

Research Article

Discovering Your Inner Bat: Echo–Acoustic Target Ranging in Humans

SVEN SCHÖRNICH,¹ ANDREAS NAGY,¹ AND LUTZ WIEGREBE¹

¹*Division of Neurobiology, Department Biologie II, Ludwig-Maximilians-Universität München, Großhadernerstr. 2, 82152 Planegg-Martinsried, Germany*

Received: 23 December 2011; Accepted: 4 June 2012; Online publication: 23 June 2012

ABSTRACT

Echolocation is typically associated with bats and toothed whales. To date, only few studies have investigated echolocation in humans. Moreover, these experiments were conducted with real objects in real rooms; a configuration in which features of both vocal emissions and perceptual cues are difficult to analyse and control. We investigated human sonar target-ranging in virtual echo-acoustic space, using a short-latency, real-time convolution engine. Subjects produced tongue clicks, which were picked up by a headset microphone, digitally delayed, convolved with individual head-related transfer functions and played back through earphones, thus simulating a reflecting surface at a specific range in front of the subject. In an adaptive 2-AFC paradigm, we measured the perceptual sensitivity to changes of the range for reference ranges of 1.7, 3.4 or 6.8 m. In a follow-up experiment, a second simulated surface at a lateral position and a fixed range was added, expected to act either as an interfering masker or a useful reference. The psychophysical data show that the subjects were well capable to discriminate differences in the range of a frontal reflector. The range–discrimination thresholds were typically below 1 m and, for a reference range of 1.7 m, they were typically below 0.5 m. Performance improved when a second reflector was introduced at a lateral angle of 45°. A detailed analysis of the tongue clicks showed that the subjects typically produced short, broadband palatal clicks with durations between 3 and 15 ms, and sound levels between 60 and

108 dB. Typically, the tongue clicks had relatively high peak frequencies around 6 to 8 kHz. Through the combination of highly controlled psychophysical experiments in virtual space and a detailed analysis of both the subjects' performance and their emitted tongue clicks, the current experiments provide insights into both vocal motor and sensory processes recruited by humans that aim to explore their environment by echolocation.

Keywords: echolocation, pitch, temporal processing, vocalizations, binaural hearing

INTRODUCTION

Echolocation is defined as imaging of the environment through the auditory analysis of precisely timed, self-generated acoustic signals and the returning echoes. Thus, echolocation allows a listener to collect information about the surroundings even in complete darkness. Therefore, this ability is found in species whose habitat or way of life renders the use of vision difficult or impossible, like toothed whales (Odontoceti, deep sea) or bats (Yangochiroptera and the Rhinolophoidea, nocturnal). These species are usually especially adapted, both anatomically and neuronally (Au 1993; Popper and Fay 1995).

There are, however, more and more findings about other species using echolocation, albeit often less sophisticated, like oilbirds (*Steatornis caripensis*) (Konishi and Knudsen 1979), cave swiftlets (*Apodidae*) (Griffin 1974) some tenrec (*Tenrecidae*) (Gould 1965), respectively shrew genera (*Soricidae*) (Gould et al. 1964; Siemers et al. 2009)—and humans.

Correspondence to: Lutz Wiegrebe · Division of Neurobiology, Department Biologie II · Ludwig-Maximilians-Universität München · Großhadernerstr. 2, 82152 Planegg-Martinsried, Germany. Telephone: +49-89-218074314 email: lutzw@lmu.de

Supa et al. (1944) conducted several experiments investigating the anecdotally reported concept of “facial vision.” Blind people, who were nonetheless able to locate door openings or obstacles in front of them reported that they felt a kind of pressure change on their facial skin when being confronted with close objects. In a series of experiments, Supa et al. had both blind and blindfolded test subjects walk towards an obstacle standing in the room, with the assignment to report the moment at which they were able to perceive the obstacle for the first time and to stop walking as close to the obstacle as possible without actually touching it. This experiment was then repeated with occluded ears respectively with heavily veiled (Supa et al. 1944) or anesthetized (Kohler 1964) facial skin—the latter did not change the performance, the occluded hearing on the other hand had the test subjects collide with the obstacle every time. Thus, the concept of “facial vision” was replaced with human echolocation. However, the subjects did not produce tongue clicks. Instead, the echo-acoustic signals originated from the scuffing of the subjects’ feet.

Since then, several studies have shown that blind or blind-folded subjects can echo-acoustically detect and discriminate objects of different shape or texture (Kellogg 1962; Rice and Feinstein 1965; Rice 1967; Schenkman and Nilsson 2010). Also, object localization has been shown to be quite precise in blind human echolocation experts (Teng et al. 2011). However, to successfully avoid an obstacle, not only its angular position but also the distance to the object, i.e. the object range must be echo-acoustically assessed. In contrast to vision, where distance information is relatively difficult to infer (Palmer 1999), the distance to a sound reflecting surface can be echo-acoustically determined by estimating the time delay between emission and echo reception (Simmons 1973; Denzinger and Schnitzler 1998; Goerlitz et al. 2010). Indeed, the experiment by Supa et al. (1944) is the first to demonstrate that humans can exploit the precision of auditory temporal encoding to estimate the range of a sound-reflecting surface. However, a formal quantification of human echo-acoustic sensitivity to target range is not available to date.

Echolocation is an active sense; the psychophysical performance of a subject will not only depend on the precision of its auditory analysis of the echoes but also on the quality of the emitted sounds and how their acoustic features are shaped to facilitate the sensory task. Second, reflection properties from real, three-dimensional objects are complex and, dependent on the spectral composition of the tongue clicks, sounds may not be completely reflected but also diffracted by the object. Finally, reflections from an object of interest may interfere with other reflections from the experimental room, if the latter is not fully anechoic.

Consequently, the detailed description of the contribution of sensory and (vocal) motor components to the psychophysical experiment in active sensing requires maximal experimental control over stimulus- and environmental experimental parameters. Here, this challenge is met by transferring echo-acoustic experiments into virtual echo-acoustic space.

METHODS

Target ranging in virtual echo-acoustic space

The principal outline of an echo-acoustic experiment is rather simple, and it is fully described in terms of a linear system. The outgoing sound, typically produced by the subject’s mouth, is reflected by an ensonified object and perceived through the subjects’ ears. The spatial characteristics of the outgoing sounds, the way it is reflected by an object and the path the reflection takes from the object to the subjects’ ear drums can be described by acoustic impulse responses (IRs). To transfer such an experiment into virtual echo-acoustic space (VEAS), these IRs have to be known and applied in real time to the sounds generated by the subject. The following section describes these stages of IR acquisition and application in detail.

The vocal IR (VIR) describes how a sound emitted from the mouth changes on its way to the ensonified object. While the human mouth is an omnidirectional sender for low frequencies, where the wavelengths are large compared to the head diameter, vocalizations become more directional with increasing vocalization frequency. Assuming a vocalization with a frequency-independent (white) power spectrum, an object directly in front of the subject will be ensonified with relatively more high frequencies than a lateral object. These features are captured in the vocal IR.

VIRs were referenced against a 1/2-inch measuring microphone (B&K 4189) at an azimuth and elevation of 0° and a distance of 1 m relative to the subject’s mouth. Subjects were required to vocalize a broadband sound (unvoiced consonant ‘s’) for at least 5 s. The sound was recorded simultaneously from a headset microphone (Sennheiser HS2) and the reference microphone. First, the headset microphone’s output was calibrated to provide the same power spectrum as the reference microphone’s output when the subject was at a horizontal angle of 0° . Then the subject was turned in angles of 15° relative to the reference microphone and the difference spectra between the calibrated headset microphone, and the reference microphone were taken as the magnitude of the VIR. Only the magnitude information of the VIR was used in the experiments; the phase delay between the headset microphone, and

the reference microphone were replaced by the (experimentally varied) echo delay.

The object IR describes how the sounds are reflected by the object. For the current experiments, we assume a simple wall-like reflector whose impulse response is a Dirac impulse. Thus, it can be omitted from the overall convolution process.

The head-related IR (HRIR) describes how the reflections from the object change on their way to the subject's eardrums. For the current study, they were individually measured for each experimental subject using a calibrated sound source (1-in. broadband speaker, Aurasound NSW1 205/8). The magnitude and phase response of this speaker was equalized against the reference microphone to provide a frequency independent magnitude and linear phase between 500 and 16 kHz. HRIRs were measured with B&K 4101 binaural microphones placed in the test subject's ear canals, as a function of the horizontal angle between the subject and the calibrated sound source. Angle step size was again 15°.

Application of the IRs to generate virtual echo-acoustic space

VEAS was created for the subjects by picking up the sounds from the headset microphone and feeding them back to the subject's ears by two paths. The first (direct) path was a direct, level-adjusted path from the headset microphone to the earphones (Ethymotic research ER4S). These earphones block external sound quite effectively and thus, the subject would not perceive the own vocalization as they would in the free sound field. The level of the direct path was set so that the subjects' percept of their own voice in the anechoic chamber was most similar to their voice as they heard it with the earphones removed from the ear canals.

The second (echo) path incorporated the IRs described above. Specifically, an echo IR was generated by first convolving the microphone output with a compensation IR such that the power spectrum would be identical to that recorded by the frontal reference microphone. Second, the VIR was applied by multiplying the linear magnitude spectrum of the convolution output with the VIR magnitude corresponding to the position of the ensonified virtual object. Third, the time-domain result of this multiplication was convolved with the left and right HRIR corresponding to the angular position of the target reflector.

Fourth, the echo delay corresponding to the required range was applied by preceding the generated IR with so many zeros that, together with the digital IO delay of the hardware and convolution software, the delay corresponding to the required reflector range was generated. Next, the IR amplitude was scaled to match the range dependent geometric attenuation of a virtual

echo. For each reference range, the geometric attenuation was globally set, i.e. compared to the reference range of 1.7 m, the IR amplitude was decreased by 6 and 12 dB for ranges of 3.4 and 6.8 m, respectively. Note that geometric attenuation was not co-varied on an interval by interval basis and thus, the subjects could not solve the psychophysical task by attending to loudness cues. The frequency-dependent effects of atmospheric attenuation were not implemented because, with the current ranges and relatively low frequencies (<16 kHz), the atmospheric attenuation has only little impact.

If a second, lateral reflector was presented (Experiment 2), all steps but the first were executed independently for each reflector, and the results were added.

Auralisation

Both the direct and echo path were implemented using an RME Audio Fireface 400 audio interface. The direct path was set via 'Asio direct monitoring', routing the microphone input directly to the two output channels of the in-built headphone amplifier. This direct path had an input-output delay of about 1.7 ms. The convolution of the microphone input with the echo IRs for each ear was implemented as a VST plugin in SoundMxPro (Hoertech, Oldenburg, Germany). The sampling rate was 48 kHz. This configuration allows a real-time convolution of the microphone input with stereo IRs with up to 60,000 coefficients per channel. The overall input-output delay of the convolution device was 4.95 ms. To simulate a simple reflection from a distance of 1.7 m in VEAS, the echo IRs were preceded with zeros such that the overall delay (input-output delay of the echo path plus digital delay in the echo IR), equalled 10 ms because, at a speed of sound of 340 m/s, the echo delay created by an object at a distance of 1.7 m is 10 ms. Consequently, echo delay and thus object distance can be manipulated by varying the digital delay preceding the echo IR, taking into account the changes in geometric and atmospheric attenuation.

All experiments were conducted in a sound- and echo-attenuated chamber (1.2×1.2×2.4 m). The chamber was lined with 10-cm acoustic wedges to provide echo attenuation >40 dB at frequencies ≥500 Hz.

Procedure

In an adaptive two-alternative, two-interval, forced-choice paradigm with audio feedback, subjects were trained to find the interval that contained the shorter of two object ranges, called the reference range. The longer of the two ranges is called the test range. Each interval began with a 50-ms, 1-kHz tone pip. Directly after the tone pip, both the direct path and the echo path were activated for 5 s, such that when the subject

produced a sound, the direct path would feed directly into its ears while the echo path provided a real-time generated echo of the sound with the appropriate range, spectral content and binaural characteristics. The end of the interval was signaled by another tone pip (50 ms, 2 kHz) which was presented directly after the direct and echo path had been switched off. After a 3-s pause, the second interval was presented in the same way, with the only difference that the target range in the echo path was changed. After the second interval, the subjects were required to respond which interval contained the shorter echo delay. Subjects were given an audio feedback consisting of a 250-ms frequency chirp which was upward-modulated for positive feedback and downward-modulated for negative feedback.

In the three experimental conditions of Experiment 1, the reference range was 1.7, 3.4 or 6.8 m, corresponding to echo delays of 10, 20 or 40 ms, respectively. This reference range was changed across trials by $\pm 5\%$ to prevent the test subjects from simply memorizing the reference range's sound. At the start of an experimental run, the test range was 50% larger than the reference range. In the adaptive track, the difference between the reference and the test range was changed by a factor of two for reversals 1–5, by a factor of 1.2 for reversals 6–8 and by a factor of 1.1 for reversals 9–11.

The subjects received extensive amounts of training on each condition in each experiment until the performance stabilized over runs. The criterion for stable performance was that the standard deviation across the last three runs was less than 25% of the average across these runs. Reaching this criterion took the test subjects between 4 and 12 weeks. During this training period, the subjects not only refined their auditory analysis of the sounds but also their individual vocalization strategies. Subjects were encouraged to discuss and compare their vocalization strategies.

Sound analysis

In contrast to a classical psycho-acoustical paradigm, the subjects' performance in an echo-acoustic task will not only depend on the fidelity of its auditory processing of the perceived sounds, but also on the subjects' capability to shape their tongue clicks such that they facilitate the echo-acoustic task. Consequently, the tongue clicks of each subject in the second interval of every fifth trial of each run were subjected to a detailed analysis: this amounted to about 3,000–8,000 tongue clicks per subject and experiment. Every tongue click was analysed in terms of its duration (defined as that time interval which contained 95% of the energy), in terms of its sound level (defined as the SPL in decibels of an 85 ms

rectangular temporal window centered on each tongue click, as measured at the position of the headset microphone) and in terms of four spectral parameters. These were the spectral centroid, the loudest frequency and the cutoff frequencies at the -15 dB points above and below the loudest frequency. The sound analysis was implemented with custom Matlab programs.

The subjects were five adults (three females and two males) aged between 24 and 26. All subjects had normal hearing, as individually confirmed by pure tone audiometry, and they were sighted. They were paid an hourly rate for their participation in the experiments.

Two experiments were performed: in Experiment 1, subjects were required to discriminate range differences of a single reflective surface presented in VEAS (Figure 1A). In Experiment 2, subjects were required to discriminate range differences of a reflective surface in the presence of a second, fixed reflective surface, presented to the left (Figure 1B).

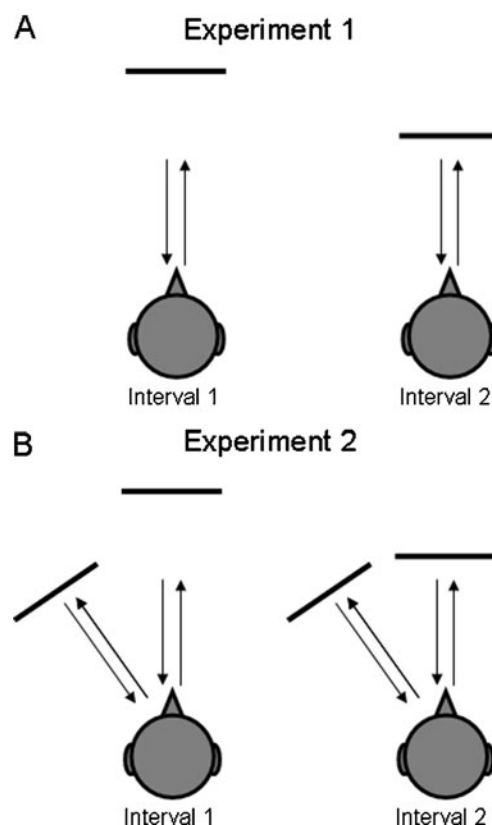


Fig. 1. Illustration of the psychophysical paradigm. In a two-interval, two-alternative forced-choice paradigm, subjects were asked to detect changes in the distance to a reflective surface positioned in virtual echo-acoustic space. Experiment 1 (upper panel) was conducted with a single frontal reflection; Experiment 2 (lower panel), was conducted in the presence of a second laterally displaced reflection at a fixed distance of 1.7 m and a lateral displacement of 15°, 30° or 45°.

RESULTS

The five subjects were successfully trained to detect changes in the range of a reflective surface, presented in VEAS. The use of VEAS ensures that the subjects did this exclusively on the basis of echoes generated from their vocal emissions. Specifically, the subjects needed to compare the time delay between the tongue click and the simulated echo of their tongue click. It appears, that subjects improved during training not only in terms of the temporal precision of their auditory analysis, but more importantly in terms of shaping their tongue click to optimally serve the echo-acoustic task. Here we will first present the psychophysical performance followed by the results of the sound analysis. Finally, we will relate the psychophysical performance to spectral and temporal parameters of their tongue clicks.

The just-noticeable target-range differences (JNDs) in Experiment 1 (cf. Figure 1A) varied considerably across subjects and thus, individual data are presented. Data are shown in Fig. 2; the individual subjects are

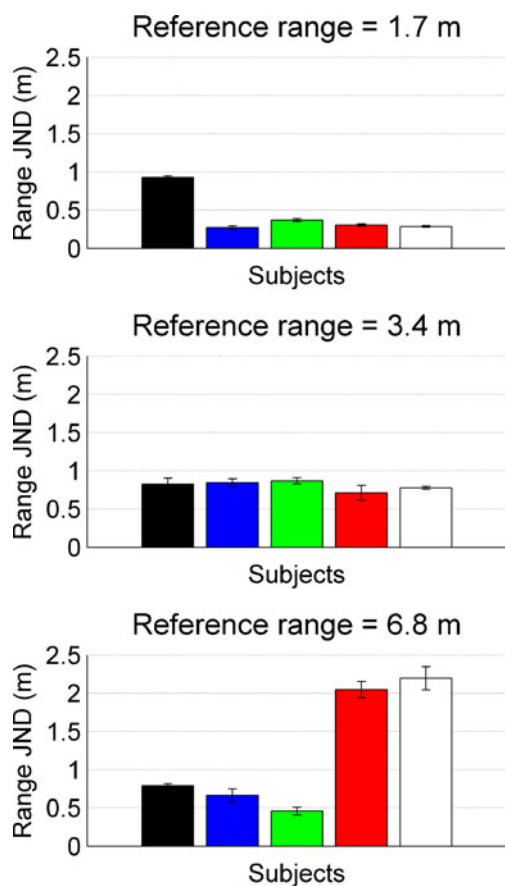


Fig. 2. Range JNDs for five different subjects, three females (*black, blue and green*) and two males (*red and white*). The different panels represent data for different reference ranges of 1.7, 3.4 and 6.8 m.

coded by the same bar colors throughout the paper. At a reference range of 1.7 m (top), most subjects could detect a change in target range of only 30–40 cm; only one subject performed significantly worse, with a range JND of 93 cm. However, while for this subject the range JND stayed approximately constant, even when the reference range was increased to 3.4 or 6.8 m, the two male subjects (red and white) showed increasing JNDs with increasing reference range.

The results of the tongue click analysis in Experiment 1 are shown in Fig. 3. Spectrograms of exemplary tongue clicks as they were produced by the subjects in the experiment are shown in the upper row. In the lower rows, the individual means of six temporal and spectral tongue click parameters are shown plotted against the individual psychophysical performance. The color code is identical to Fig. 2. All subjects produced relatively short tongue clicks with durations between 3 and 12 ms. The tongue click SPLs were more variable ranging from about 60 to 105 dB. Also, the number of tongue clicks produced to evaluate the reflection properties in an interval of the 2AFC task varied across the subjects with individual averages between only six and up to 23 tongue clicks. Most subjects produced relatively high-frequency tongue clicks with a -15 -dB bandwidth starting around 3 kHz, peaking at 6 to 7 kHz and ending around 10 to 15 kHz. Only one subject (white symbols) produced a high number of faint and low-frequency tongue clicks. The high-pass and low-pass frequencies were only 0.4 and 3 kHz, respectively. Note, however, that this subject's psychophysical performance is very similar to another subject (red symbols) who produced typical, high-frequency transient sounds.

In Experiment 2, a second reflector was introduced into the VEAS at a fixed distance of 1.7 m and a left-lateral angle of 15, 30 or 45 degree (cf. Figure 1B). The reference range of the target reflector was also 1.7 m. The range JNDs as a function of lateral reflector angle are shown in Fig. 4. Four of the five subjects from Experiment 1 took part in the experiment; their individual results are coded by the same colors as for Experiment 1. The data show that compared to Experiment 1, introducing a lateral reflector leads to an improvement in range JNDs. For subjects green and red, this effect is highly significant, for the other subjects, this effect is not significant, but the trend is in the same direction (Wilcoxon rank-sum test).

For most of the subjects, performance improved with increasing angle of the lateral reflector. A Kruskal–Wallis nonparametric ANOVA shows that this improvement is significant for subjects black and red ($p < 0.05$). For the other two subjects, the data show the same

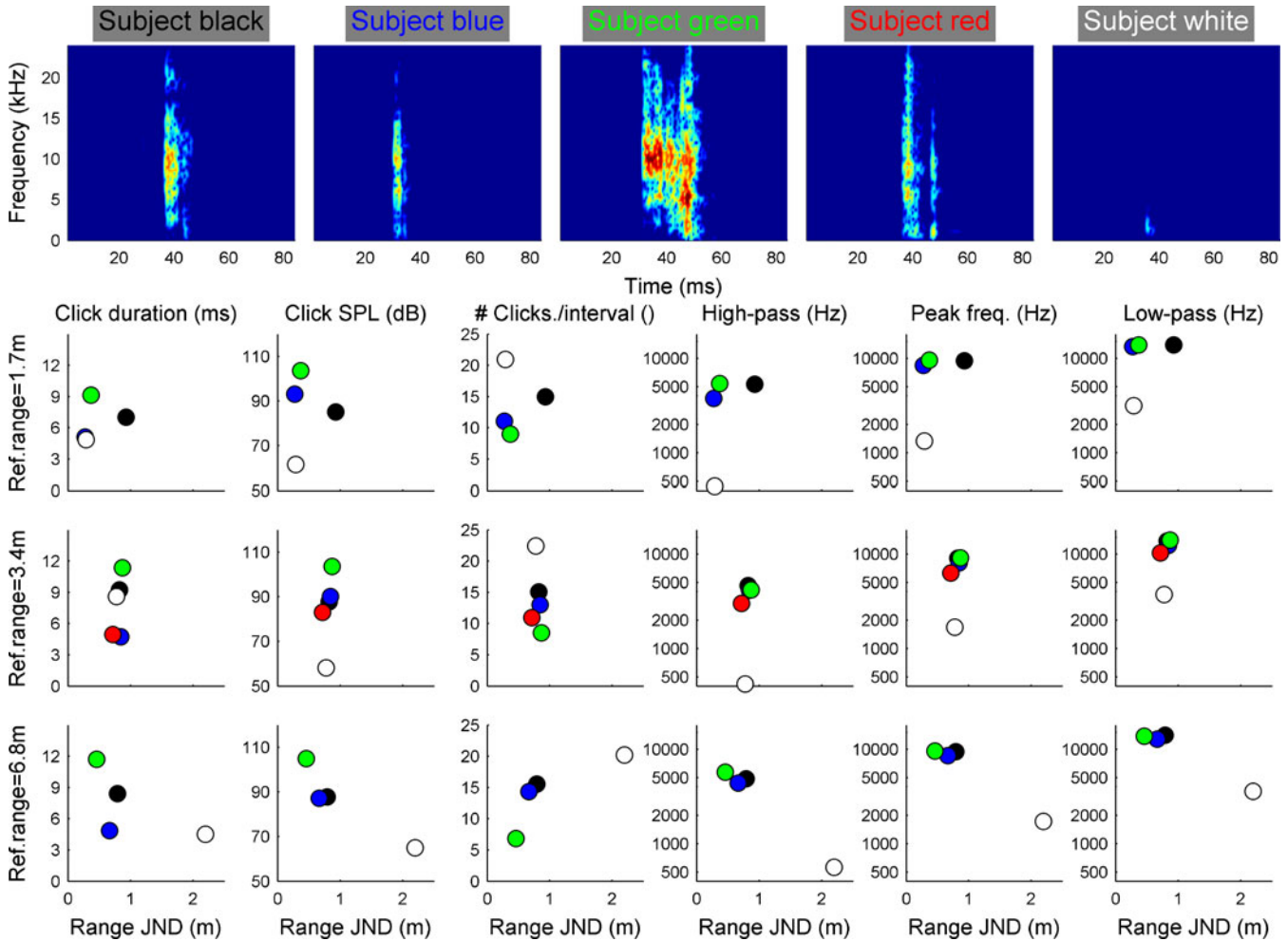


Fig. 3. Acoustic analysis of the tongue clicks emitted by the subjects to solve the psychophysical task in relation to their performance. The upper row shows spectrograms of exemplary tongue clicks by the subjects in the experiment. The sound level is color-coded; the color axis spans 80 dB. The lower columns represent different tongue click parameters as depicted in the column

title. Individual averages are shown using the same colors as in Fig. 2. Overall, the analysis shows that within a single interval subjects produced between 5 and 22 short (4–12 ms) tongue clicks. Apart from one subject (*white*) tongue clicks were typically loud (85–108 dB SPL) and high frequency (peak frequencies around 6 to 9 kHz).

pattern but the data just miss significance ($p=0.07$ and 0.11 for subjects blue and green, respectively).

The sound analysis of the individual tongue clicks in Experiment 2 is shown in Fig. 5. A significant correlation between the range JND and the peak frequency of the test subject's tongue click is seen for the lateral reflector being positioned at 45 degrees (red line in Figure 5, bottom row, second panel from right). The subject with the worst performance (subject black) had the lowest frequency tongue clicks. Additional evidence for the effect of tongue click frequency on range JNDs comes from the comparison of the range JNDs of subject red in Experiments 1 and 2: This subject increased the frequencies of his tongue clicks strongly (cf. red dots in right three columns of Figures 3 and 5). This frequency increase coincided with overall improved range JNDs.

DISCUSSION

The current psychophysical experiments show that sighted subjects can be successfully trained to echo-acoustically detect changes in the range of a reflector positioned in virtual echo-acoustic space. The subjects accomplish this task by vocally emitting short broadband sounds and evaluating the echoes generated by the reflector or reflectors. Placing the experiments in virtual echo-acoustic space allows for unprecedented experimental control of stimulus parameters and detailed documentation of the sensory-motor interactions underlying echolocation in humans. The data show that range JNDs were typically below 1 m and, for a reference range of 1.7 m, they were typically below 0.5 m.

For all subjects, performance improved when a second reflector was introduced at a lateral angle of

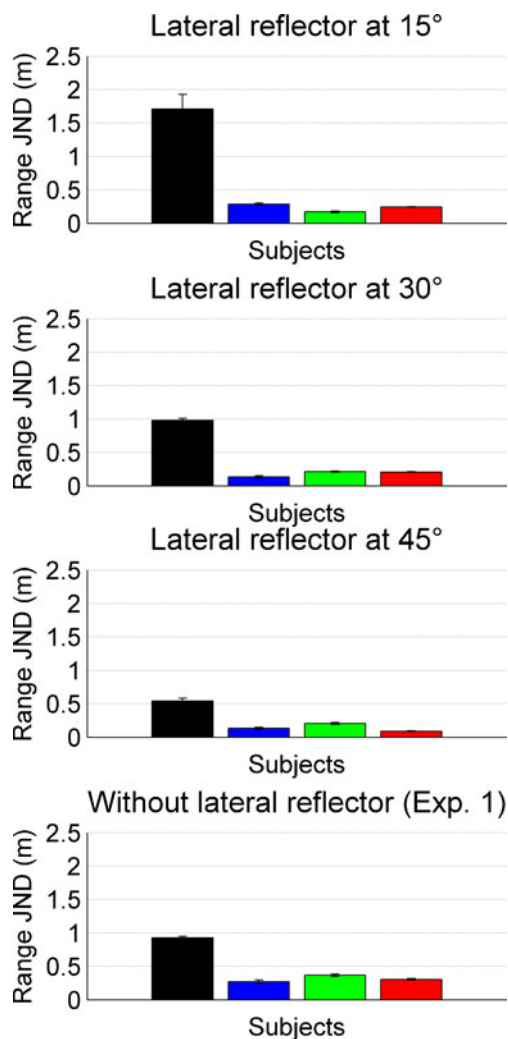


Fig. 4. Range JNDs for four subjects in the presence of a lateral reflector. Bars represent individual thresholds; the subject color code is identical to that of the previous figures. Compared to Experiment 1 (without lateral reflector, bottom panel), all subjects benefit from the second reflector when it is positioned at 45°. For lower angular separations (15° or 30°), this benefit is seen only for two subjects (green and red) while for the other two subjects, the lateral reflector is not beneficial.

45°. With smaller angular separations between the target and lateral reflector, effects were heterogeneous, i.e. some subjects still benefited from the lateral reflector (green and red) while others did not (black and blue).

A detailed sound analysis of the tongue clicks produced by the subjects during psychophysical data acquisition reveals that the tongue clicks had durations between 3 and 15 ms, and sound levels between 60 and 108 dB SPL. Except for one subject (subject white in Experiment 1 and subject black in Experiment 2), the tongue clicks had a relatively high peak frequency around 6 to 8 kHz. The number of tongue clicks produced in each 5-s interval ranged from 6 to 23.

Unlike all other mammalian senses, echolocation is an active sense. Subjects are not only the receivers of a sensory input, they are also responsible for sending out the signal exciting their habitat. The current task sets high demands on the temporal aspects of echolocation. Subjects are required to estimate the delay between the tongue click and the received echo with greatest possible accuracy. The spectral excitation patterns generated by tongue click and echo are similar but not identical. Note that the tongue click reaches the ear not only through air-borne sound but also through bone conduction (Stenfelt and Goode 2005). This typically increases excitation at lower frequencies. The auditory representation of the tongue click and its echo will also depend on frequency. Auditory filter ringing is quite prolonged at low frequencies often exceeding the 10-ms tongue click–echo delay corresponding to a target range of 1.7 m (Patterson 1976; Wiegand and Krumbholz 1997). At those (low) frequency regions where the excitation produced by tongue click and echo are of similar magnitude and where auditory filter ringing significantly exceeds 10 ms, the pulse-echo delay will produce a spectral interference pattern with a fundamental frequency of 100 Hz (the reciprocal of 10 ms). Consequently, subjects may listen for changes in the fundamental frequency of this interference pattern to solve the psychophysical task (Bassett and Eastmond 1964). Subjects could thus exploit the extraordinary human sensitivity for changes in fundamental frequency (Moore 1997; Plack et al. 2005).

Alternatively, listeners may recruit mainly information from high frequencies, where auditory filter ringing is much shorter than the tongue click–echo delay and consequently, tongue click and echo are temporally resolved. The results of the tongue click analysis indicate that four of the five subjects in Experiment 1 pursued this strategy; these four subjects all produced tongue clicks with a relatively high spectral center of gravity, as reflected in the three rightmost columns of Figs. 3 and 5. Only one subject (white) produced tongue clicks that would be low enough to produce a spectrally resolved interference pattern at the shortest reflector range (1.7 m).

Note that overall, the spectral center of gravity of the tongue clicks emitted in the current study are quite high. In another study (Flanagin et al. unpublished), where sighted listeners were required to discriminate changes in the size of a virtual room through echolocation, the peak frequencies of their tongue clicks were much lower (around 1 to 5 kHz as opposed to typically 5 to 10 kHz in the current study). In summary, the current echo-acoustic task appears to be facilitated by emitting high-frequency tongue clicks where tongue click and echo are temporally resolved by the peripheral auditory system.

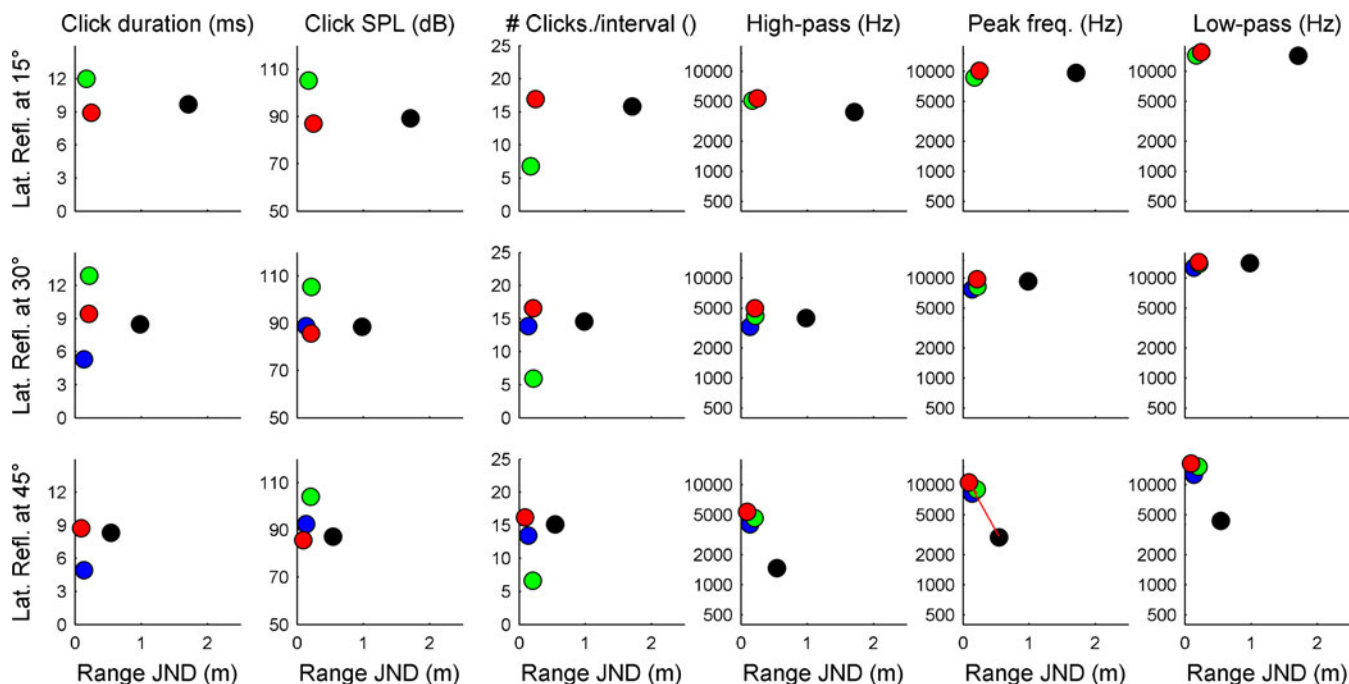


Fig. 5. Acoustic analysis of the tongue clicks emitted by the subjects in Experiment 2. Data are shown in the same format as in Fig. 3. The *red line* (bottom row, second panel from right) indicates a significant

correlation between the tongue click's peak frequency and the psychophysical performance. Note that subject *red* produced much higher-frequency tongue clicks in Experiment 2 than in Experiment 1.

In the light of the psychophysical literature on temporal processing of spectrally unresolved harmonics (for a review, see Plack et al. (2005)), the current data suggest that subjects that produce 'sharper' tongue clicks, i.e. short vocalizations with a high crest factor, may resolve range differences better than listeners with longer, more noise-like vocalizations; Houtsma and Smurzynski (1990) showed that pitch difference limens of spectrally unresolved harmonics are larger when the harmonics are in random or Schroeder phase than when they are in Cosine phase.

Similarly, research on the sensitivity to interaural time differences (ITDs) in the envelopes of high-frequency sounds have shown that stimuli with stronger envelope fluctuations result in better envelope ITD JNDs (Bernstein and Trahiotis 2002; Ewert et al. 2012). It is conceivable that subjects would also benefit from stronger envelope fluctuations to detect changes in tongue click–echo delay.

Nevertheless, the tongue click duration appears not to have a strong influence on psychophysical performance in the current data. For the three reference ranges of 1.7, 3.4 and 6.8 m, corresponding to tongue click–echo delays of 10, 20 and 40 ms, respectively, individual vocalization durations are quite stable. Thus, the subjects obviously did not adjust their vocalization duration to the vocalization–echo delay. An adaptive change of vocalization duration may also be difficult to implement as the subjects produced palatal clicks whose duration is probably very difficult to change. It is likely that the

current tongue click durations already present the lower limit of what is vocally achievable.

Introducing a second, fixed reflector at a lateral angle of 45° into the virtual echo–acoustic space led to improved range JNDs for all subjects. When, however, the lateral angle was decreased, performance deteriorated again for three of the four subjects.

The second, fixed reflector may serve as a second temporal reference, which, together with the first reference, i.e. the time of tongue click, may lead to improved range JNDs. But also in spectral term, introducing a second reflector can potentially facilitate the task because reflections from the fixed lateral reflector and the target reflector produce a comb filter effect which humans may perceive as a pitch. Changes in target–reflection delay would result in pitch changes which humans are quite sensitive too, as argued above. As the temporal separation at threshold is much shorter between the lateral and frontal reflection than between the tongue click and the target reflection, the comb filter effect generated between the lateral and the target reflector will have a higher fundamental frequency which should further increase its salience.

However, again the data argue against the hypothesis that the subjects recruited these spectral cues: The comb filter is the more pronounced the closer the lateral reflection is to the target reflection. This is so because with decreasing lateral displacement, the correlation between the echoes from the two reflec-

tors, as they would add up on the subjects' eardrums, increases. Thus, based on spectral cues, one would assume that performance should improve with decreasing spatial separation of lateral and target reflector. The data show the opposite.

Instead, the data are in line with the hypothesis that the subjects recruited temporal, timbre and spatial cues to solve the task; it appears that the subjects used the lateral reflector, based on its divergent binaural and timbre qualities, as another temporal reference against which to measure the delay of the target reflection. The timbre differences arise from differences in both the vocal- and head-related transfer functions with which the two reflections are generated. The binaural differences arise solely from the head-related transfer functions which impose both ITDs and ILDs on the echo from the lateral reflector. Both the timbre and binaural differences decrease with decreasing lateral displacement of the second reflector. Thus, the two reflections are more likely to fuse perceptually which, as it appears, makes the task more difficult for the subjects.

A direct comparison of the psychophysical performance of our human subjects with echolocation specialists, i.e. bats and dolphins show that range JNDs in humans are worse by at least an order of magnitude; at a reference range around 60 cm, bats can discriminate range differences on the order of 1.3 cm (Simmons 1973). Similarly, dolphins can discriminate range differences of 0.9, 1.5 and 3 cm for reference ranges of 1, 3 and 7 m, respectively (Murchison 1980). Apart from the neural specializations for target-range estimation (O'Neill and Suga 1979; Covey 2005), these dramatic performance differences are likely related to the exceptionally high-frequency vocalizations with which bats and dolphins solve this task. Note that the bat *Eptesicus fuscus* (Simmons 1973) uses calls no longer than 1.5 ms with a spectral peak around 60 kHz and the porpoise uses 20 μ s clicks with a peak frequency around 130 kHz.

As it is true for technical sonar, ranging accuracy scales with emission bandwidth. Considering that the bandwidth in humans is only about 10 % to 20 % of those in bats, one may expect range discrimination thresholds on the order of 10 to 20 cm. Some subjects approach these levels in some of the experimental conditions: Subject blue could discriminate 27 cm range difference at a reference range of 1.7 m in Experiment 1 and in Experiment 2 (with the additional lateral reflector) subject blue could discriminate 13 cm range difference at a reference range of 1.7 m. Also, subject green was around 20 cm in Experiment 2.

In summary, the current experiments show that subjects reliably detect relatively small changes in the range of an ensonified reflector positioned in virtual

echo-acoustic space. A detailed analysis of the tongue clicks emitted during the experiments shows that most subjects produced relatively loud, short, broadband tongue clicks to solve the task. Both the high spectral center of gravity and the effects of introducing a second, fixed reflector at a lateral angle provide at least circumstantial evidence that subjects did not rely on pitch cues as it was argued before. Instead, subjects appear to rely on temporal and possibly timbre cues (in Experiment 2). Thus, through the combination of highly controlled psychophysical experiments in virtual space and a detailed analysis of both the subjects' performance and their emitted tongue clicks, the current experiments shed first light onto the psychophysical basis of target ranging through echolocation in humans. The current techniques allow exploring the detailed adjustment of both vocal sound generation and auditory analysis to the many environmental challenges imposed on human navigation through echolocation.

ACKNOWLEDGMENTS

The authors would like to thank Daniel Kish and Magnus Wahlberg for many fruitful discussions on the topic of human echolocation. We would also like to thank the anonymous reviewers for the detailed and constructive comments. This work was supported by a grant from the 'Deutsche Forschungsgemeinschaft' (Wi 1518/9) to L.W.

REFERENCES

- AU WWL (1993) The sonar of dolphins. Springer, New York
- BASSETT IG, EASTMOND EJ (1964) Echolocation: measurement of pitch versus distance for sounds reflected from a flat surface. *J Acoust Soc Am* 36:911-916
- BERNSTEIN LR, TRAHOTIS C (2002) Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli". *J Acoust Soc Am* 112:1026-1036
- COVEY E (2005) Neurobiological specializations in echolocating bats. *Anat Rec A Discov Mol Cell Evol Biol* 287:1103-1116
- DENZINGER A, SCHNITZLER HU (1998) Echo SPL, training experience, and experimental procedure influence the ranging performance in the big brown bat, *Eptesicus fuscus*. *J Comp Physiol [A]* 183:213-224
- EWERT SD, KAISER K, KERNSCHMIDT L, WIEGREBE L (2012) Perceptual sensitivity to high-frequency interaural time differences created by rustling sounds. *J Assoc Res Otolaryngol* 13(1):131-143
- GOERLITZ HR, GEBERL C, WIEGREBE L (2010) Sonar detection of jittering real targets in a free-flying bat. *J Acoust Soc Am* 128:1467-1475
- GOULD E (1965) Evidence for echolocation in the *Tenrecidae* of Madagascar. *Proc Am Phil Soc* 109:352-360
- GOULD E, NOVICK A, NEGUS NC (1964) Evidence for echolocation in shrews. *J Exp Zool* 156:19-37
- GRIFFIN DR (1974) Listening in the dark: acoustic orientation of bats and men. Dover Publications Inc., New York
- HOUTSMA AJM, SMURZYNSKI J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87:304-310

- KELLOGG WN (1962) Sonar system of the blind. *Science* 137:399–404
- KOHLER I (1964) Orientation by aural clues. *Am Found Blind, Res Bull* 4:14–53
- KONISHI M, KNUDSEN EI (1979) The oilbird: hearing and echolocation. *Science* 204:425–427
- MOORE BCJ (1997) *Introduction to the psychology of hearing*. Academic Press, San Diego
- MURCHISON AE (1980) Detection range and range resolution of porpoise. In: Busnel RG, Fish JF (eds) *Animal sonar systems*. Plenum Press, New York, pp 43–70
- O'NEILL WE, SUGA N (1979) Target range-sensitive neurons in the auditory cortex of the mustache bat. *Science* 203:69–73
- PALMER SE (1999) *Vision science: photons to phenomenology*. MIT Press, Cambridge
- PATTERSON RD (1976) Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am* 59:640–654
- PLACK CJ, OXENHAM AJ, FAY RR (2005) *Pitch: neural coding and perception*. Springer, New York
- POPPER AN, FAY RR (1995) *Hearing by bats*. Springer, New York
- RICE CE (1967) Human echo perception. *Science* 155:656–664
- RICE CE, FEINSTEIN SH (1965) Sonar system of the blind: size discrimination. *Science* 148:1107–1108
- SCHENKMAN BN, NILSSON ME (2010) Human echolocation: blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object. *Perception* 39:483–501
- SIEMERS BM, SCHAUERMANN G, TURNI H, VON MERTEN S (2009) Why do shrews twitter? Communication or simple echo-based orientation. *Biol Lett* 5:593–596
- SIMMONS JA (1973) The resolution of target range by echolocating bats. *J Acoust Soc Am* 54:157–173
- STENFELT S, GOODE RL (2005) Bone-conducted sound: physiological and clinical aspects. *Otol Neurotol* 26:1245–1261
- SUPA M, COTZIN M, DALLENBACH KM (1944) Facial vision the perception of obstacles by the blind. *Am J Psychol* 57:133–183
- TENG S, PURI A, WHITNEY D (2011) Ultrafine spatial acuity of blind expert human echolocators. *Exp Brain Res* 216:483–488
- WIEGREBE L, KRUMBHOLZ K (1997) Temporal integration of transient stimuli and the auditory filter. *Br J Audiol* 31:107–108