshold is the distance at which a person may detect an object with a probability of 75%. As the fitted psychometric function is discrete, it was not always possible for the fit to have an exact value of 0.75. Therefore, the threshold values at the proportion of correct responses within the range of 0.73 to 0.76 were calculated and the mean of the threshold values was determined as the actual threshold.

Our reanalysis of the data showed a higher sensitivity of both blind and sighted test persons than the values in study [6]. The distance threshold in [6] as calculated by us for which the blind and the sighted could detect the object using echolocation, with a proportion of correct responses between 0.73 and 0.76, are shown in Table 1. The distances at which the blind participants could detect the reflecting object were farther away than for the sighted in both rooms and for all three sound signals. The threshold is positively related to the signal duration for both groups, i.e. the longer durations give a longer range of detection. One can also see that the blind persons could detect objects farther away in the conference room for all signals, but for the sighted this was only the case for the 500 ms signal. In the original study SN2010 [6, Table 3], the calculations of the distance where a blind or a sighted person might detect a reflecting object, based on a parametric approach, yielded in general lower distance values for thresholds, i.e. lower sensitivity, than our non-parametric approach.
A 3-way mixed analysis of variance was conducted on the distance thresholds for the data from SN2010 for all the test persons. The two groups, blind and sighted, was the between variable, while the signal duration and rooms were the within variables. The analysis was calculated on the mean of the distance thresholds corresponding to proportion correct of 0.73 to 0.76, since the curve fitting is based on discrete values, while excluding one sighted test person who did not reach this threshold. We therefore used an unbalanced variance for the analysis of variance, where the sum of squares was calculated as type 3. The same considerations for the analysis and for the exclusion of this test person were made in the subsequent analysis of loudness, pitch and timbre. Mauchly’s test of sphericity indicated that the assumption of sphericity had been violated for the effects of duration and for the interaction of duration with test groups. For these effects, the Greenhouse-Geisser correction was applied.

The analysis showed that there was a significant difference between the difference thresholds for the main effect of sound durations (F(1.32, 22.44) = 26.22, p < 0.05, Geisser-Greenhouse epsilon = 0.66), for the two-way interaction for duration with test...
groups (F(1.32, 22.44) = 11.90, p < 0.001, Geisser-Greenhouse epsilon = 0.66). The main effect of rooms (F(1,17) = 10.41, p < 0.05) was significant, as well as the two-way interaction between rooms and test groups (F(1, 17) = 5.50, p < 0.05).

4.6.2. Loudness thresholds, absolute and difference, for object detection

Loudness is one of the information sources for detecting reflecting objects by echolocation [2,20]. A common psycho-acoustical measure to express loudness is in the unit of sone [46,49]. Therefore, we used sone values for the local linear fitting to determine the absolute and difference threshold of loudness, where blind and sighted could detect a reflecting object. As elsewhere, the criterion was detection with percentage correct between 73 and 76%. The mean loudness values, and the mean percentage of correct responses calculated from study SN2010 [6] were used as inputs for the psychometric fit. The resulting absolute threshold values of loudness for detecting the object are presented in Table 2. However, the absolute threshold values were only used for calculation of the difference threshold values, and conclusions on echolocation based on an analysis of these values would be misleading. Therefore the statistical analysis was performed on the difference thresholds. The reasons are as follows.

The values in Table 2 may misleadingly lead a reader to infer that the shortest sounds had the lowest threshold. This is not the case. If we look at Fig. 6, the recording without object for the 5 ms signal had a loudness of approximately 13 sone and the 500 ms signal had approximately 45 sone. Considering these values, it was more appropriate to use the difference rather than the absolute thresholds, since the detection based on loudness in SN2010 [6], as a consequence of the 2AFC method that had been used, was based on a relative judgment, a comparison, and not on an absolute judgment. The difference threshold values were calculated by subtracting the absolute threshold values, shown in Table 2 with their corresponding loudness values for the recording without object, also shown in Table 2.

A 3-way mixed analysis of variance was conducted on the difference thresholds for loudness for the data from SN2010. In SN2010 the analysis had been on the percentage correct of the test persons. As mentioned, an analysis of variance for the absolute thresholds is not presented here, since these values were only used when calculating the difference thresholds. The two groups, blind and sighted, was the between variable, while the sound duration and rooms were the within variables. Since these thresholds were computed for each individual by using the distance variable, this variable is not present in these analyses. The analysis was calculated on the mean of the difference thresholds corresponding to proportion correct responses of 0.73 to 0.76. Since, as noted before, one of the sighted test persons did not reach this criterion of proportion we used an unbalanced variance for the analysis of variance, where the sum of squares were calculated as type 3. The same considerations for the analysis hold for the subsequent analysis of pitch and timbre. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated for the main effect of duration, for which the Greenhouse-Geisser correction was applied.

The analysis showed that there was a significant difference between the difference thresholds for sound durations (F(1.43, 24.39) = 5.60, p < 0.05, Geisser-Greenhouse epsilon = 0.72), and for the two-way interaction of duration with room (F(2, 34) = 11.34). There was also a two-way interaction between rooms and test groups (F(1, 17) = 5.74, p < 0.05). The mean loudness difference thresholds for the blind and the sighted were 5.17 and 7.20 sone, respectively, thus indicating a higher sensitivity of the blind for loudness differences.

The data presented in Table 2 and the analysis of variance may indicate that the blind could detect objects at lower loudness values in both rooms, and that both groups could detect with lower

Table 1

| Sound duration (ms) | Room     | Blind | Sighted | Blind | Sighted | Blind | Sighted |
|---------------------|----------|-------|---------|-------|---------|-------|---------|
| 5                   | Anechoic | 150   | 130     | 166   | 160     | 172   | 166     |
|                     | Conference | 158   | 121     | 176   | 147     | 247   | 207     |

Table 2

Mean proportion of correct responses for the 500 ms recordings of the blind participants in SN2010 as a function of distance: (A) anechoic room, (B) conference room.

Fig. 13.
relative loudness levels (loudness difference between the recording with and without object) for the 500 ms duration signals in the conference room. The loudness model used to compute the mean loudness was the same for both test groups, and therefore the apparent lower thresholds of the blind persons are an effect of their perceptual ability. This conclusion is further addressed in section 5, “Discussion”. When analyzing the difference threshold values, one should remember that these values could be similar for both sighted and blind persons, while the absolute values could vary. One could thus make the erroneous conclusion that the sensitivities of both groups are the same. These considerations are a consequence of the 2AFC method that had been used in the original tests. They also hold true for when discussing the pitch and sharpness differential thresholds below.

4.6.3. Pitch thresholds, absolute and difference, for object detection

The absolute and difference threshold values of pitch strength, as calculated by the autocorrelation index for which the blind and the sighted test persons in study SN2010 [6] could detect the reflecting object, are presented in Table 3. We will first discuss the results in terms of the absolute thresholds, but the more appropriate conclusions will be based on the difference thresholds.

The absolute threshold varies for blind and sighted persons depending on signal durations and room conditions. The blind, for all conditions, had lower thresholds, the pitch strength increased with signal duration and the thresholds were lower in the anechoic room. It is possible that for shorter duration signals, the person may be inattentive and miss the signal and thus also the pitch information. The performance (percentage of correct responses) of the participants with 5 and 50 ms signals may thus not only be based on pitch strength but also on cognitive factors such as attention.

Schenkman and Nilsson [20] showed that when pitch and loudness information were presented together, for distances up to 200 cm to the reflecting object, the participants’ performance was almost 100 percent correct. The 500 ms recordings with the object at 50 and 100 cm in study SN2010 [6] had almost 100 percent correct responses for both the blind and the sighted. Therefore, for the 500 ms sound condition the likelihood to miss a signal and its pitch information because of non-attention is lower, and the perceptual results of the participants are probably based mostly on pitch information.

There are two possible theoretical ways to regard how the hearing system may treat the ACF values. We will here focus the analysis on the 500 ms signal, since as noted, the 5 ms and 50 ms signals may have cognitive aspects that could bias the auditory model analysis. (1) Based on the above reasoning, and if we assume that the auditory system analyses the pitch information absolutely, i.e. it does not compare the peak heights in the ACF between the recordings (when presented in a 2AFC manner), then the results indicate that the absolute threshold for detecting a pitch based on an autocorrelation process should be greater than 1.10 and 1.23 (as indicated by the autocorrelation index for the 500 ms signal) for the blind and the sighted, respectively, as shown in Table 3. (2) On the other hand, if we assume that the auditory system analyses the pitch information relatively, i.e. it compares the peak heights in the ACF between the recordings (when presented in a 2AFC manner) then the results indicate that the difference threshold for detecting the pitch based on autocorrelation should be greater than 0.27 and 0.49 (autocorrelation index) for the blind and the sighted, respectively, as shown in Table 3. For all cases, the blind persons could detect echo reflections of objects having lower peak heights in the ACF, than the sighted could.

A similar analysis of variance as for the difference thresholds of loudness was conducted for the difference thresholds of the pitch values, i.e. an unbalanced analysis with sum of squares of type 3, where one sighted test person had been excluded. Mauchly’s test of sphericity indicated that the assumption of sphericity had been

### Table 2
Absolute (top) and difference (below) threshold values of loudness (sone) for duration, room, and listener groups in study SN2010 [6]. The threshold values were calculated from the psychometric function of the blind and sighted participants’ responses at the mean proportion of correct response values of 0.73–0.76.

| Room         | Sound duration (ms) | 5     | 50    | 500   |
|--------------|---------------------|-------|-------|-------|
|              | Blind | Sighted | Blind | Sighted | Blind | Sighted |
| Anechoic     |       |         |       |         |       |         |
| Conference   | 16.8  | 17.5    | 43.7  | 45.1    | 52.9  | 53.2    |
|              | 22.6  | 24.1    | 49.4  | 53.1    | 53.6  | 55.3    |
| Anechoic     | 3.4   | 4.1     | 3.6   | 5     | 4.8   | 5.1     |
| Conference   | 3.2   | 4.7     | 4.3   | 8     | 1.1   | 2.8     |

### Table 3
Absolute (top) and difference (below) threshold values of pitch strength (autocorrelation index) for duration, room, and listener groups in study SN2010 [6]. The threshold values were calculated from the psychometric function of the blind and sighted participants’ responses at the mean proportion of correct response values of 0.73 to 0.76.

| Room         | Sound duration (ms) | 5     | 50    | 500   |
|--------------|---------------------|-------|-------|-------|
|              | Blind | Sighted | Blind | Sighted | Blind | Sighted |
| Anechoic     |       |         |       |         |       |         |
| Conference   | 0.77  | 0.88    | 0.8   | 0.96   | 1.1   | 1.23    |
|              | 1.54  | 2.21    | 1.07  | 1.69   | 1.14  | 1.41    |
| Anechoic     | 0.39  | 0.50    | 0.35  | 0.51   | 0.36  | 0.49    |
| Conference   | 0.75  | 1.42    | 0.54  | 1.16   | 0.27  | 0.54    |
violated for the effects of duration and for all its interactions. For these effects, the Greenhouse-Geisser correction was applied. For the pitch values, there was a significant difference for the different sound durations, \(F(1.30, 22.10) = 6.95, p < 0.05\), Geisser-Greenhouse epsilon = 0.65. The means for the 5, 50 and 500 ms signal durations were 1.94, 1.24 and 0.78, respectively, on the autocorrelation index. Also for pitch there was a significant interaction of sound duration and room \(F(1.12, 19.04) = 5.54, p < 0.05\), Geisser-Greenhouse epsilon = 0.56. The main factor of room was significant, \(F(1, 17) = 8.24, p < 0.05\), while the interaction of room and group did not reach significance in the unbalanced analysis. We may add that the blind and the sighted had mean difference threshold values for pitch of 0.87 and 1.78, respectively, indicating a higher sensitivity for the blind.

4.6.4. Sharpness thresholds, absolute and difference, for object detection

We chose to study one aspect of timbre, sharpness, as a potential information source for object detection by echolocation. In analog to the previous psycho-acoustical parameters, we calculated the absolute and difference threshold values of sharpness for which the blind and the sighted test persons in study SN2010 [6] could detect a reflecting object using echolocation with a correct response value of 0.73 to 0.76.

For quantitative values for sharpness, we used the psychophysical unit acum (e.g. [43]). Table 4 shows that for the blind and sighted test persons both their absolute and difference thresholds, the sharpness values were about the same. However, unlike loudness and pitch strength the sharpness information need not be greater in value for the participants to detect an object. For sharpness, a listener must distinguish between timbres. This may include cognition, involving e.g. memory processes.

When a participant in SN2010 [6] was presented with two stimuli in the 2AFC method, they distinguished the recording with the object from the recording without the object by identifying the one with the higher loudness level, stronger pitch strength or both. However, when a person uses sharpness for echolocation it is not necessary that the sound with the reflecting object has the higher sharpness value. The reflecting object might be perceived as duller, i.e. having a lower value of sharpness, than the sound without the object. A person might use this information to detect or identify an object.

Fig. 12 shows that the sharpness values for the 500 ms duration recordings, with reflecting object at 400 and 500 cm in the anechoic room, had smaller sharpness values than the recordings without an object. As mentioned earlier, these sharpness values are a perceptual measure computed by using an auditory model. Interestingly, two blind participants (no. 2 and no. 6) performed better at these conditions than all the remaining participants, i.e. their proportion correct were approximately 0.7, even at 400 and 500 cm. We looked deeper into the performance of these two high-performing echolocators by making a local linear fit for the proportion correct of these two persons and the sharpness values of the 500 ms recordings in the anechoic room that were calculated and shown in Fig. 12. Cross validation was used to find the bandwidth of the local linear fit kernel. Fig. 14 shows the corresponding local linear fits.

When the proportion correct was approximately equal to 0.7, there were two absolute threshold values for sharpness, one higher and one lower. If we consider the mean of the sharpness of the two No object recordings at this condition (i.e. anechoic room with 500 ms signal duration) it was about 1.87 (Fig. 12). Hence the difference threshold for the blind participant no. 2 would be 1.94–1.87 = 0.07 acum and 1.83–1.87 = -0.04 acum. Similarly, the difference threshold for the blind participant no. 6 would be 1.97–1.87 = 0.10 acum and 1.83–1.87 = -0.04 acum. Perceptually, this means that the two high-performing blind participants could detect the object even when the recording with the object was duller than the recording without an object. A more detailed discussion on timbre, sharpness and human echolocation is presented in section 5, Discussion.

For completeness to the analysis of loudness and pitch, an analysis of variance for the difference thresholds of timbre was also conducted, with a similar statistical model and with the same assumptions. Mauchly’s test of sphericity indicated that no violations of sphericity had occurred. There was a significant difference between the rooms, \(F(1, 17) = 128.0, p < 0.001\). The perception of sounds in the two rooms thus differed much in timbre to the listening test persons. The anechoic and the conference room had mean threshold difference values of 0.12 and 0.05 acum, respectively. The interaction of sound duration with room was also significant \(F(2, 34) = 10.61, p < 0.01\). The blind and the sighted test persons had mean difference thresholds for timbre of 0.08 and 0.09 acum, respectively.

5. Discussion

We wanted to study other aspects of human echolocation than had been done in the original study [6] where the empirical data had been collected. Signal analysis was conducted on the physical signals in order to find the physical information that could be used for echolocation and to analyze the effects of room acoustics on human echolocation. We studied sound pressure levels, auto correlations and spectral centroids. The results give further support to the effect of room acoustics on the sounds and thereby the physical attributes that are associated with it. However, the information represented in the auditory system is complex and this physical information is processed through the auditory neural system. To understand better what takes place in the auditory system of a person using echolocation, we used what we consider are the most relevant auditory models for human echolocation, that today are available in the literature. We thus studied how the corresponding perceptual attributes of sound pressure level, auto correlation and spectral centroid are processed in the human auditory system.

The results of the auditory models suggest that loudness, repetition pitch and sharpness all provide potential information for people to echolocate at distances shorter than 200 cm. At longer distances, we propose that sharpness may be used for human echolocation. A detailed discussion of these issues is presented below.

5.1. Echolocation and loudness

Of the existing loudness models, we chose the model by Glasberg and Moore [34], since it has a good fit to the equal loudness contours in ISO 2006. The results of the model were related to the proportion of correct responses of the listeners in study SN2010 [6] for calculating estimates of threshold values based on loudness. The loudness values are presented in Fig. 6 and the resulting threshold values for detecting a reflecting object, when based on the percentage correct, are shown in Table 2.

The difference in loudness level between the loudness threshold and loudness level of the recordings without the object for the 5, 50 and 500 ms duration sounds in the anechoic room were approximately 4.2, 5 and 5 sone for the sighted persons. As an example, the 5 ms signal in the anechoic room (Table 2), had a threshold of 17.5 sone for the sighted persons, while the mean for loudness with no object in this room was 13.3 sone, as shown in Fig. 6. The difference of these two values, 17.5 and 13.3, is 4.2 sone. For the conference room those differences were 5.8 and 3 sone, respectively (Fig. 7 and Table 2). The analysis of variance of the difference
threshold values for loudness showed that there were significant differences between the conditions, including some of the higher order interaction effects. These differences in loudness levels make it possible for persons to echolocate, making loudness a potential information source for echolocation [2,20]. Comparing the loudness thresholds of sighted and blind persons, the thresholds of the blind persons seem to be lower than those of the sighted test persons (Table 2). If loudness information is processed in the same manner for both groups of test persons, which is a reasonable assumption, then this analysis suggests that blind persons may echolocate at lower loudness levels than sighted persons.

5.2. Echolocation and pitch

Repetition pitch is one of the important information sources that blind people use to detect a reflecting object at shorter distances, e.g. [6,12]. We studied how this information is represented in the auditory system. For this purpose, we analyzed the time averaged spectrum and frequency weighted over time information, presented in section 4.2. “Autocorrelation Function (ACF) and Pitch analysis”. The results suggest that repetition pitch can be explained by the peaks in the temporal profile rather than by peaks in the spectral profile of the autocorrelation function. This is in agreement with a study by Yost [26], and the thesis by Bilsen [12] where the peaks in the temporal domain of the autocorrelation function form the basis for explaining the perception of repetition pitch.

However, the time averaged spectrum and frequency weighted over time analysis was not sufficient to determine the strength of the perceived pitch as the peaks appeared random in the temporal profile of the autocorrelation function. A measure of pitch strength was therefore used that showed if the peaks were random or not, and thereafter the pitch strength was computed (Eq. (6)). The means of the resulting pitch strengths are shown in Fig. 10 and the thresholds for detecting objects in Table 3. Only the pitch strength values obtained from the 500 ms duration recordings, as shown in Fig. 10, were considered for this measure, since these recordings are not likely to be influenced by cognitive factors. The pitch strength threshold was lower for blind persons than for sighted. A reasonable assumption is that pitch information is processed in the same manner for both blind and sighted persons. Then it appears that blind persons may echolocate at a lower pitch strength than sighted persons. The auditory models were used without changing its parameters for the analysis of the two groups and it is thus not possible to infer what the factors are that determine the perceptual differences.
5.3. Echolocation and timbre

To determine to what extent that sharpness is useful for echolocation, we computed the weighted centroid of the sounds from study SN2010 [6] for specific loudness using the code of Pysound3 [38], Pedrielli, Carletti, and Casazza [44] showed in their analysis that the just noticeable difference for sharpness in their study was 0.04 acm. We used this value as a criterion of sharpness for detecting reflecting objects. The results presented in Fig. 12 show that the difference in sharpness was greater than 0.04 acm between recordings with no object and with the object at distances of 50, 100, 150 and 200 cm. However, at these distances both loudness and pitch information were more prominent. Hence, at distances shorter than 200 cm, sharpness might not be a major information source for echolocation.

One may note that, in study SN2010 [6], for the 500 ms recording in the anechoic room, with the reflecting object at 400 cm and 500 cm, the sharpness difference was approximately equal to 0.04 acm when compared to the mean of the two recordings without the object (Fig. 12). A few of the blind test persons in SN2010 were able to detect objects at 400 cm. If the just noticeable difference for sharpness of 0.04 acm as found by Pedrielli, Carletti, and Casazza [44] also is a difference threshold for sharpness when echolocating sounds at longer distances than about 2 m, then sharpness can be used as vital information for blind people to detect objects at 400 cm. One might expect that there should be a linear relationship between sharpness and percentage correct, i.e. if there is a higher value of sharpness then a higher probability of detection will result. However, as discussed previously, in distinction to loudness or pitch, sharpness does not need to be larger to indicate detection, since this information may be indicated by either being perceived as dull or sharp. A psychometric function cannot depict this. An experiment controlling for sharpness information of the sound could clarify its role for echolocation.

Timbre has also been studied in the form of early reflections altering the pitch (e.g. [25]) and as “coloration” [31]. The coloration detection threshold is the level of a test reflection relative to the direct sound at which a coloration becomes just audible. It seems that coloration has a very similar meaning to timbre. Our focus has been on one aspect of timbre, namely sharpness.

Besides the information for echolocation provided by pitch, loudness and timbre, other sources are also relevant. Binaural information, inter-aural level and time differences provide information for echolocation. Papadopoulos et al. [50] argued that information for obstacle discrimination were found for the frequency dependent interaural level differences especially in the range from 5.5 to 6.5 kHz. Nilsson and Schenkman [33] found that the blind people in their study used interaural level differences more efficiently than the sighted. Self-generated sounds [51,52], as well as binaural information [53] are beneficial for echolocation. The recordings of the analyzed study in this report, SN2010 [6], had the reflecting object directly in front of the recording microphones of the dummy head, and very little binaural information was therefore provided to the test persons. As can be seen in Fig. 7 the SPL values at both ears were very similar.

We may here add that the static nature of the recordings might have resulted in poorer echolocation of the test persons. Arias et al. [54] saw echolocation as a combination of action and perception, and Thaler and Goodale [3] in their review stressed that echolocation is an active process. Tonelli, Campus and Brayda [55] showed in tests with sighted, blindfolded, test persons that various body motions like head movements, affected their echolocation. Rosenblum et al. [56] showed advantages of walking when echolocating. Furthermore, self-motion has been found to be beneficial for echolocation, as shown by Wallmeier and Wiegrebe [5]. In a study on blind children walking along a path, Ashmead, Hill and Talor [57] found that they avoided a box by utilizing non-visual information.

In psychoacoustic experiments a sound is usually presented to participants in a controlled manner and the perceptual or behavioral responses are measured. This makes it possible to identify cause-effect relationships between stimuli and response. However, although the stimuli are presented in a controlled manner in e.g. echolocation studies, the underlying cause for the echolocation is not obvious, e.g. whether it is biological, perceptual or psychological. For example, in the study SN2010 [6], the blind test persons performed better than the sighted, but the main cause for the higher performance is not evident, except that the cause is related to blindness.

The high echolocation ability of the blind may be the outcome of physiological changes in the neural system, as some studies have indicated [61,62]. To investigate this in further detail, one should change the parameters of the auditory models and analyze the results together with data from neuroimaging and psychoacoustic experiments. If it is established that the underlying ability of blind persons is based on certain physiological conditions, then the parameters of the auditory models can be varied until the results from the auditory models agree with the psychoacoustic results.

There are a number of models for animal sonar that have been developed and some of these would warrant a comparison with models used for understanding human echolocation. Simmons and Stein [63] simulated a number of different signals that are used for echolocation by bats, using cross-correlation as a basic construct. Another such model is one by Wiegrebe [21] who simulated the bat species Phyllostomus discolor with processes based on autocorrelation. It took into account both spectral and temporal processes. The model is explicitly said to have been inspired by the AIM model for human pitch perception as developed by Patterson et al [16]. We note that Wiegrebe analyzed one aspect of timbre, roughness, for the temporal performance while we chose another, viz. sharpness.

Among bat researchers today there is a debate on which receiver model that best explains target ranging, see Simmons [64] for a
review. For example, Yovel, Geva-Sagiv and Ulansky [65] used cross-correlation between outgoing pulse and returning echo to predict the echolocating performance of the Egyptian fruit bat, Rousettus aegyptiacus. There is a substantial amount of knowledge on target detection, ranging, localization and recognition for echolocation by bats that could be useful for understanding how human echolocation accomplishes similar functions.

To address representation and processing, we implemented a number of auditory models. We chose the binaural loudness model of Moore and Glasberg [37] since it agrees well with the equal loudness contours of ISO 2006 and also gives an accurate representation of binaural loudness [36]. We chose the auditory image model, AIM, of Patterson et al [16] since it uses a dynamic compressive gammachirp filterbank (dcGC) module to depict both the frequency selectivity and the compression performed by the basilar membrane. The implementation of the AIM model by Bleeck, Ives, and Patterson [39] was used to analyze repetition pitch. Finally, we used the loudness model of Glasberg and Moore [34] and the sharpness model of Fastl and Zwicker [43] to analyze sharpness. The sharpness information was obtained from the weighted centroid of the specific loudness.

The signal analysis performed on the physical stimuli showed how sound pressure level, autocorrelation and spectral centroid varied with the recordings. The results with AIM showed that the peaks in the temporal information was the likely source for echolocation at shorter distances, and this explanation is in line with the analysis by Bilsen [12] and Yost [26] in how the perception of repetition pitch is represented in people, i.e. that the information necessary for pitch perception is represented temporally in the auditory system. The analysis performed with the sharpness model showed that some blind participants in our analyzed study could have used sharpness to detect objects at longer distances and that both temporal and spectral information are required to encode this attribute. Our analysis has some similarities to that of Rowan et al [66] in utilizing models to analyze the perception of level information. Similar to their analysis we saw cross channel cues or spectral spread information as relevant for object detection, which we used in our quantification of sharpness.

Our analyses with the auditory models do not fully explain how information necessary for the high echolocation ability of blind persons is represented in the auditory system. However, we have assumed that the high echolocation ability was due mainly to a perceptual ability common to both groups. Therefore, the thresholds for the blind and the sighted persons were obtained by comparing the results of the auditory models with the perceptual results of both test groups in SN2010 [6].

The analysis with the auditory models confirmed that repetition pitch and loudness are important information sources for people when echolocating at shorter distances, which is in agreement with earlier results (e.g. [2,6,20]). Sharpness is a candidate for being an important source for echolocation both at short and long distances. Psychoacoustic experiments could determine the usefulness of sharpness and other timbre qualities such as roughness for human echolocation. Highest ecological validity is probably reached by experiments with real objects in actual environments, but laboratory studies are a viable alternative. Today, simulations of rooms and objects provide another option for how to study human echolocation, as done by e.g. Pelegrín-García, Rychtáriková and Glorieux [29].

6. Conclusions

We found support for three of the four hypotheses outlined in section 1.1 in the Introduction. The fourth hypothesis, on the use of timbre, such as sharpness, at longer distances, seems reasonable, but has to be supported by further empirical evidence. The main results of our analysis are the following. (1) Detection thresholds for echolocation based on repetition pitch, loudness and sharpness depend on the room acoustics and type of sound that is used. (2) At shorter distances, <200 cm, between person and reflecting object, repetition pitch, loudness and also sharpness provide information to detect objects by echolocation. Our analysis confirmed that repetition pitch can best be represented in the auditory system by the peaks in the temporal profile rather than by the spectral profile (see also [26]). (3) Loudness provides additional information at the shorter distances. (4) At longer distances, greater than about 300 cm, sharpness information might be used for the same purpose, but this conclusion has to be justified experimentally by varying in particular the sharpness characteristics of echolocation sounds.

For our analysis we have assumed that the auditory information for both blind and sighted persons is represented and processed in the same way. However, this assumption may not be true.

The analysis indicate that blind persons may have lower perceptual thresholds than sighted persons and could echolocate at both lower loudness and lower pitch strength levels. The recordings that form the basis for our analysis were recorded in static positions, but in real life a blind person would likely be moving, or the object would be in motion. In addition, the person is often using his/her own sounds, which is advantageous for echolocation (e.g. [51,52]). When movement is provided, and self-produced sounds are used, we believe that the thresholds for the blind persons would be significantly lower. These ideas are in alignment with the concept of surplus information in [6] that more information makes perceptual tasks easier to perform, while lack of information makes perception ambiguous and difficult. This concept follows from Gibsons [67] theory of ecological perception. In summary, we have shown the importance of pitch, loudness and timbre for human echolocation. These three characteristics have to be further studied, including their interaction and relative weights for echolocation. Additionally, timbre attributes like sharpness, needs a deeper understanding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The original data were collected together with Mats E. Nilsson. This work was partially supported by the Swedish Research Council for Health, Working Life and Welfare, (grant number 2008-0600) https://forte.se/, and by Promobilia (grant numbers 12006 and 18003) https://www.promobilia.se/, both to BS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We also thank the anonymous reviewers for constructive comments.

Appendix A. Pitch strength calculations

Pitch strength calculations using the AIM module

In Fig. A.1 is shown how pitch strength is calculated for a 200 Hz tone, when using the AIM module. Local maxima and local minima are identified, and the pitch strength is calculated between these by using Eq. (8).

The peak with the greatest peak height has the greatest pitch strength and would be the perceived frequency of repetition pitch,
where in Eq. (8) $PS$ is the calculated pitch strength, $PH$ is the height of the peak and $PHLM$ is the mean of the peak height between two adjacent local minima.

**Pitch strength calculations using strobe temporal integration**

The temporal profile of the stabilized auditory image for a recording of a 500 ms in the conference room in SN2010 [6] is shown in Fig. A.2. As stated in section 3.2, the stabilized auditory image was implemented with two modules, sf2003 and ti2003. A brief description of this implementation is given below.

Initially, the sf2003 module uses an adaptive strobe threshold to issue a strobe on the Neural Activity Pattern (NAP). After the strobe is initiated, the threshold initially rises along a parabolic path and then returns to the linear decay to avoid spurious strobes (cf. Fig. 5). When the strobes have been computed for each frequency channel of the NAP, the ti2003 module uses the strobes to initiate a temporal integration.

The time interval between the strobe and the NAP value determines the position where the NAP value is entered into the Stabilized Auditory Image (SAI). For example, if a strobe is identified in the 200 Hz channel of the NAP at a 5 ms time instant, then the level of the NAP sample at the 5 ms time instant is added to the first position of the 200 Hz channel in the SAI. The next sample of the NAP is added to the second position of the SAI. This process of

---

**Fig. A.1.** An illustration of the pitch strength measure computed using the pitch strength module of the AIM for a 200 Hz tone. The dot indicates the local maxima and the two stars are the corresponding local minima. The vertical dashed line is the pitch strength calculated using Eq. (8). The frequency in Hz was computed by inverting the time delay, $f = 1/t$.

**Fig. A.2.** The temporal profiles of the stabilized auditory image for a 500 ms signal recorded in the conference room (SN2010 [6]) at a 495 ms time frame. The dot indicates the highest peak and the corresponding values indicate the pitch strength (calculated by Eq. (6)) and frequency in Hz (using the inverse relationship of time and frequency, $f = 1/t$). The subfigure A is for the no object recording and subfigures B, C and D are for the recordings with object at 50, 100 and 200 cm respectively.
adding the levels of the NAP samples continues for 35 ms and terminates if no further strobes are identified.

In the case of strobes detected within the 35 ms interval, each strobe initiates a temporal integration process. To preserve the shape of the SAI to that of the NAP, ti2003 uses weighting, viz. new strobes are initially weighted high (also the weights are normalized so that the sum of the weights is equal to 1), making older strobes contribute relatively less to the SAI. In this way the time axis of the NAP is converted into a time interval axis of the SAI.

The temporal profile in the subfigures of Fig. A.2 was generated by summing the SAI along the center frequencies. Fig. A.2 shows that the recording with no reflecting object had a pitch strength of 0.07, while the recording with the object at 200 cm (the fourth subfigure in Fig. A.2) had a pitch strength of 0.1 at the corresponding frequencies of the repetition pitch. If this is the case for all the recordings is a matter for verification.

Previous researchers [26,35,42] analyzed the perception of repetition pitch by the autocorrelation function. We followed the same approach, since autocorrelation appears to be a good description of how repetition pitch is processed in the auditory system. To determine whether it is autocorrelation or strobe temporal integration that best accounts for human echolocation, repetition pitch and the relevant processes of the auditory system, further analysis is needed, where these two concepts are studied and compared in a number of different conditions.

References

[1] Stoffregen TA, Pittenger JB. Human echolocation as a basic form of perception and action. Ecol Psychol 1995;7:181–216.
[2] Kolarik AJ, Cirilo S, Savage S, Gardiner S, Moore BC. A summary of research investigating echolocation abilities of blind and sighted humans. Hear Res 2014;310:60–8.
[3] Thaler L, Goudale MA. Echolocation in humans: an overview. WIREs Cogn Sci 2016;7:192–220. https://doi.org/10.1002/wsc.1366.
[4] Wilbon JP. (1967) Psychoacoustics of obstacle detection using ambient or self-generated noise. In: Busnel RC, editor, Animal Sonar Systems. Biology and Bionics, vol. 1 (pp 89 – 114) Jouv.en-Josas, France: Laboratoire de Physiologie Acoustique, INRA-CNRS.
[5] Moore BC. Self-motion facilitates echo-acoustic orientation in humans. R Soc Open Sci 2014;1:140185.
[6] Schenkman BN, Nilsson ME. Human echolocation: Acoustic gaze for burst trains and acoustic noise. J Acoust Soc Am 2016;139:1059–65. https://doi.org/10.1121/1.4955803.
[7] Thaler L, Goudale MA. Echolocation in humans: an overview. WIREs Cogn Sci 2016;7:192–220. https://doi.org/10.1002/wsc.1366.
[8] Wilbon JP. Echolocation models and the relevant processes of the auditory system, further analysis is needed, where these two concepts are studied and compared in a number of different conditions.

References

[1] Stoffregen TA, Pittenger JB. Human echolocation as a basic form of perception and action. Ecol Psychol 1995;7:181–216.
[2] Kolarik AJ, Cirilo S, Savage S, Gardiner S, Moore BC. A summary of research investigating echolocation abilities of blind and sighted humans. Hear Res 2014;310:60–8.
[3] Thaler L, Goudale MA. Echolocation in humans: an overview. WIREs Cogn Sci 2016;7:192–220. https://doi.org/10.1002/wsc.1366.
[4] Wilbon JP. (1967) Psychoacoustics of obstacle detection using ambient or self-generated noise. In: Busnel RC, editor, Animal Sonar Systems. Biology and Bionics, vol. 1 (pp 89 – 114) Jouv.en-Josas, France: Laboratoire de Physiologie Acoustique, INRA-CNRS.
[5] Moore BC. Self-motion facilitates echo-acoustic orientation in humans. R Soc Open Sci 2014;1:140185.
[6] Schenkman BN, Nilsson ME. Human echolocation: Acoustic gaze for burst trains and acoustic noise. J Acoust Soc Am 2016;139:1059–65. https://doi.org/10.1121/1.4955803.
[61] Thaler L, Arnott SR, Goodale MA. Neural correlates of natural human echolocation in early and late blind echolocation experts. PLoS ONE 2011;6. e20162.

[62] Thaler L, Milne JL, Arnott SR, Kish D, Goodale MA. Neural correlates of motion processing through echolocation, source hearing, and vision in blind echolocation experts and sighted echolocation novices. J Neurophysiol 2014;111:112–27.

[63] Simmons JA, Stein RA. Acoustic imaging in bat sonar: echolocation signals and the evolution of echolocation. J Comparat Physiol A 1980;135:61–84.

[64] Simmons JA. Theories about target ranging in bat sonar. Acoust Today 2017;13:43–51.

[65] Yovel Y, Geva-Saviv M, Ulanovsky N. Click-based echolocation in bats: not so primitive after all. J Comparat Physiol A 2011;197:515–30.

[66] Rowan D, Papadopoulos T, Archer L, Goodhew A, Cozens H, Guzman Lopez R, et al. The detection of ‘virtual’ objects using echoes by humans: spectral cues. Hear Res 2017;350:205–16.

[67] Gibson JJ. The ecological approach to visual perception. Hillsdale, NJ: Lawrence Erlbaum; 1986.