A DEVICE FOR HUMAN ULTRASONIC ECHOLOCATION

Anonymous ICME submission

ABSTRACT

Representing and interacting with the external world constitutes a major challenge for persons who are blind or visually impaired. Often, non-visual sensory modalities such as audition and somato-sensation acquire increased importance in sampling the environment. In many non-human species operating without vision, this is achieved by active echolocation, in which acoustic pulses are emitted and their reflections interpreted. The ultrasonic pulses employed by, e.g., bats and dolphins are highly informative but cannot be heard by humans. Here we present a device that combines principles of echolocation and spatial hearing to present users with environmental cues that are i) not otherwise available to the human auditory system and ii) richer in object and spatial information than similar systems employing extensive processing of echo information.

Index Terms— echolocation, ultrasound, navigation aid, naturalistic stimuli

1. INTRODUCTION

Representing and interacting with the external world constitutes a major challenge for persons who are blind or visually impaired. Often, non-visual sensory modalities such as audition and somato-sensation acquire increased importance in sampling the environment. In many non-human species operating without vision, this is achieved by active echolocation, in which acoustic pulses are emitted and their reflections interpreted [1]. The ultrasonic pulses employed by, e.g., bats and dolphins are highly informative but cannot be heard by humans. Here we present a device that combines principles of echolocation and spatial hearing to present users with environmental cues that are i) not otherwise available to the human auditory system and ii) richer in object and spatial information than similar systems employing extensive processing of echo information.

1.1. Echolocation in animals (bats/cetaceans)

In environments where vision is ineffective some animals have evolved the capacity for echolocation. The most sophisticated echolocation is found in microchiropteran bats (micro-bats), and odontocetes (dolphins and toothed whales). Micro-bats are able to fly through dense forest and catch insects on the wing in total darkness. Likewise, dolphins can detect and catch fish in opaque water. Also, arguably simpler echolocation is also found in oilbirds, swiftlets, Rousettas megabats, and some shrews and tenrecs, and even rats [2]. There is even some evidence that mole rats may use a form of seismic echolocation, by sending sound through earth [3].

Microbats, which use sound frequencies in echolocation ranging from 25-150 kHz, use several different types of echolocation call [4] One type of call, the broadband call or chirp, is a tone of less than 5ms duration, rapidly sweeping downward over a wide range of frequencies over its time-course. This is used for localization at close range. Another call is the narrowband call, which tends be over a narrower range of frequencies, and is longer in duration. This is used for detection and classification of objects such as flying insects, typically at longer range. Odontocetes use sound frequencies in echolocation ranging up to 200kHz [5]. In contrast to microbats, odontocetes use clicks, these are shorter in duration than the calls of bats. Dolphin may need to use these shorter calls as sound travels about 4 times faster in water, whereas bats may need longer signals to have sufficiently energetic calls for echolocation in air. Dolphins can even use echolocation to detect features that are invisible to vision: for example, dolphins can tell visually identical hollow objects apart based on their thickness [6].

1.2. Spatial hearing and echolocation in humans

1.2.1. Major cues informing human spatial hearing

The cues underpinning human auditory spatial perception remain the subject of considerable study and are very briefly summarized here. Localization in the azimuthal (horizontal) dimension depends heavily on binaural cues: timing and intensity differences between sounds arriving at each of the two ears; for vertical localization, the major cues are monaural spectral transformations induced by the asymmetric shape of the pinna, the visible outer portion of the ear [7]. Auditory distance perception is less well characterized than the other two dimensions, though evidence suggests that intensity and the ratio of direct to reverberant energy play major roles in distance judgments [8]. Notably, most research on distance perception does not include pulse-echo delays, which serve as primary distance cues for echolocating animals.
1.2.2. Active echolocation in humans

Humans are not typically considered among the echolocating species. Remarkably, however, some blind persons have demonstrated the use of active echolocation in their daily lives, interpreting reflections from self-generated tongue clicks for such tasks as obstacle detection [9], distance discrimination [10], and object localization [11, 12]. The underpinnings of human echolocation in blind (and sighted) people remain poorly characterized, though some informative cues [13], neural correlates [14] and models [15] have been proposed. Most evidence in the existing literature suggests that human echolocation ability, even in blind, trained experts, does not approach the precision and versatility found in organisms with highly specialized echolocation mechanisms.

While the practice of active echolocation is currently not commonly taught, it is recognized as an orientation and mobility method, though recommended techniques, evaluative measures, and acceptance vary widely [16, 17]. Similarly, numerous electronic travel aids that utilize sonar have been developed (e.g., [18, 19, 20, 21]); however, very few provide information other than rangefinding or a processed localization cue, and none appears to be in common use.

1.2.3. Distance resolution

The relative and absolute distances of objects is determined from the timing of echoes. [22] measured how well humans can estimate the gaps between sounds (5ms noise bursts) depending on the size of the gap. We can use this to estimate how well humans might be able to estimate the distance between echo source depending on the size of the distance. To do this we multiply the gap size and just-noticable-difference (a measure of estimation capacity) by the speed of sound in air over two. The results are plotted in Figure 3; solid line for the unmodified human, dotted line for with the device (where time is dilated by a factor of 20). Observe that it improves discrimination over large distances (0.5-5m) by 1.5-3 times better. It is in discriminating between close echoes, such as might come from a complex object, that the device brings the most advantage. Whereas without the device, Figure 3 suggests that he smallest distance that can be discriminated over is about 25cm (the floor on the y-axis of the solid line), with the device this distance is 1.3 cm. This has the potential to provide rich information about the object.

Another consideration is psychological phenomenon known as the precedence effect; if two sounds are heard in quick succession, the location of the sound it dominated by the first sound. This is obviously a problem for human echolocation. For brief sounds such as clicks this occurs if the interval between the sounds is less than 5ms. This means without the device, echoes from objects with a relative distance of ∼85cm will be localized to the position of the first sound. With the device, slowing time by 20 times, the smallest relative distance becomes ∼4.3 cm.

1.2.4. Directional resolution

The first point with regards to determining direction is the minimum size of an object that will cause significant reflections. This will be approximately the wavelength of the sound. Human vocalizations used by echolocators have most power around 3kHz, but do have some power at higher frequencies [14]. Also, humans can only hear up to 20 kHz. Thus the smallest width of an object that sounds producible by humans will reflect off will be about 2cm-10cm. The device produces chirps with substantial power up to 50kHz. This gives a smallest size of 0.7 cm, 3-25 times smaller than for the unaided human.

The second point is the discrimination capacity of the human auditory system for auditory directional cues. For discriminating sounds along the left-right axis there are two main cues; the interaural time difference (ITD) and the interaural intensity difference (IID). The use of high frequencies will greatly increase the IIDs, as high frequencies will diffract around the head less. The time dilation will increase the ITDs 20-fold. The largest ITD produced by the human head is about 0.690 ms. The distance around the head between the microphones will be the main determinant of their maximum ITD. This distance is approximately 20cm, given a maximum ITD of 0.6 ms. With the time dilation, this gives a maxi-
mum ITD of about 12 ms. Surprisingly, humans can lateralise ITDs out to about 15 ms [23]. The jnd to broadband noise, for these supra-ecological ITDs plateus as one increases reference ITD at a value of about 400 us, up to 3ms, the largest value that has been measured. The jnd for humans within the ecological range is about 20 us. This suggests that the time dilation will not give much improvement in ITD discrimination, however with training or careful microphone positioning this may change. For discriminating along the up-down axis we have the pinna. Humans can learn to adapt to new pinna shapes rapidly, and remember them for a long period of time [24, 25].

1.3. Description of the present device

In this paper we present a device, which we refer to as the Sonic Eye, that facilitates the use of ultrasonic echo information by humans. A forehead-mounted speaker emits ultrasonic chirps (FM sweeps) modeled after bat echolocation calls. The echoes are recorded by ultrasonic microphones on each side of the head. Each microphone is mounted inside an artificial pinna, also modeled after bat pinnae i.e., optimized to produce spectrally based spatial cues. The recorded waveforms are then played back to the user at 1/20 of normal speed. This magnifies all temporally based cues by a factor of 20 and lowers frequencies into the human audible range.

The minimal processing of the signal and the use of artificial pinnae distinguish our approach from previous work.

2. SPECIFICATIONS

The flow of information through the Sonic Eye is illustrated in Figure 1, and the device is pictured in Figure 6. Recordings of a sound waveform moving through the system are presented in Figure ???. A video including helmet-cam video of the device experience is included in Supplemental Material.

The majority of the components are mounted within a baby carrier backpack, which provides ventilation and a solid frame. A tweeter speaker and two ultrasonic microphones in artificial pinnae are head-mounted on a bicycle helmet. The pinna are hand moulded from clay to resemble bat ears.

All signal processing is performed by a a small enclosure mini-ITX computer running Windows 7 and a custom MATLAB program. An ESI Juli@ soundcard with stereo 192 kHz input and output is used to drive the ultrasonic speaker and receive input from B&K ultrasonic microphones. A lithium-ion wheelchair battery is used to power the equipment.

Step 1: The computer generates a chirp waveform, consisting of a 3 ms sweep from 25 kHz to 50 kHz with a constant sweep rate in log frequency. The initial and final 0.3 ms are tapered using a cosine windowing function.

Step 2: The chirp is played through the head mounted tweeter speaker. In order to play the chirp, it is output through the ESI Juli@ soundcard, amplified using a Lepai TRIPATH TA2020 12 volt stereo amplifier, and finally emitted by a Fostex FT17H Realistic SuperTweeter speaker.

Step 3: The computer records 30 milliseconds of audio through the helmet mounted B&K microphones. This captures the initial chirp, and echoes from objects up to 10 meters away. The signal from the microphones passes through a B&K 2670 preamp followed by a B&K Nexus conditioning Amplifier before being digitized by the ESI Juli@ soundcard.

Step 4: The recorded signal is bandpass filtered from 25 to 50 kHz, and time-dilated by a factor of 20 such that the recorded ultrasonic chirp and echoes now lie between 1,250 Hz and 2,500 Hz.

Step 5: The processed signal is played to the subject through AirDrives open ear headphones, driven by a Gigaport HD USB soundcard.

3. EXPERIMENTAL RESULTS

3.1. Measurement of transfer functions

Angular transfer functions were measured for the ultrasonic speaker and microphone in an anechoic chamber. As shown in Figure 4, the full width half max (FWHM) angle for speaker power was ~50 degrees, and for the microphone ~160 degrees. Power was measured using bandpass white noise between 25 kHz and 50 kHz. An image of the test configuration is supplied in Figure 6.

3.2. Performance testing

3.2.1. Psychophysics of object localization

A psychophysical localization experiment was conducted with one blindfolded sighted user after approximately 6 hours
Fig. 4. Measured transfer functions for ultrasonic microphones and ultrasonic speaker as a function of angle. For the microphone the sensitivity relative to the sensitivity at zero degrees is plotted, for the speaker the emission power relative to the emission power at zero degrees is plotted. A photograph of the experimental setup is given in Figure 6.

of self-guided practice. The task was to localize a plate, approximately 30 cm (16.7) in diameter, held at one of 9 positions relative to the user (see Figure 5). In each of 100 trials, the plate was held at a randomly selected position at a distance of 1 meter, or removed for a 10th "absent" condition. The random positions were chosen using Numpy’s randint procedure. The grid of positions spanned 1 meter on a side, such that the horizontal and vertical offsets from the center position subtended 18.3 degrees. Responses consisted of a verbal report of grid position.

Overall performance was 48% correct, significantly greater than chance performance [stats?]. As illustrated in the confusion matrix of Figure 5b, the correct position was indicated with high probability. For all non-absent trials, 72% of localization judgments were within one horizontal or one vertical position of the veridical target position.

We note that performance on this test likely underestimates the sensitivity achievable by the Sonic Eye for several reasons. For example, the task constraint of keeping the head fixed forward would not apply to a user in a naturalistic context. Additionally, the user we tested is sighted and relatively untrained; blind users would benefit from superior auditory capabilities (e.g., [26, 27]) as well as extensive training. Finally, ongoing development of the prototype continues to improve the quality of the emitted, received, and processed signal.

This result suggests that both vertical and horizontal localization cues (see Section ??) were available to the user.

It is also qualitatively consistent with previous measures of spatial resolution in blind and sighted subjects performing unaided spatial echolocation tasks [11, 28]. While further research is needed to validate such comparisons and, more generally, characterize the behavioral envelope of Sonic Eye-aided echolocation, we consider these preliminary results encouraging.

3.2.2. Psychophysics of texture discrimination

A lego wall was constructed, with one surface completely smooth, and the other surface containing a single protruding lego, but otherwise identical. The helmet was mounted with the speaker 21.6 cm from the wall surface. A photo of the experimental setup is shown in Figure 6. Three subjects attempted to identify the orientation of the wall using only the audio output of the Sonic Eye. 24 trials were performed for each subject. For each trial, the orientation of the lego wall was assigned randomly by use of Numpy’s randint procedure. There were no ‘practice’ trials, but the subject was informed after each trial whether their identification was correct or incorrect. Any initial learning period for this task is thus included in the experimental data. Though more data is required, the results show a trend towards an ability to detect very subtle texture cues. Subject 1 had between 5 and 10 hours self-guided practice, Subject 2 had spent less than 1 hour, and subject 3 had spent less than ten minutes. Additionally, this task was also entirely unpracticed – so the results

Fig. 5. Localization. A subject was asked to identify the position of an approximately 30 cm plastic plate held at one m distance. (a) Illustrates the ten possible configurations of the plate, including nine spatial locations and a tenth ‘absent’ condition. (b) Each sub-figure corresponds to a location of the plate, and the intensity map within each subfigure indicates the fraction of trials the subject identified each position for each plate location. Black corresponds to a subject never indicating a location, and white corresponds to a location always being indicated. Observe that the correct position or an adjacent position is usually given. A photograph of the experimental setup is given in Figure 6.
are also suggestive that skills with the device are easily trans-
ferrable to new environments. The statistical significance of 
the experiment, using a two-tailed binomial test, is given in 
Table 1.

### Table 1. Table caption

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number correct</td>
<td>16</td>
<td>15</td>
<td>12</td>
<td>43</td>
</tr>
<tr>
<td>Number incorrect</td>
<td>8</td>
<td>9</td>
<td>12</td>
<td>29</td>
</tr>
<tr>
<td>p-value</td>
<td>0.15</td>
<td>0.31</td>
<td>1</td>
<td>0.12</td>
</tr>
</tbody>
</table>

3.2.3. **Subjective testing description**

Three subjects have spent greater than an hour engaged in 
self-guided use of the sonic eye. Informal experiments with 
two subjects suggest an ability to navigate through hallways 
while blindfolded. All subjects report a capacity to detect the 
presence of and judge distance to walls and stairs. Two sub-
jects were informally tested on their ability to distinguish be-
tween open and closed fists while blindfolded, and appeared 
able to make the distinction.

The supplemental video provides audio and video from 
the subjects perspective during Sonic Eye use.

4. REFERENCES

[1] JA Thomas, CF Moss, and M Vater, *Echolocation in 


use of reflected self-generated seismic waves for spatial 
orientation in a blind subterranean mammal,” *Journal of 

by bats*, 1995.

[5] AN Popper, “Sound emission and detection by del-
phinids.” *Cetacean behavior: Mechanisms and . . .*, 
1980.

difference discrimination by an echolocating Atlantic 
bottlenose dolphin,” *Journal of Comparative Physiol-

by Human Listeners,” *Annual review of psychology*, 

[8] Pavel Zahorik, “Auditory distance perception in hu-
mans: A summary of past and present research,” *Acta 
Acustica united with Acustica*, vol. 91, no. February 

[9] Michael Supa, Milton Cotzin, and KM Dallenbach, 
“”Facial Vision”: The Perception of Obstacles by the 
Blind,” *The American Journal of Psychology*, vol. 57, 
no. 2, pp. 133–183, 1944.

1962.


[12] Santani Teng, Amrita Puri, and David Whitney, “Ultra-
fine spatial acuity of blind expert human echolocators.,” 
*Experimental brain research*, vol. 216, no. 4, pp. 483–8, 


