

Human echolocation for target detection is more accurate with emissions containing higher spectral frequencies, and this is explained by echo intensity

L. J. Norman¹ and L. Thaler¹

¹*Department of Psychology, Durham University, Durham, DH1 3LE*

Abstract

Humans can learn to use acoustic echoes to detect and classify objects. Echolocators typically use tongue clicks to induce these echoes, and there is some evidence that higher-spectral frequency content of an echolocator's tongue click is associated with better echolocation performance. This may be explained by the intensity of the echoes. The current study tested experimentally (1) if emissions with higher spectral frequencies lead to better performance for target detection, and (2) if this is mediated by echo intensity. Participants listened to sound recordings that contained an emission and sometimes an echo from an object. The peak spectral frequency of the emission was varied between 3.5 and 4.5 kHz. Participants judged whether they heard the object in these recordings, and did the same under conditions in which the intensity of the echoes had been digitally equated. Participants performed better using emissions with higher spectral frequencies, but this advantage was eliminated when the intensity of the echoes was equated. These results demonstrate that emissions with higher spectral frequencies can benefit echolocation performance in conditions where they lead to an increase in echo intensity. The findings suggest that people who train to echolocate should be instructed to make emissions (e.g. mouth clicks) with higher spectral frequency content.

1 **Introduction**

2 Echolocation describes the process by which an organism perceives their external
3 environment through reflected sound waves (Griffin, 1944). Typically it involves the
4 creation of a sonic emission (e.g. an oral click) by the echolocator that ensonifies the
5 environment (Schörmich, Nagy & Wiegrebe, 2012), thus inducing the echoes that are
6 reflected back to the echolocator. It is a mode of perception that is most often associated
7 with certain non-human animal species, such as bats and some marine animals (Griffin &
8 Galambos, 1941; Jones, 2005), but it is also well-documented (since Supa, Cotzin &
9 Dallenbach, 1944) that humans are also capable of using echolocation (Kolarik, Cristea,
10 Pardhan & Moore, 2014; Thaler & Goodale, 2016). Specifically, some humans with vision
11 loss have been known to develop echolocation skills to an exceptional level (e.g. Teng, Puri
12 & Whitney, 2012), often without formal instruction, and typically use a tongue click as their
13 preferred type of emission to achieve this. Through this, blind individuals are able to access
14 many properties of distal objects in the environment that would otherwise be accessed
15 through vision, such as distance (Schörmich et al, 2012; Tonelli, Brayda & Gori, 2016),
16 position (Thaler, Arnott & Goodale, 2011; Teng et al, 2012), size (Teng & Whitney, 2011;
17 Thaler, Wilson & Gee, 2014), shape (Milne, Goodale & Thaler, 2014) and material of distal
18 objects (Milne, Goodale, Arnott, Kish & Thaler, 2005).

19 One important and practical use of echolocation lies in the detection and localisation of
20 objects in space, as this allows the echolocator to navigate their environment safely (Kolarik,
21 Scarfe, Moore & Pardhan, 2016). To achieve this, expert echolocators may extract a number
22 of different cues from echoes. As the distance between the echolocator and the target
23 object increases, for example, there is an increase in the delay between the onset of the

1 emission (e.g. mouth click) and that of the echo, as well as a decrease in the overall intensity
2 of the echo. These could be used as cues by an echolocator when judging an object's
3 distance. It is also possible that the emission and echo sounds will fuse, giving rise to the
4 perceptual experience of a single sound carrying a particular pitch. This fusion may be the
5 result of acoustic interference, i.e. when emission and echo temporally overlap, or it might
6 be the result of perceptual interference, i.e. "repetition pitch", where two brief sounds
7 separated by a short gap attain the quality of a single sound carrying a pitch that is inversely
8 related to the duration of the gap (Bilsen, 1966). Cues such as intensity and pitch, however,
9 are not unambiguous in their indication of object distance, as they are also affected by the
10 reflecting object's size, shape and material. In comparison, when echolocators localise an
11 object in the horizontal plane they may rely on binaural cues such as the interaural level
12 difference of the reflected sound (Rowan et al, 2013; Rowan, Papadopoulos, Edwards &
13 Allen, 2015).

14 In order for echolocation to be successfully used as a mode of perception, however, the
15 echolocator must be able to not only interpret the relevant acoustic cues in the echo, but
16 also be able to create an emission with effective signal properties. Since the early studies
17 on echolocation there has been considerable progress in our understanding of the acoustic
18 properties of tongue clicks common among expert echolocators. Mouth clicks of expert
19 echolocators, for example, typically last 3 ms and contain energy at multiple parts of the
20 audible spectrum, with peaks between 2 - 5 kHz and an additional local peak at 10 kHz
21 (Thaler et al, 2017; Zhang et al, 2017). There is also individual variability across echolocators
22 in terms of peak spectral frequency and click duration.

1 The question arises if there are certain acoustic features of the emission that confer a
2 general performance advantage. Louder emissions, for example, are associated with better
3 performance in detecting changes in room size (Flanagin et al, 2017) and in detecting the
4 presence of a single object (Thaler & Castillo-Serrano, 2016). In contrast, there are some
5 mixed findings regarding the optimum duration of the emission. Rowan and colleagues
6 (2013), for example, found that object localisation improved with the duration of the
7 emission, and Schenkman and colleagues (2016) found that object detection improved as
8 the number of emissions made increased (from 1 to 64) within a 500 ms interval. In
9 comparison, Thaler and Castillo-Serrano (2016) found that shorter emissions were
10 associated with better performance in a detection task. This discrepancy can perhaps be
11 explained by differences in the methodologies of the studies. In Thaler and Castillo-
12 Serrano's (2016) study participants used their own mouth clicks to actively detect a real
13 object in a real environment. In the other two studies, however, participants either
14 passively localised an object in virtual acoustic space (Rowan et al, 2013) or passively
15 detected an object in pre-recorded sounds (Schenkman and colleagues, 2016), in both cases
16 using artificially generated bursts of noise as emissions. The studies also differed in the
17 durations of emissions that they used. In Thaler and Castillo-Serrano's (2016) study the
18 clicks varied in their duration from 3.4 to 18.1 ms, whereas in Rowan and colleagues' (2013)
19 study the noise bursts varied from 10 to 400 ms. It is important, therefore, to determine
20 the extent to which such findings generalise across different types of emission.

21 Regarding the spectral content of the emission – that is the energy contained within the
22 emission at different frequency ranges within the audible spectrum – it has been observed
23 that individuals who produce clicks containing higher spectral frequencies perform better in

1 tasks of object detection and localisation. Thaler and Castillo-Serrano (2016), for example,
2 found that within a mixed sample of sighted participants naïve to echolocation and two
3 expert echolocators there was a positive correlation between the peak frequency of the
4 click generated by participants and their performance in a simple object detection task. This
5 correlation, however, was driven by the performance of the two expert echolocators, who
6 both produced clicks with higher spectral frequencies compared to the naïve echolocators.
7 Nonetheless, within a small sample of expert echolocators, two who produced clicks with
8 average peak spectral frequencies of 3.39 and 3.63 kHz, respectively, had smaller perceptual
9 thresholds in an angular discrimination task in comparison to a third expert echolocator
10 whose clicks had a lower average peak spectral frequency of 2.07 kHz (Thaler et al, 2017;
11 Teng et al, 2012). These results, however, only provide correlational evidence and it
12 remains to be tested experimentally whether clicks containing higher spectral frequencies
13 do in fact lead to better echolocation performance. Rowan and colleagues (2013, 2017)
14 found that participants' ability to localise and detect an object in virtual acoustic space was
15 better with noise emissions containing higher spectral frequencies. Specifically, in those
16 studies participants were more accurate in detecting or localising an object in virtual
17 acoustic space using high-pass (>3 kHz), compared to low-pass (<3 kHz), filtered noise. Yet,
18 they also found that for object azimuth localization the addition of low frequency sound (i.e.
19 use of broadband noise as compared to high-pass noise only) impaired echo processing
20 (Rowan et al., 2013), whilst for target detection the addition of low frequency sound
21 actually improved performance (Rowan et al., 2017). Thus, even though there seems to be
22 an advantage of high frequency sound, it is unclear to what degree this advantage persists
23 when low spectral frequency components are present at the same time. Stimuli in these
24 studies were noise bursts, rather than clicks, and echo onset cues had been digitally

1 removed. Thus, it is also unclear whether this result also applies to click emissions typically
2 used by echolocators.

3 There are potential reasons why an emission containing higher spectral frequencies might
4 be preferable over one containing lower spectral frequencies. Rowan and colleagues (2013)
5 suggested that the presence of low frequency sound (i.e. <3 kHz) in the emission may
6 interfere with processing of the returning echo. More generally, however, emissions that
7 contain higher spectral frequencies might simply confer a general advantage for
8 echolocation because they elicit stronger echoes for objects of finite size. Specifically, an
9 object will reflect sound more effectively if the wavelength of the sound is smaller than the
10 object's size. Given that sounds of higher frequency have shorter wavelengths, it follows
11 that objects of finite size will lead to stronger echoes when the emission contains higher
12 frequencies. This effect is due to both specular and diffraction effects. This might explain
13 the possible advantage of using emissions containing higher frequencies in tasks of object
14 detection.

15 The first aim of this study was to test the hypothesis that performance in an object
16 detection task using echolocation improves with the use of emissions containing higher
17 spectral frequencies, regardless of whether the emission is a click or noise burst, and
18 regardless of the presence of low frequency sound (i.e. <3 kHz). If this advantage were
19 found, the second aim was to determine whether it could be eliminated by equating the
20 intensity of the echoes across levels of spectral frequency. If the advantage could be
21 eliminated this way, it would suggest that the underlying cause of the association between
22 spectral frequency content of the emission and echolocation performance lies in differences
23 in echo intensity (at least in tasks of object detection). In order to test this, we first made

1 sound recordings in an anechoic room using a human model manikin. A loudspeaker was
2 fixed to the mouth of the manikin, which emitted artificially generated emissions varying in
3 their peak spectral frequency (3.5, 4.0 and 4.5 kHz). We used both artificial clicks (lasting ~5
4 ms and modelled after real human mouth clicks) as well as noise bursts (lasting 500 ms) as
5 our emissions. Recordings were made with a wooden disk at 1, 2 or 3 m distance from the
6 loudspeaker, or with no object present at all. Sighted participants then listened to these
7 recordings and judged whether or not they heard the reflecting object. They also did the
8 same under conditions in which the sounds had been digitally altered to equate the
9 intensity of echoes at each distance across emissions. We varied the object distance in
10 order to test whether any high frequency advantage generalises to situations where
11 different acoustic cues might be used to detect the object. We used the noise bursts in
12 addition to the clicks in order to determine if the results hold regardless of emission type
13 used.

14 15 **Methods**

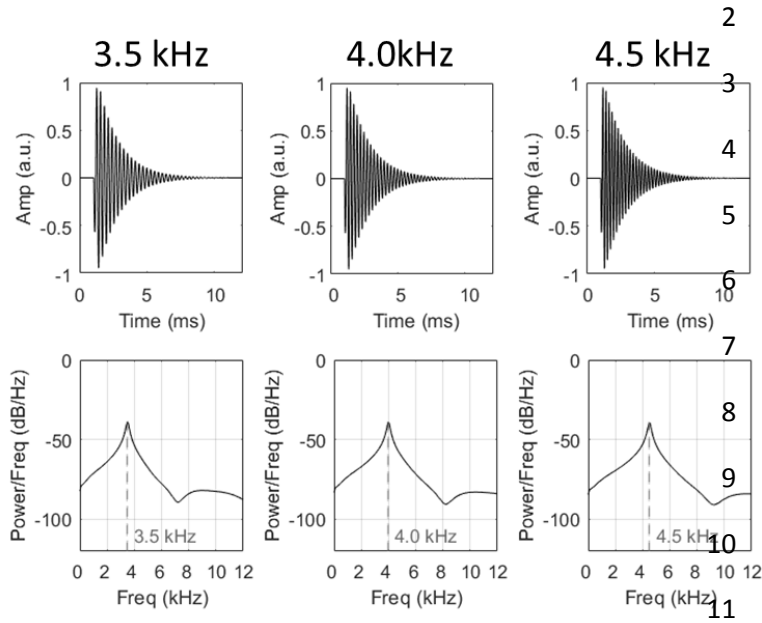
16 All Procedures followed the British Psychological Society code of practice and the World
17 Medical Association's Declaration of Helsinki. The experiment had received ethical approval
18 by the Ethics Advisory Sub-Committee in the Department of Psychology at Durham
19 University. All participants gave written informed consent to take part in this study.

Sound recordings

Emissions

The emissions used in the recordings took the form of either a click or noise burst, and were artificially generated as wav-files at a sampling rate of 96 kHz and resolution of 24-bit using MATLAB R2015b (The Mathworks, Natick, MA). The click was constructed by first creating a 10 ms sinusoid of the desired frequency (3.5, 4.0 or 4.5 kHz) and then multiplying all values up until the first half period by 0.6. This effectively simulates the rising intensity of a natural click. Then, all values after the first 1.5 periods were multiplied by the output of the decaying exponential function $y=e^{-6x}$, where x is a series of linear equally spaced values between 0 and 1 that is equal in length to the number of values in the sinusoid between the first 1.5 periods and its end. This effectively simulates the fall in intensity of a natural click. This type of sound (a sinusoid multiplied by a decaying exponential) has been suggested previously to be a good approximation of the waveform created by a human echolocator's mouth click (Martinez-Rojas, Hermosilla, Montero & Espi, 2009; Thaler et al., 2017) and has been used successfully as an artificial emission in previous tasks of echolocation (e.g. Thaler & Castillo-Serrano, 2016). The noise emissions were 500 ms bursts of noise with a 9 dB boost centred at either 3.5, 4.0 or 4.5 kHz. The 9 dB boost was created by first filtering white noise with a Kaiser window bandpass filter (1 kHz passband centred at the desired frequency, with 0.1 kHz transition bands either side). This filtered noise was then added to unfiltered white noise, at a relative amplitude ratio of 3:1. The waveforms for all emissions, as well as their spectral power functions, are shown in figure 1.

Clicks



Noise

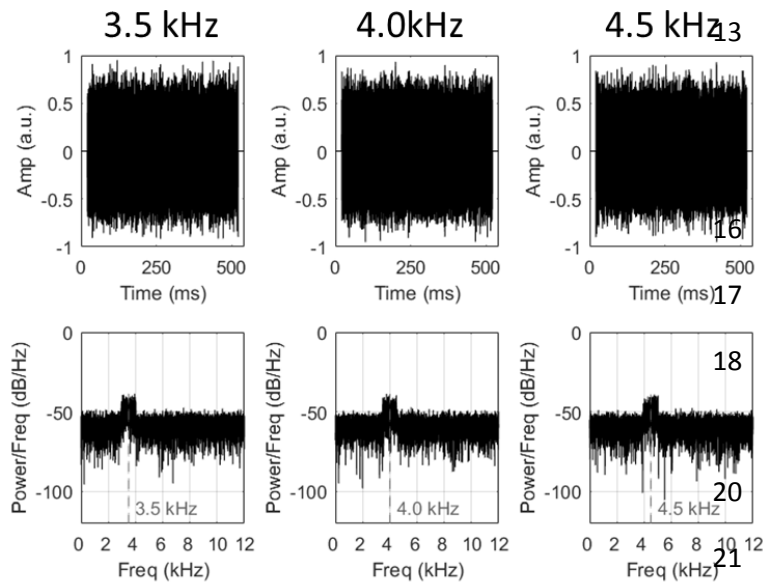


Figure 1 Waveforms and power spectral density functions of the click emissions (upper panels) and noise emissions (lower panels). The clicks were generated by multiplying a sine wave of a particular frequency (3.5, 4.0 or 4.5 kHz) with a decaying exponential. The noise was generated by selectively adding a 9 dB boost (1 kHz passband) to particular frequency components (3.5, 4.0 or 4.5 kHz) of white noise.

Apparatus and setup

All sound recordings were made in a sound-insulated and echo-acoustic dampened room (approx. 2.9 m x 4.2 m x 4.9 m) lined with foam wedges (cut-off frequency 315 Hz). Binaural sound recordings were made at a sampling rate of 96 kHz and resolution of 24-bit using a portable digital recorder (Tascam DR-100 MK2, TEAC Corporation, Japan) and in-ear microphones (Bruel & Kjaer model 4101, Denmark). The microphones were placed in the ears of a manikin, the mouth of which was positioned behind a loudspeaker (Fostex FE103En) mounted on a metal pole (1 cm diameter) and used to emit artificially generated sounds. The manikin was custom-made, consisting of a torso and head made of high-density foam covered with soft plastic having a skin like texture. The manikin was constructed based on a commercially available head and torso simulator (BodyRip Punching Dummy Torso), but head shape had been modified to better match anthropometric head and torso measurements (Algazi, Duda, Tompson & Avendano, 2001). See table 1 for the model's head, neck and ear measurements. Furthermore, 5 mm diameter holes had been drilled inside the ears to function as artificial ear canals into which binaural microphones could be placed. Woollen clothing was placed over the torso and a woollen hat was placed on the head to resemble hair. The loudspeaker was driven by a Dell Latitude E7470 laptop (Intel Core i56300U CPU 2.40 GHz, 8 GB RAM, 64-bit Windows 7 Enterprise) through a USB Soundcard (Creative Sound Blaster X-Fi HD Sound Card; Creative Technology Ltd., Creative Labs Ireland, Dublin, Ireland) and amplified by a Kramer 900N Stereo Power Amplifier (Kramer Electronics Ltd., Jerusalem, Israel). Level of amplification in all electronic equipment was held constant for the recording of all sounds across all conditions.

Table 1 Anthropometric details of the human model manikin used in the sound recordings.

Head width	14.50 cm
Head height	18.50 cm
Head depth	16.00 cm
Neck width	13.50 cm
Neck height	6.50 cm
Neck depth	14.00 cm
Pinna offset down	4.50 cm
Pinna offset back	1.50 cm
Pinna height	6.50 cm
Pinna width	3.75 cm

Separate recordings were made with a reflecting object positioned at a distance of 1, 2 or 3 m from the loudspeaker, or with no reflecting object present at all, for each emission type (~5ms click, 500ms noise) and each level of the emission's spectral frequency (3.5, 4.0, 4.5 kHz). The reflecting object was chosen to be a .8 cm thick wooden disk (50 cm diameter, made from plywood, double coated with matte emulsion paint) also mounted on a metal pole (1 cm diameter) directly facing the loudspeaker, with the height of the disk's centre matching that of the loudspeaker. The size and shape of the object had been chosen based on previous research in this area (Schenkman & Nilsson, 2010, 2011; Thaler & Castillo-Serrano, 2016) and because this size is relevant to people who use echolocation in everyday life (e.g. to detect the side panel of a bus shelter, a large tree, or a person). The set-up of the recording apparatus is shown in figure 2. Based on the speed of sound (in air at 20°C), and given the diameter of the reflecting object used, frequencies of 690 Hz (.69 kHz) and below will not return echoes, whilst higher spectral frequencies will interact with the object and form echoes through specular and diffraction effects. Thus, the object we used in our study will result in echoes from all emissions we used. In this way, the object we used is a 'worst case' scenario for testing the effects of emission spectral frequency. Yet, emissions

containing higher spectral frequencies are still expected to lead to stronger echoes because of diffraction. Thus, there will be more intense echoes as the peak spectral frequency of the emission is increased from 3.5 to 4.5 kHz.

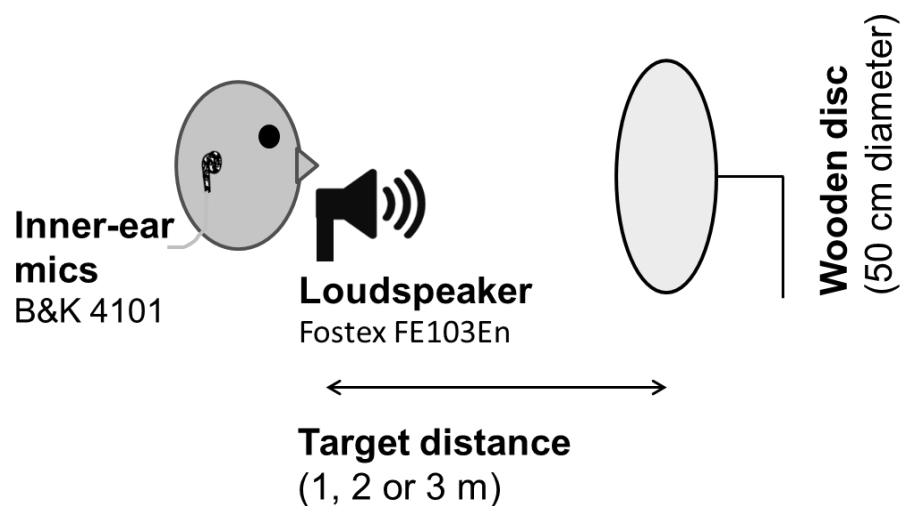


Figure 2 Sketch of the apparatus setup used for making the sounds recordings. A manikin was positioned with their mouth behind a loudspeaker. The loudspeaker emitted either a click or noise burst at one of three major spectral frequencies. A wooden disc was used as a reflecting object and was positioned at a distance of either 1, 2 or 3 m from the loudspeaker, or not present at all. Recordings of the emission and echoes (when the object was present) were made using inner-ear microphones.

Processing the sound recordings

For each type of emission (click, noise) 12 recordings were made (3 frequency ranges x 4 target conditions). In a first processing step we applied a level correction to equate recorded sound intensity of the emission across the three spectral frequency conditions. Discrepancies in the recorded emissions arose because the beam pattern of the speaker generating the emissions differed across frequencies thus leading to lower intensity of higher frequency emissions measured at the ear. To avoid the possibility that differences in the intensity of the emission would bias processing of the subsequent echo via forward

masking/echo suppression (Litovsky, Colburn, Yost, & Guzman, 1999; Wallach, Newman & Rosenzweig, 1949), we matched the intensity of emissions across conditions. This was done by multiplying the 3.5 and 4.5 kHz emission recordings with scaling factors (i.e. single numerical values) in order to equate the peak intensity of the emission to that in the 4 kHz conditions (in the target-absent recordings). We define the term “peak intensity” as the maximum absolute recorded sound value. These scaling factors are shown in table 2, and had been obtained based on the difference in peak intensity of recordings in the target absent conditions. Furthermore, a scaling factor of 1.128 was applied to the noise emission recordings to equate the intensity of the noise emission (without the echo present) to that of the click. The sounds following these alterations are shown in figure 3. As can be seen in the images in figure 3, although the intensity of the emission is matched in these recordings, the intensity of the echo increases with increasing spectral frequency of the emission. This effect is expected based on the fact that sound waves of higher frequencies are composed of shorter wavelengths, which lead to stronger echoes from an object of the size used here (see Apparatus and Setup).

Table 2 Scaling factors applied to the recordings, prior to normalisation of echo intensity. These scaling factors were applied in order to address the discrepancies in the emissions recorded at the ear that arose from irregularities in the speaker’s beam pattern across different frequencies. The recordings of the 3.5 and 4.5 kHz clicks and noise were scaled in order to equate their peak intensity to that of the 4.0 kHz recordings.

Clicks				Noise		
3.5 kHz	4.0 kHz	4.5 kHz		3.5 kHz	4.0 kHz	4.5 kHz
0.794	1.000	1.288		0.732	1.000	1.672

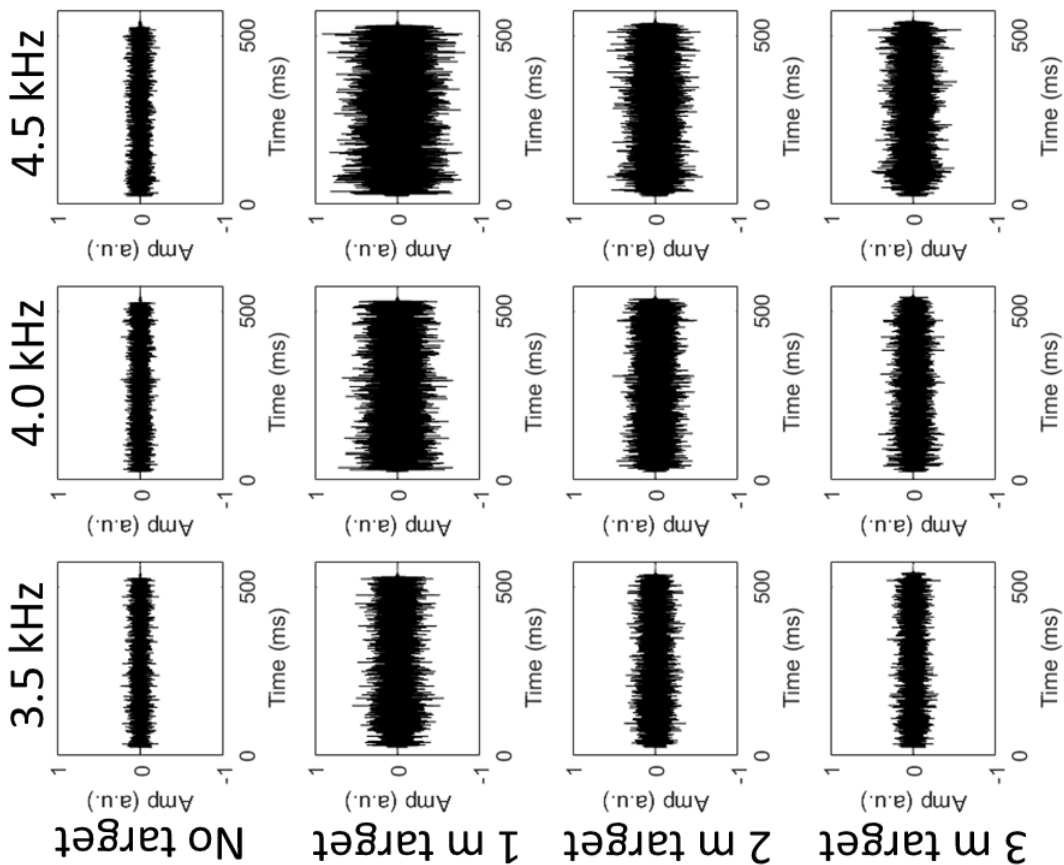
In order to test the hypothesis that this increase in echo intensity might underlie superior performance for emissions with higher spectral frequencies, we used these sounds to create a further set of stimuli in which the peak intensity of the echo had been equated across levels of spectral frequency (separately for each level of target distance). In order to do this, the temporal onset of the echo at each level of target distance had to be identified. This was done by visual inspection of the waveforms for the 4.0 kHz click recordings, with the point at which the waveform first rose above the noise floor being taken as the temporal onset of the echo. This point, taken by inspection of the recorded click waveforms at each level of target distance, was also taken as the onset point for the echo in the equivalent noise recordings. Any sound data following these identified time points in the recordings of the 3.5 kHz and 4.5 kHz emissions were then multiplied by respective scaling factors in order to equate their peak intensity to that in the recording of the 4.0 kHz emission. These scaling factors are shown in table 3, and the scaled waveforms are shown in Figure 4. Thus, two sets of sound recordings were used in the experiment – one in which the peak intensity of the echoes differed and one in which the peak intensity of the echoes had been digitally equated.

Table 3 Scaling factors applied to parts of the recordings containing the echo in order to equate the intensity of the echo. The recordings of the 3.5 and 4.5 kHz clicks and noise were scaled in order to equate the peak intensity of the echo to that of the 4.0 kHz recordings.

	Clicks				Noise		
Distance	3.5 kHz	4.0 kHz	4.5 kHz		3.5 kHz	4.0 kHz	4.5 kHz
1 m	1.315	1.000	0.792		1.282	1.000	0.788
2 m	1.280	1.000	0.897		1.364	1.000	0.880
3 m	1.315	1.000	0.758		1.270	1.000	0.727

As an example, here is the step-by-step processing that was applied to the recorded 3.5 kHz click sounds. The peak intensities of the original 3.5 kHz and 4.0 kHz click recordings with no object present were computed, where peak intensity refers to the maximum absolute sound value. Based on these values, all recordings of the 3.5 kHz clicks (i.e. target absent, 1 m target, 2 m target and 3 m target) were multiplied by 0.794 in order to bring the peak intensity of the 3.5 kHz click to that of the 4.0 kHz click. In order to create a second set of sound recordings, in which the peak intensity of the echo was equated across different values of the emission frequency, the peak intensity of the echo at each target distance was first calculated for both the 3.5 kHz and 4.0 kHz recordings. All recorded sound values after the initial onset of the echo in the 3.5 kHz click recordings were then multiplied by either 1.315 (target at 1 m), 1.280 (target at 2 m) or 1.315 (target at 3 m) in order to equate the peak intensity of the echo at 3.5 kHz to that at 4.0 kHz, but leaving the clicks unchanged.

Noise



Clicks

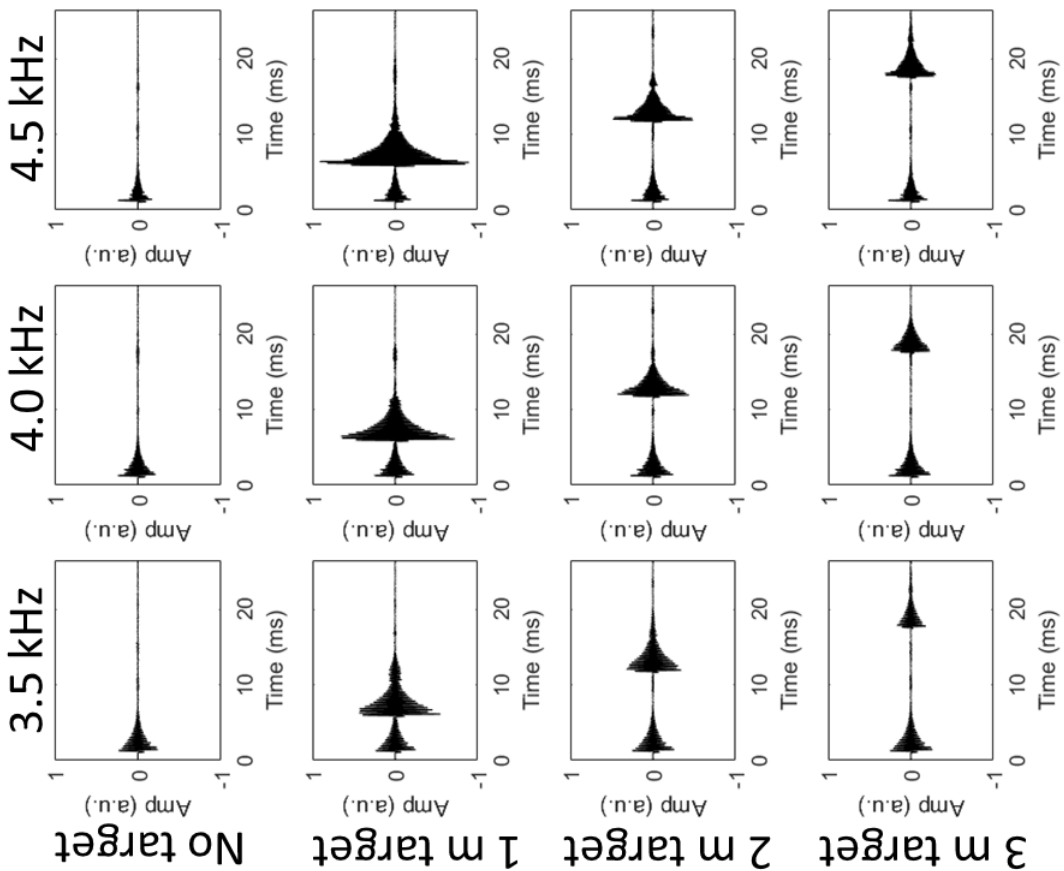
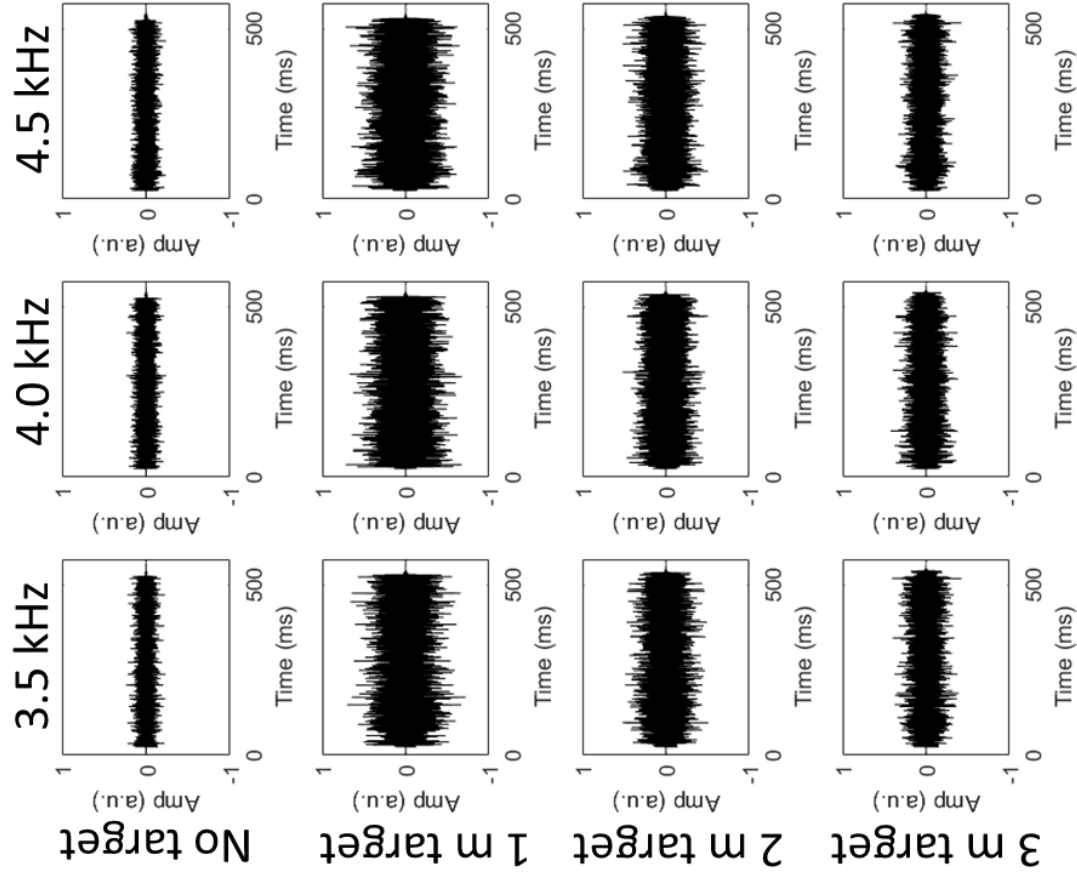


Figure 3 Waveforms of the recorded click and noise emissions. The clicks are shown in the left set of images, and the noise in the right set. From top to bottom: no target, 1 m target, 2 m target and 3 m target. From left to right (within each set): 3.5 kHz emission, 4.0 kHz emission and 4.5 kHz emission. In the click recordings, the emission and echo are temporally separated, whilst for the noise recordings they temporally overlap due to the longer emission duration (500 ms). It can be seen that emissions with higher spectral frequency content lead to echoes with higher intensity. The abbreviation a.u. refers to “arbitrary units”.

Noise



Clicks

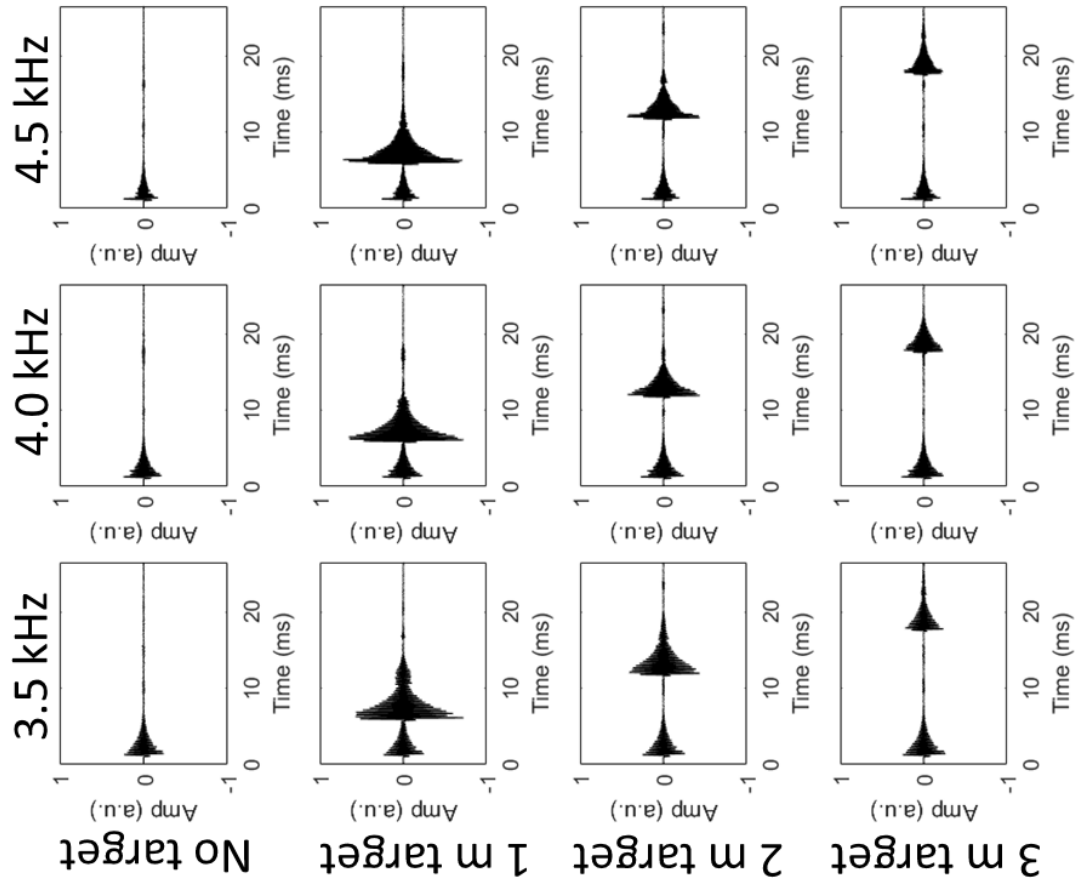


Figure 4 Waveforms of the recorded click and noise emissions that have been digitally altered to equate the intensity of the echoes across emission frequencies. The clicks are shown in the left set of images, and the noise in the right set. From top to bottom: no target, 1 m target, 2 m target and 3 m target. From left to right (within each set): 3.5 kHz emission, 4.0 kHz emission and 4.5 kHz emission. In the click recordings, the emission and echo are temporally separated, whilst for the noise recordings they temporally overlap due to the longer emission duration (500 ms). The intensity of the echoes for the 3.5 and 4.5 kHz emissions have been digitally altered to match that of the echo for the 4.0 kHz emission. The abbreviation a.u. refers to “arbitrary units”.

10

11

12

13

14

15

16

17

18

19

20

21

22

23

1 Behavioural experiment

2 *Participants*

3 12 participants (8 females, 4 males; age range 19 – 41, $M = 27.8$ years) completed all
4 components of the behavioural experiment. All participants reported having normal or
5 corrected to normal vision and hearing, and reported no prior experience using
6 echolocation. Sighted participants who are not familiar with echolocation have been
7 successfully trained to a good level of performance in previous tasks of echolocation (e.g.
8 Schenkman, Nilsson and Grbic, 2016; Tonelli et al, 2016; Rowan et al, 2013; 2015).
9 Participants were compensated either at a rate of £6/hour or with the equivalent
10 participant pool credit.

11

12 *Task and procedure*

13 Participants were tested in the same sound-insulated and echo-acoustic dampened room in
14 which the sound recordings had been made. Sounds were played to participants through
15 binaural in-ear headphones (Etymotic Research ER4B MicroPro) driven by a Dell Latitude
16 E7470 laptop (Intel Core i56300U CPU 2.40 GHz, 8 GB RAM, 64-bit Windows 7 Enterprise)
17 through a USB Soundcard (Creative Sound Blaster X-Fi HD Sound Card; Creative Technology
18 Ltd., Creative Labs Ireland, Dublin, Ireland). Sounds were played to participants at a level at
19 which the highest peak intensity (the 4.5 kHz noise emissions with target present at 1 m,
20 with echo intensity not equated) was presented at 80 dB SPL. Participants sat upright
21 wearing a blindfold and gave their response using a keyboard.

22 Trials were presented in blocks that were defined by two factors: emission type (click, noise)
23 and echo intensity (not equated, equated), and the order of these blocks was fully

counterbalanced across participants. Six repetitions of each of these blocks were run across two separate testing sessions, with 3 repetitions in the first session and 3 in the second, and the two sessions were carried out on separate days. At the start of each session, participants completed a training block of 16 trials for each of the four block permutations. Each block contained 72 randomly presented trials of two factors: emission frequency (3.5, 4.0, 4.5 kHz in the ratio 1:1:1) and target condition (absent, 1, 2, 3 m in the ratio 3:1:1:1). Participants pressed a key to initiate the onset of each trial, and were then presented with a single sound (e.g. a 4.5 kHz click with target at 2 m with echo intensity not equated). After hearing the sound they then judged using a 6-rating confidence scale (by pressing one of 6 keys on the keyboard) whether they heard the target object. The scale ranged from “very confident object absent” (1) to “very confident object present” (6), with intermediate responses to indicate less confident judgments. Participants received auditory feedback (50 ms tone) on each trial to indicate whether they were correct or not (1200 or 900 Hz tone, respectively), with a rating of 1-3 being classed as a correct response for target-absent trials and a rating of 4-6 being classed as a correct response for target-present trials. Participants were instructed to prioritise accuracy over speed in their response.

Results

Participants’ performance in the click and noise conditions was analysed separately, and participants’ sensitivity to the target was calculated separately for each permutation of the following experimental conditions: echo intensity (equated, not equated), emission frequency (3.5, 4.0, 4.5 kHz) and target distance (1, 2, 3 m). Sensitivity was calculated by

tabulating the number of responses for each of the six confidence levels for both target-absent trials and target-present trials, and using the software RScorePlus (Harvey, 2002) to fit a Gaussian unequal-variance signal detection model and derive a discriminability index (d_a). d_a does not assume equal variance between the participants' underlying *noise* and *signal + noise* Gaussian probability distributions, but is equivalent to d' (d-prime) in the case of equal variance. A higher d_a indicates a greater sensitivity to the target, and a d_a of zero indicates no sensitivity. Table 4 shows the response probabilities (mean, minimum and maximum values across all 12 participants) for each of the 6 confidence responses that participants could give on a single trial. Although the probabilities are low for some of the response categories, all participants did use the full scale. Furthermore, a chi-square statistic was calculated for each of the fitted models, computed from the log likelihood of the fit, as an indication of each of the model's goodness-of-fit. None of these statistics were significant (i.e. all p values > 0.05), indicating that the variance in the response data is well-described by the fitted models.

Table 4 Responses probabilities statistics, taken from all participants and stimulus conditions.

	1 (highly confident target absent)	2 (slightly confident target absent)	3 (guessing target absent)	4 (guessing target present)	5 (slightly confident target present)	6 (very confident target present)
Mean	0.245	0.129	0.112	0.086	0.117	0.311
Min	0.006	0.023	0.007	0.006	0.031	0.023
Max	0.413	0.298	0.477	0.259	0.330	0.451

Clicks

Behavioural results for the click emissions are shown in Figure 5 (top two panels). An initial 2 (echo intensity: equated or not) x 3 (emission frequency: 3.5, 4, 4.5 kHz) x 3 (target distance: 1m, 2m, 3m) repeated measures ANOVA was carried out, with d_a as the dependent variable. This revealed a significant interaction between echo intensity and emission frequency ($F(2,22)=10.146$, $p=0.001$, $\eta_p^2 = 0.480$), implying that clicks of different spectral frequencies affected performance differently depending on whether the intensity of the echoes was equated or not. To follow up this interaction, two separate 3 (emission frequency: 3.5, 4, 4.5 kHz) x 3 (target distance: 1m, 2m, 3m) repeated measures ANOVAs were carried out for conditions in which the echo intensity had been equated, or not.

When the echo intensity was not equated across levels of emission frequency, there was a significant effect of emission frequency ($F(2,22)=13.338$, $p<0.001$, $\eta_p^2 = 0.548$), with mean d_a values of 1.942, 2.313 and 2.543 for 3.5, 4.0 and 4.5 kHz clicks, respectively. This shows that participants' sensitivity to the target was higher for clicks containing higher spectral frequencies. Pairwise comparisons across emission frequencies (with p values adjusted for multiple comparisons using the Bonferroni method) revealed that d_a was significantly higher for 4.5 kHz clicks compared to 4.0 kHz clicks ($p=0.026$), for 4.5 kHz clicks compared to 3.5 kHz clicks ($p=0.004$) and for 4.0 kHz clicks compared to 3.5 kHz clicks ($p=0.046$). There was a significant effect of target distance ($F(1.265,13.913)=6.639$, $p=0.017$, Greenhouse-Geisser corrected, $\eta_p^2 = 0.376$), with mean d_a values of 2.500, 2.233 and 2.066, respectively for target distances of 1, 2 and 3 m, respectively. This shows that participant's sensitivity to the target decreased with target distance. None of the individual pairwise comparisons across

target distances (Bonferroni corrected) were significant, however. There was no significant interaction between emission frequency and target distance ($F(4,44)=0.946$, $p=0.447$, $\eta_p^2 = 0.079$).

In comparison, when echo intensity was equated across levels of emission frequency, there was no significant effect of emission frequency ($F(2,22)=0.900$, $p=.421$, $\eta_p^2 = 0.076$), with mean d_a values of 2.336, 2.317 and 2.216 for 3.5, 4.0 and 4.5 kHz clicks, respectively. This shows that the advantage for using clicks with higher spectral frequencies was not observed when the clicks with higher spectral frequencies did not result in more intense echoes.

There was a significant effect of target distance ($F(2,22)=7.769$, $p=0.003$, $\eta_p^2 = 0.414$), with mean d_a values of 2.528, 2.304 and 2.036 for target distances of 1, 2 and 3 m, respectively.

Again, this shows that participant's sensitivity to the target decreased with target distance.

Pairwise comparisons across target distances (Bonferroni corrected) showed that d_a was significantly higher ($p=0.014$) only for the detection of 1 m targets compared to 3 m targets.

There was no significant interaction between emission frequency and target distance ($F(4,44)=2.277$, $p=0.076$, $\eta_p^2 = 0.171$).

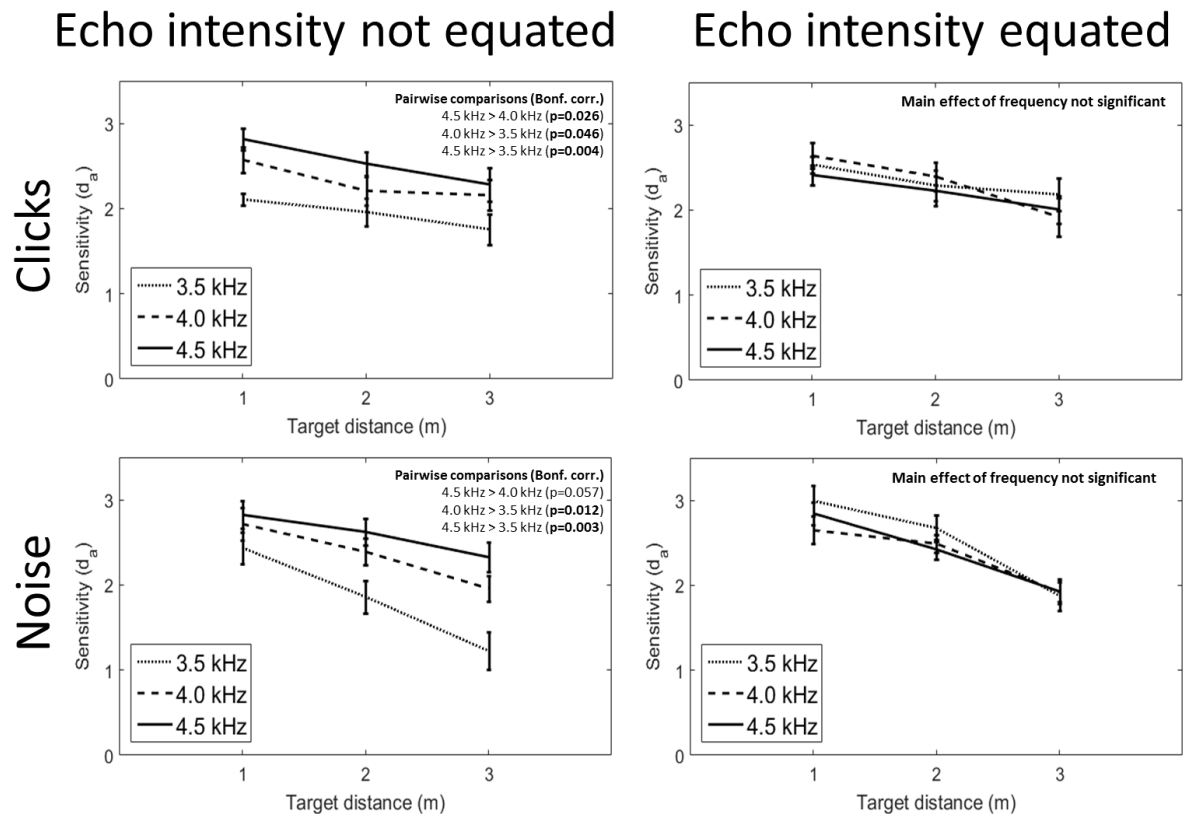


Figure 5 Results from the behavioural experiment. Participants' performance in detecting the target when the intensity of the echoes was not equated is shown in the left panels, with click emissions represented in the upper panel and noise emissions in the lower panel. As can be seen, there was an advantage in using emissions (both clicks and noise) containing higher spectral frequencies (from 3.5 to 4.5 kHz). In contrast, participants' performance in detecting the target when the intensity of the echoes was equated is shown in the right panels. In these conditions, there was no advantage for using clicks or noises containing higher spectral frequencies. Error bars represent the standard error of the mean, with between-subject variance removed.

Noise

Behavioural results for the noise emissions are shown in Figure 5 (bottom two panels). A 2 (echo intensity: equated or not) x 3 (emission frequency: 3.5, 4, 4.5 kHz) x 3 (target distance: 1m, 2m, 3m) repeated measures ANOVA was carried out, with d_a as the dependent variable. This revealed a significant interaction between echo intensity and emission frequency

($F(1.219,13.405)=18.898$, $p<0.001$, Greenhouse-Geisser corrected, $\eta_p^2 = 0.632$), implying that noise of different spectral frequencies affected performance differently depending on whether the intensity of the echoes was equated or not. To follow up this interaction, two separate 3 (emission frequency: 3.5, 4, 4.5 kHz) x 3 (target distance: 1m, 2m, 3m) repeated measures ANOVAs were carried out for conditions in which the echo intensity had been equated, or not.

When the echo intensity was not equated across levels of emission frequency, there was a significant effect of emission frequency ($F(2,22)=15.885$, $p<0.001$ $\eta_p^2 = 0.591$), with mean d_a values of 1.842, 2.355 and 2.592 for 3.5, 4.0 and 4.5 kHz noise, respectively. Pairwise comparisons across emission frequencies (Bonferroni corrected) revealed that d_a was significantly higher for 4.0 kHz noise compared to 3.5 kHz noise ($p=0.012$) and for 4.5 kHz noise compared to 3.5 kHz noise ($p=0.003$), but was not significantly higher for 4.5 kHz noise compared to 4.0 kHz noise ($p=0.057$). There was a significant effect of target distance ($F(1.276,14.037)=27.122$, $p<0.001$, Greenhouse-Geisser corrected, $\eta_p^2 = 0.711$), with mean d_a values of 2.659, 2.292 and 1.837 for target distances of 1, 2 and 3 m, respectively. Again this shows that participant's sensitivity to the target decreased with target distance. Pairwise comparisons across target distances (Bonferroni corrected) showed that d_a was significantly higher for 1 m targets compared to 2 m targets ($p=0.005$), for 1 m targets compared to 3 m targets ($p=0.001$) and for 2 m targets compared to 3 m targets ($p=0.001$). There was a significant interaction between emission frequency and target distance ($F(4,44)=4.946$, $p=0.002$, $\eta_p^2 = 0.310$). The bottom left panel in figure 5 illustrates that this interaction arose because performance differences across emission frequencies became more pronounced as target distance increased. This is also confirmed with three post-hoc

1 ANOVAs, each assessing the effect of emission frequency at one of the three target
2 distances (with p values adjusted for multiple comparisons using the Bonferroni method).
3 There was no significant effect of emission frequency at a target distance of 1 m
4 ($F(2,22)=2.921$, $p=0.225$, $\eta_p^2 = 0.210$), but there was a significant effect of emission
5 frequency at a target distance of 2 m ($F(2,22)=12.914$, $p<0.001$, $\eta_p^2 = 0.450$) and at 3 m
6 ($F(2,22)=20.257$, $p<0.001$, $\eta_p^2 = 0.648$).

7 In comparison, when echo intensity was equated across levels of emission frequency, there
8 was no significant effect of emission frequency ($F(2,22)=1.959$, $p=0.165$, $\eta_p^2 = 0.151$), with
9 3.5 kHz clicks giving a mean d_a value of 2.516, 4.0 kHz a mean value of 2.354 and 4.5 kHz a
10 mean of 2.400. This shows that the advantage for using noise with higher spectral
11 frequencies was not observed when the noise with higher spectral frequencies did not
12 result in more intense echoes. There was a significant effect of target distance
13 ($F(2,22)=36.763$, $p<0.001$, $\eta_p^2 = 0.770$), with mean d_a values of 2.831, 2.530 and 1.190 for
14 target distances of 1, 2 and 3 m, respectively. Again, this shows that participants' sensitivity
15 to the target decreased with target distance. Pairwise comparisons across target distances
16 (Bonferroni corrected) showed that d_a was significantly higher for 1 m targets compared to
17 2 m targets ($p=0.044$), for 1 m targets compared to 3 m targets ($p<0.001$) and for 2 m
18 targets compared to 3 m targets ($p=0.001$). There was no significant interaction between
19 emission frequency and target distance ($F(4,44)=2.566$, $p=0.051$, $\eta_p^2 = 0.189$).

20

21

22

General discussion

In this experiment, participants listened to pre-recorded sounds that contained an emission (either a click or noise) and judged whether they could hear an object that was present at a distance of either 1, 2 or 3 m, or absent altogether. Results showed that participants' sensitivity to the target increased as the peak spectral frequency of the emission (both clicks and noise) increased from 3.5 to 4.5 kHz. This advantage was not present, however, when the sounds were digitally altered such that the intensity of the echoes was equated across emissions of different frequency ranges. We conclude, therefore, that emissions with higher spectral frequencies can benefit echolocation performance in conditions where they lead to an increase in echo intensity.

This is the first study to directly test whether emissions containing higher spectral frequencies confer an advantage in echolocation. A previous study did show that the spectral content of the emission was associated with echolocation performance (Thaler & Castillo-Serrano, 2016), but this was only correlational evidence and the result was driven by the performance of a small number of expert echolocators who, relative to the non-expert controls, produced clicks containing higher spectral frequencies and performed significantly better. There was also evidence in a small sample of echolocators that those who produced clicks containing higher spectral frequencies performed better in a task of angle discrimination (Thaler et al, 2017; Teng et al, 2012) but, again, this was only correlational evidence. In contrast, the design of the current study allowed us to show that participants' ability to detect an object using echolocation improved when they used emissions containing higher spectral frequencies. Importantly, we included both clicks (lasting ~5 ms) and noise bursts (lasting 500 ms) as emissions in our experiment, and the effect was the

1 same for both emissions. With respect to the noise emissions, our results generally agree
2 with those by Rowan and colleagues (2013; 2017). Specifically, in those studies participants
3 were more accurate in detecting or localising an object in virtual acoustic space using high-
4 pass (>3 kHz), compared to low-pass (<3 kHz), filtered noise. Yet, they also found that for
5 object azimuth localization the addition of low frequency sound (i.e. use of broad band
6 noise as compared to high pass noise only) impaired echo processing (Rowan et al., 2013),
7 whilst for target detection the addition of low frequency sound actually improved
8 performance (Rowan et al., 2017). Importantly, in our stimuli low frequency sound (<3 kHz)
9 was always present. Thus, if any effects from low frequency sound occurred they would
10 have been the same across conditions. As such, our results clearly highlight the importance
11 of high frequency emissions due to higher echo intensity, above and beyond the effects of
12 low frequency sound as described by Rowan and colleagues (2013, 2017).

13 It should be noted, however, that producing emissions of higher spectral frequencies may
14 not confer an advantage in all situations. Flanagan and colleagues (2016), for instance,
15 found that participants' ability to estimate the size of a virtual room did not correlate with
16 the spectral content of their clicks. This makes sense, however, because emissions of higher
17 spectral frequencies will only increase echo intensity when echolocating an object of finite
18 size, but not when echolocating an enclosed space. Our current findings are therefore
19 consistent with what Flanagan et al (2016) found.

20 It is important to address whether the results of the present study can be generalised to the
21 use of echolocation in a more ecologically valid setting. Importantly, we used emissions
22 with peak frequencies between 3.5 and 4.5 kHz – a range that includes frequencies
23 contained in natural human mouth clicks of expert echolocators (Thaler et al, 2017; Zhang et

al, 2017). Furthermore the click emissions we used were similar to those that people make (Thaler et al, 2017; Zhang et al, 2017. Also, the object size we used was relevant to people who use echolocation in everyday life (e.g. to detect side panel of a bus shelter, a large tree or a person). In sum, it is reasonable to assume that our findings generalise to natural human echolocation. It was a necessity in the design of this study, however, that participants did not actively generate their own emissions, as otherwise we would have lacked control over acoustics of emissions. It has been shown in a previous study (Thaler & Castillo-Serrano, 2016), however, that when expert echolocators detect a target of the same size used here and at distances also used here, there is no difference in their performance when they create their own emissions compared to when they use artificial ones similar to those used here. We expect, therefore, that the current results generalise to active echolocation.

It is important to address whether the use of feedback tones in our experiment could limit the generalisability of our results. The positive/negative feedback tones were used as a proxy for the real feedback that echolocators would receive in a realistic setting, whose perceptual judgments would be positively or negatively reinforced based on their accuracy (e.g. physical feedback from touching objects, or from colliding with an undetected obstacle). The use of feedback in our task, therefore, was not entirely arbitrary, but we do expect that it accelerated participants' learning of echolocation.

An additional point to address is whether participants did in fact use "true" echolocation to solve this task, or whether they instead relied on a heuristic based on a simple acoustic property (e.g. a judgment based purely on loudness or pitch). Two aspects of the design of our experiment make the use of such a heuristic unlikely. Firstly, participants made a

1 judgment only on a single sound on each trial, as opposed to an alternative design in which
2 they would identify which of two sounds contained an echo. Such an alternative design
3 might conceivably allow participants to use a heuristic such as “Choose the sound that is
4 loudest”, but it would be difficult to apply such a heuristic in our single-interval-trial design
5 as there are no two sounds to compare. Secondly, the random presentation of trials in the
6 present study ensured that both the spectral content of the sound as well as the target
7 distance were unpredictable from trial to trial. The overall perceived loudness and pitch of
8 the echo, therefore, were unpredictable to participants, making it difficult for them to rely
9 on a simple heuristic to detect the object.

10 To conclude, using emissions containing higher spectral frequencies improved echolocation
11 performance in an object-detection task. Importantly, this was true for both click and noise
12 emissions, and generalised to different target distances. Emissions containing higher
13 spectral frequencies induced louder echoes, but when sound intensity was equated across
14 emissions of different frequency ranges participants no longer showed an advantage in
15 using the higher frequency emissions. This shows that emissions with higher spectral
16 content can benefit echolocation performance in conditions where they lead to an increase
17 in echo intensity. The findings suggest that people who train to echolocate should be
18 instructed to make emissions (e.g. mouth clicks) with higher spectral frequency content.

21 **Acknowledgments**

22 Supported by BBSRC (BB/M007847/1) to LT.

1

2 **References**

3

4 Algazi, V.R. & Duda, R.O. & Thompson, D.M. & Avendano, C. (2001). The CIPIC HRTF
5 database. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.
6 99 - 102. 10.1109/ASPAA.2001.969552.

7 Bilsen, F. (1966). Repetition pitch: monaural interaction of a sound with the repetition of the
8 same, but phase shifted sound. *Acustica*, 17, 295-300.

9 Griffin, D. R. (1958). *Listening in the dark: the acoustic orientation of bats and men*. New
10 Haven, CT: Yale University Press.

11 Griffin, D. & Galambos, R. (1941). The sensory basis of obstacle avoidance by flying bats.
12 *Journal of Experimental Zoology*, 86, 481-506

13 Harvey, L. (2002). RscorePlus, v 5.6.1. Parameter estimation of signal detection models.
14 <http://psych.colorado.edu/wlharvey/html/software.html>.

15

16 Jones, G. (2005). Echolocation. *Current Biology*, 15, 484-488.

17 Kolarik, A. J., Cirstea, S., Pardhan, S., & Moore, B. C. (2014). A summary of research
18 investigating echolocation abilities of blind and sighted humans. *Hearing Research*, 310C,
19 60–68.

20 Kolarik, A., Scarfe, A., & Moore, B. (2016). An assessment of auditory-guided locomotion in
21 an obstacle circumvention task. *Experimental Brain Research*, 234, 1725-1734

22 Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman S. (1999). The precedence
23 effect. *Journal of the Acoustical Society of America*, 106, 1633–1654.

24 Martinez-Rojas J. A., Hermosilla J. A., Montero R. S., & Espí P. L. L. Physical analysis of several
25 organic signals for human echolocation: Oral vacuum pulses. *Acta Acustica united with*
26 *Acustica*, 2009; 95, 325–330

- 1 Milne, J. L., Goodale, M. A., Arnott, S.R., Kish, D. & Thaler, L. (2015). Parahippocampal cortex
2 is involved in material processing through echolocation in blind echolocation experts. *Vision*
3 *Research*, 109, 139-148
- 4 Milne, J. L., Goodale, M. A. & Thaler, L. (2014). The role of head movements in the
5 discrimination of 2-D shape by blind echolocation experts. *Attention, Perception &*
6 *Psychophysics*, 76, 1828–1837.
- 7 Rowan, D., Papadopoulos, T., Archer, L., Goodhew, A., Cozens, H., Lopez, R. G., et al. (2017).
8 The detection of ‘virtual’ objects using echoes by humans: Spectral cues. *Hearing Research*,
9 350, 205-216
- 10 Rowan, D., Papadopoulos, T., Edwards, D., Holmes, H., Hollingdale, A., Evans, L., et al.
11 (2013). Identification of the lateral position of a virtual object based on echoes by humans.
12 *Hearing Research*, 300, 56–65.
- 13 Rowan, D., Papadopoulos, T., Edwards, D. & Allen, R. (2015). Use of binaural and monaural
14 cues to identify the lateral position of a virtual object using echoes. *Hearing Research*, 323,
15 32–39.
- 16 Schenkman B. N., & Nilsson M. E. (2010). Human echolocation: Blind and sighted persons'
17 ability to detect sounds recorded in the presence of a reflecting object. *Perception*, 39, 483–
18 501
19
- 20 Schenkman, B. N. & Nilsson, M. E. (2011). Human echolocation: pitch versus loudness
21 information. *Perception*, 40, 840-852.
- 22 Schenkman, B. N., Nilsson, M. E., & Grbic, N. (2016). Human echolocation: Acoustic gaze for
23 burst trains and continuous noise. *Applied Acoustics*, 106, 77-86
- 24 Schörnich, S., Nagy, A. & Wiegrebe, L. (2012). Discovering your inner bat: echo-acoustic
25 target ranging in humans. *Journal of the Association for Research in Otolaryngology*, 13,
26 673-682.
- 27 Supa, M., Cotzin, M., & Dallenbach, K. M. (1944). “Facial vision”: the perception of obstacles
28 by the blind. *American Journal of Psychology*, 57, 133-183.
- 29 Teng, S., Puri, A. & Whitney, D. (2012). Ultrafine spatial acuity of blind expert human

echolocators. *Experimental Brain Research*, 216, 483–888.

Teng, S. & Whitney, D. (2011). The acuity of echolocation: Spatial resolution in sighted persons compared to the performance of an expert who is blind. *Journal of Visual impairment and Blindness*, 105, 20-32

Thaler, L., Arnott, S. R. & Goodale, M. A. (2011). Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS ONE*, 6, e20162.

Thaler, L. & Castillo-Serrano, J. G. (2016). People's ability to detect objects using click-based echolocation: A direct comparison between mouth-clicks and clicks made by a loudspeaker. *PLoS ONE*, 11, e0154868.

Thaler, L. & Goodale, M. A. (2016). Echolocation in humans: an overview. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7, 382-393

Thaler, L., Reich, G.M., Zhang, X., Wang, D., Smith, G.E., Tao, Z., Abdullah, R., Cherniakov, M., Baker, C.J., Kish, D. & Antoniou, M. (2017). Mouth-Clicks used by Blind Human Echolocators – Signal Description and Model Based Signal Synthesis. *PLoS Computational Biology*, 13(8): e1005670.

Thaler, L., Wilson, R. C. & Gee, B. K. (2014). Correlation between vividness of visual imagery and echolocation ability in sighted, echo-naïve people. *Experimental Brain Research*, 232, 1915–1925.

Tonelli, A., Brayda, L. & Gori, M. (2016). Depth Echolocation Learnt by Novice Sighted People. *PLoS ONE*, 11, e0156654.

Wallach H., Newman, E. B., Rosenzweig, M. R. (1949). The precedence effect in sound localization. *American Journal of Psychology*, 52, 315–336

Zhang, X., Reich, G., Antoniou, M., Cherniakov, M., Baker, C.J., Thaler, L., Kish, D., Smith, G.E. (2017). Human echolocation: Waveform Analysis of Tongue Clicks. *Electronics Letters*, 53, 580-582

1

2