A device for human ultrasonic echolocation

Jascha Sohl-Dickstein*, Santani Teng*, Benjamin M. Gaub, Chris C. Rodgers, Crystal Li, Michael R. DeWeese
and Nicol S. Harper

Abstract—*Objective*: **We present a device that combines principles of ultrasonic echolocation and spatial hearing to provide human users with environmental cues that are i) not otherwise available to the human auditory system and ii) richer in object and spatial information than the more heavily processed sonar cues of other assistive devices. The device consists of a wearable headset with an ultrasonic emitter and stereo microphones with affixed artificial pinnae. The goal of this study is to describe the device and evaluate the utility of the echoic information it provides.** *Methods*: **The echoes of ultrasonic pulses were recorded and time-stretched to lower their frequencies into the human auditory range, then played back to the user. We tested performance among naive and experienced sighted volunteers using a set of localization experiments in which the locations of echo-reflective surfaces were judged using these time-stretched echoes.** *Results*: **Naive subjects were able to make laterality and distance judgments, suggesting that the echoes provide innately useful information without prior training. Naive subjects were generally unable to make elevation judgments from recorded echoes. However trained subjects demonstrated an ability to judge elevation as well.** *Conclusion*: **This suggests that the device can be used effectively to examine the environment and that the human auditory system can rapidly adapt to these artificial echolocation cues.** *Significance*: **Interpreting and interacting with the external world constitutes a major challenge for persons who are blind or visually impaired. This device has the potential to aid blind people in interacting with their environment.**

Index Terms—**Echolocation, ultrasonic, blind, assistive device**

I. INTRODUCTION

A. Echolocation in Animals

In environments where vision is ineffective, some animals have evolved echolocation – perception using reflections of self-made sounds. Remarkably, some blind humans are also able to echolocate to an extent, frequently with vocal clicks. However, animals specialized for echolocation typically use much higher sound frequencies for their echolocation, and have specialized capabilities to detect time delays in sounds.

The most sophisticated echolocation abilities are found in microchiropteran bats (microbats) and odontocetes (dolphins and toothed whales). For example, microbats catch insects on the wing in total darkness, and dolphins hunt fish in opaque water. Arguably simpler echolocation is also found in oilbirds, swiftlets, Rousettas megabats, some shrews and tenrecs, and even rats [2]. Evidence suggests microbats form a spatially structured representation of objects in their environment using their echolocation [3].

Microbats use sound frequencies ranging from 25-150 kHz in echolocation, and use several different kinds of echolocation calls [4]. One call – the broadband call or chirp, consisting of a brief tone (< 5 ms) sweeping downward over a wide frequency range – is used for localization at close range. A longer duration call – the narrowband call, named for its narrower frequency range – is used for detection and classification of objects, typically at longer range.

In contrast to microbats, odontocetes use clicks; shorter in duration than bat calls and with sound frequencies up to 200 kHz [5]. Odontocetes may use shorter calls as sound travels ~ 4 times faster in water, whereas bat calls may be longer to have sufficient energy for echolocation in air. Dolphins can even use echolocation to detect features that are unavailable via vision: for example, dolphins can tell visually identical hollow objects apart based on differences in thickness [6].

B. Echolocation in Humans

Humans are not typically considered among the echolocating species. However, some blind persons have demonstrated the use of active echolocation, interpreting reflections from self-generated tongue clicks for such tasks as obstacle detection [7], distance discrimination [8], and object localization [9], [10]. The underpinnings of human echolocation in blind (and sighted) people remain poorly characterized, though some informative cues [11], neural correlates [12], [13], [14], and models [15] have been proposed. While the practice of active echolocation via tongue clicks is not commonly taught, it is recognized as an orientation and mobility method [16], [17]. However, 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. For instance, due to their shorter wavelengths the ultrasonic pulses employed by echolocating animals yield higher spatial resolution, stronger directionality, and higher bandwidth than pulses at human-audible frequencies [18].

An understanding of the cues underpinning human auditory spatial perception is crucial to the design of an artifi-

cial echolocation device. Left-right (laterality) localization of sound sources depends heavily on binaural cues in the form of timing and intensity differences between sounds arriving at the two ears. For elevation and front/back localization, the major cues are direction-dependent spectral transformations of the incoming sound induced by the convoluted shape of the pinna, the visible outer portion of the ear [19]. Auditory distance perception is less well characterized than the other dimensions, though evidence suggests that intensity and the ratio of direct-to-reverberant energy play major roles in distance judgments [20]. Notably, the ability of humans to gauge distance using pulse-echo delays has not been well characterized, though these serve as the primary distance cues for actively echolocating animals [21].

Studies of human hearing suggest that it is very adaptable to altered auditory cues, such as those provided by remapped laterality cues [22] or altered pinna shapes [23], [24]. Additionally, in blind subjects the visual cortex can be recruited to also represent auditory cues [12], [25], further illustrating the plasticity of human auditory processing.

C. The Sonic Eye Device

Here we present a device, referred to as the Sonic Eye, that uses a forehead-mounted speaker to emit ultrasonic "chirps" (FM sweeps) modeled after bat echolocation calls. The echoes are recorded by bilaterally mounted ultrasonic microphones, each mounted inside an artificial pinna, also modeled after bat pinnae to produce direction-dependent spectral cues. After each chirp, the recorded chirp and reflections are played back to the user at $\frac{1}{m}$ of normal speed, where m is an adjustable magnification factor. This magnifies all temporally based cues linearly by a factor of m and lowers frequencies into the human audible range. For empirical results reported here, m is 20 or 25 as indicated. That is, cues that are normally too high or too fast for the listener to use are brought into the usable range simply by replaying them more slowly.

Although a number of electronic travel aids that utilize sonar have been developed (e.g., [26], [27], [28], [29]), none appear to be in common use, and very few provide information other than range-finding or a processed localization cue. For example in [27], distance to a single object is calculated and then mapped to a sound frequency, providing only extremely limited information about the world. The device presented in [26] is the most similar to the Sonic Eye. In [26] ultrasonic downward frequency sweeps are emitted, and then time stretched before presentation to the user. However the signals are time stretched in 2 μs chunks sampled every 100 μs, the overall playback of the echoes is not time stretched, no pinnae are used, the binaural microphones are placed only 2 cm apart, and microphone and transducer fidelity is unknown.

In contrast, the Sonic Eye provides a minimally processed input which, while initially challenging to use, has the capacity to be much more informative and integrate better with the innate human spatial hearing system. The relatively raw echoes contain not just distance information but horizontal location information and also vertical location information (from the pinnae), as well as texture, geometric, and material cues.
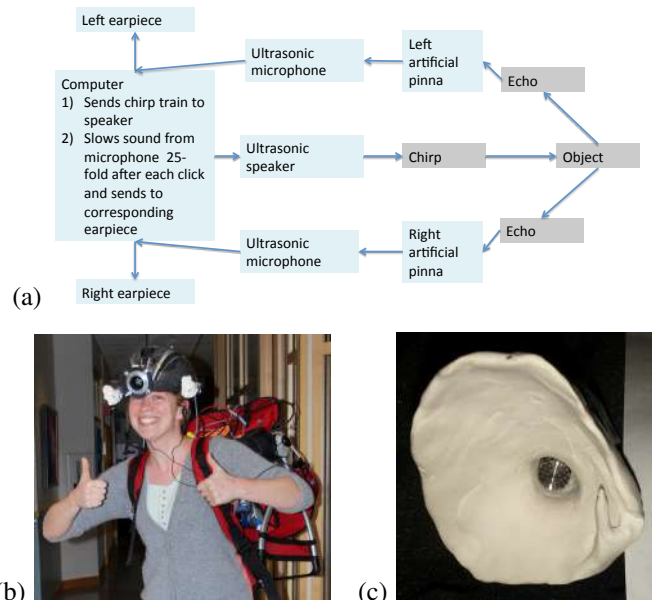

Fig. 1. *(a)* Diagram of components and information flow. *(b)* Photograph of the current hardware. *(c)* Photograph of one of the artificial pinnae used, modeled after a bat ear.
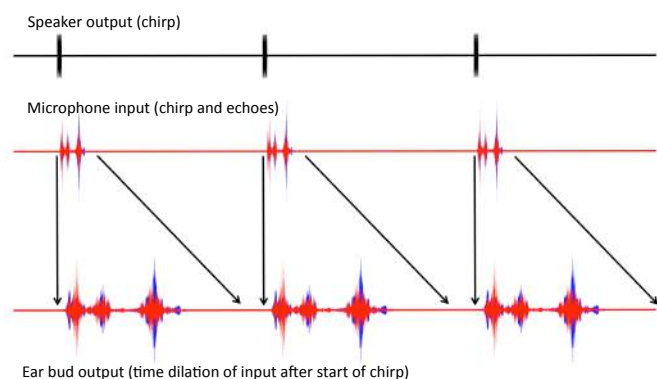

Fig. 2. Schematic of waveforms at several processing stages, from ultrasonic speaker output to stretched pulse-echo signal headphone output presented to user. Red traces correspond to the left ear signal, and blue traces to right ear signal. Note that the relative temporal scales are chosen for ease of visualization, and do not correspond to the temporal scaling used experimentally.

Behavioral testing suggests that novice users can quickly judge the laterality and distance of objects, and with experience can also judge elevation, and that the Sonic Eye thus demonstrates potential as an assistive mobility device.

A sample of audio and video from the Sonic Eye from the user's perspective is provided in the supplemental video to this manuscript, and is also available at http://youtu.be/md-VkLDwYzc.

II. SPECIFICATIONS AND SIGNAL PROCESSING

The flow of information through the Sonic Eye is illustrated in Figure 1a, and the device is pictured in Figure 1b. Recordings of a sound waveform moving through the system are presented in Figure 2. A video including helmet-cam video of the device experience is included in Supplemental Material.

The signal processing steps performed by the Sonic Eye, and the hardware used in each step, are as follows:

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 ramp function. The computer, in a small enclosure mini-ITX case, runs Windows 7 and performs all signal processing using a custom Matlab program.

Step 2: The chirp is played through the head-mounted tweeter speaker. In order to play the chirp, it is output through an ESI Juli@ soundcard with stereo 192 kHz input and output, 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 audio through the helmet mounted B&K Type 4939 microphones. For all experiments, the recording duration was 30 ms, capturing the initial chirp and the resulting echoes from objects up to 5 m 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 using Butterworth filters from 50 to 25 kHz, and time-dilated by a factor of m. For $m = 25$, the recorded ultrasonic chirp and echoes now lie between 1 and 2 kHz.

Step 5: The processed signal is played to the user through AirDrives open-ear headphones, driven by a Gigaport HD USB sound card. Critically, the open-ear design leaves the ear canal unobstructed, ensuring safety in applied situations. (Note that in Experiments 1 and 2 described below, conventional headphones were used for stimulus delivery.)

The chirps are played at a steady rate with a period of approximately 1.5 s. This is a sufficient delay that in all experiments the echoes from the previous chirp have attenuated before the next chirp is played. In the current version of the device, the speaker and two ultrasonic microphones housed in artificial pinnae are mounted on a bicycle helmet. The pinnae are hand-molded from clay to resemble bat ears. The rest of the components are mounted within a baby carrier backpack, which provides portability, ventilation, and a sturdy frame. A lithium-ion wheelchair battery is used to power the equipment. We note that in its current form, the Sonic Eye prototype is a proof-of-principle device whose weight and size make it unsuited to everyday use by blind subjects and extensive open-field navigation testing. To overcome these limitations we are developing a low-cost miniaturized version that retains all the functionality, with a user interface specifically for the blind. However, user testing with the current version has provided a proof of principle of the device's capabilities, as we describe below.

A. Measurement of Transfer Functions

We measured angular transfer functions for the ultrasonic speaker and microphone in an anechoic chamber (Figure 3). The full-width half-max (FWHM) angle for speaker power was $\sim 50°$, and for the microphone was $\sim 160°$. Power was measured using bandpass Gaussian noise between 25 kHz and
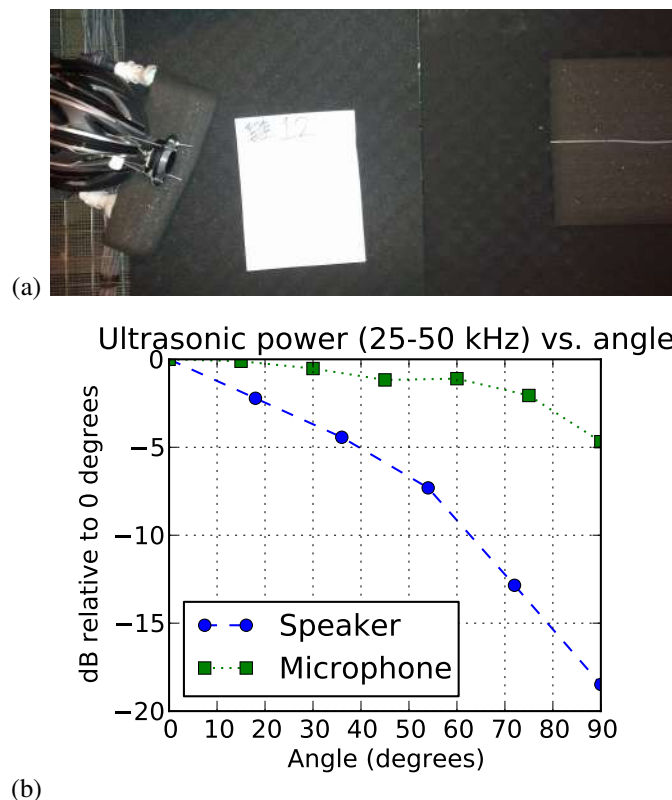
(a)

(b)

Fig. 3. Measurement of transfer functions for ultrasonic microphones and ultrasonic speaker as a function of angle. *(a)* Angular transfer function measurement setup. *(b)* Angular transfer function data. 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.

50 kHz. We expect the FWHM of the speaker and microphone to determine the effective field of view of The Sonic Eye.

III. EXPERIMENTAL METHODS

To explore the perceptual acuity afforded by the artificial echoes, we conducted three behavioral experiments: two in which we presented pulse-echo recordings (from the Sonic Eye) via headphones to naive sighted participants, and a practical localization test with three trained users wearing the device. In both Experiments 1 and 2, we tested spatial discrimination performance in separate two-alternative forced-choice (2AFC) tasks along three dimensions: i) laterality (left-right), ii) depth (near-far), and iii) elevation (high-low). The difference between Experiments 1 and 2 is that we provided trial-by-trial feedback in Experiment 2, but not Experiment 1. This allowed us to assess both the intuitive discriminability of the stimuli (Experiment 1) as well as the benefit provided by feedback (Experiment 2).

In Experiment 3 we tested laterality and elevation localization performance in a separate task on three users each of whom had between four and six hours of total experience wearing the Sonic Eye.

A. Methods, Experiment 1

1) Stimuli: For each of the three spatial discrimination tasks (laterality, depth and elevation), echoes were recorded from

an 18-cm-diameter plastic disc placed in positions appropriate to the stimulus condition, and with the plate face normal to the emitter's line of sight, as illustrated in Figure 4a. For laterality judgments, the disc was suspended from the testing room ceiling via a thin (< 1 cm thick) wooden rod 148 cm in front of the emitter and 23.5 cm to the left or right of the midline. The "left" and "right" conditions were thus each $\sim 9°$ from the midline relative to the emitter, with a center-to-center separation of $\sim 18°$. For depth judgments, the disc was suspended on the midline directly facing the emitter at a distance of 117 or 164 cm, separating the "near" and "far" conditions by 47 cm. Finally, for elevation judgments, the disc was suspended 142 cm in front and 20 cm above or below the midline, such that the "high" and "low" conditions were $\sim 8°$ above and below the horizontal median plane, respectively, separated by $\sim 16°$. In all cases, the helmet with microphones and speakers was mounted on a Styrofoam dummy head.

To reduce the impact of any artifactual cues from a single echo recording, we recorded five "chirp" pulses (3-ms rising frequency sweeps, time dilation factor $m = 25$, truncated to 1 s length) and the corresponding echoes from the disc for each stimulus position (pulse-echo exemplars). Additionally, pulse-echo exemplars from each stimulus position were recorded with and without the artificial pinnae attached to the microphones. Thus, for each of the six stimulus positions, we had 10 recorded pulse-echo exemplars, for a total of 60 stimuli.

2) Procedure: Sighted participants ($N = 13$, 4 female, mean age 25.5 y) underwent 20 trials for each of the three spatial discrimination tasks, for a total of 60 trials per session. The trials were shuffled such that the tasks were randomly interleaved. Sound stimuli were presented on a desktop or laptop PC using closed-back circumaural headphones (Sennheiser HD202) at a comfortable volume, ~ 70 dB SPL. Assessment of these headphones using modified Sennheiser KE4-211-2 microphones (from AuSIM) in the ear canal showed at least ~ 30dB attenuation (no distinguishable sound above the microphone noise floor) at one ear when a 70dB SPL 1-2kHz passband noise was played through the headphone speaker at the other ear. Thus there was negligible sound transfer between the ears. No visual stimuli were presented; the screen remained a neutral gray during auditory stimulus presentation. On each trial, the participant listened to a set of three randomly selected 1-s exemplars (pulse-echo recordings) for each of two stimulus conditions. Depending on the spatial task, the participant then followed on-screen instructions to select from two options; whether the second exemplar represented an object to the left or right; nearer or farther; or above or below relative to the echoic object from the first exemplar. Upon the participant's response, a new trial began immediately, without feedback.

B. Methods, Experiment 2

1) Stimuli: Stimuli in Experiment 2 were nearly identical to those in Experiment 1, except that we now provided trial-by-trial feedback. To prevent participants from improving their performance based on artifactual noise that might be present in our specific stimulus set, we filtered background noise from the original recordings using the spectral noise gating function in the program Audacity (Audacity Team, http://audacity.sourceforge.net/). All other stimulus characteristics remained as in Experiment 1.

2) Procedure: Sighted volunteers ($N = 12$, 5 female, mean age 23.3 y) were tested on the same spatial discrimination tasks as in Experiment 1. After each response, participants were informed whether they had answered correctly or incorrectly. All other attributes of the testing remained the same as in Experiment 1.

C. Methods, Experiment 3

We conducted a psychophysical localization experiment with three sighted users (all male, mean age 33.7 y), who had between four and six hours of self-guided practice in using the device, largely to navigate the corridors near the laboratory. The participants were blindfolded throughout the experiment, and they wore the Sonic Eye device. The task was to localize a plate, ~ 30 cm ($17°$) in diameter, held at one of 9 positions relative to the user (see Figure 5), with the face of the plate oriented to be approximately normal to the emitter's line of sight. In each of 100 trials, the plate (on a long thin pole) was held at a randomly selected position at a distance of 1 m, or removed for a 10th "absent" condition. Each of the 10 conditions was selected with equal probability. The grid of positions spanned 1 m on a side, such that the horizontal and vertical offsets from the center position subtended $\sim 18°$. The subjects stood still and initially fixated centrally, but were able to move their head during the task (although Subject 1 kept their head motionless). Responses consisted of a verbal report of grid position. After each response the participant was given feedback on the true position. The experiment took place in a furnished seminar room, a cluttered echoic space.

The hardware configuration for Subjects 2 and 3 was identical to that in Experiments 1 and 2. Subject 1 used an earlier hardware configuration, which differed as follows. The output of the B&K Nexus conditioning amplifier was fed into a NIDAQ USB-9201 acquisition device for digitization. Ultrasonic audio was output using an ESI GIGAPORT HD sound card. The temporal magnification factor m was set to 20. The backpack used was different, and power was provided by extension cord. Subject 1 did not participate in Experiments 1 or 2, although Subjects 2 and 3 did.

IV. EXPERIMENTAL RESULTS

A. Results, Experiment 1

Laterality judgments were robustly above chance for pinna (mean 75.4% correct, $p < 0.001$, $n = 130$, two-tailed binomial test, Bonferroni-Holm multiple comparison correction over 6 tests) and no-pinna conditions (mean 86.2% correct, $p < 0.001$, $n = 130$, two-tailed binomial test, Bonferroni-Holm multiple comparison correction over 6 tests), indicating that the binaural echo input produced reliable, intuitive cues for left-right judgments. Depth and elevation judgments, however, proved more difficult; performance on both tasks was not different from chance for the group. The presence or absence of the artificial pinnae did not significantly affect performance in any of the three tasks: logistic regression results were
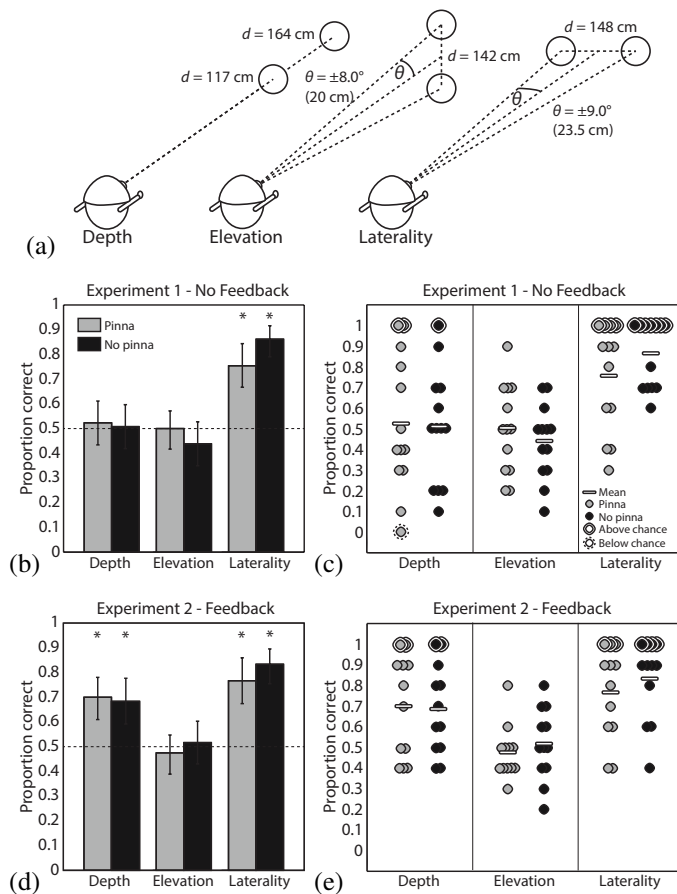
Fig. 4. Two alternative forced choice spatial localization testing. *(a)* A diagram of the configurations used to generate stimuli for each of the depth, elevation, and laterality tasks. *(b)* The fraction of stimuli correctly classified with no feedback provided to subjects $(N = 13)$. Light gray bars indicate results for stimuli recorded with artificial pinnae, while dark gray indicates that pinnae were absent. The dotted line indicates chance performance level. Error bars represent 95% confidence intervals, computed using Matlab's binofit function. Asterisks indicate significant differences from 50% according to a two-tailed binomial test, with Bonferroni-Holm correction for multiple comparisons. *(c)* The same data as in (b), but with each circle representing the performance of a single subject, and significance on a two-tailed binomial test determined after Bonferroni-Holm correction over 13 subjects. *(d)* and *(e)* The same as in (b) and (c), except that after each trial feedback was provided on whether the correct answer was given $(N = 12)$.

nonsignificant for the effect of pinnae $(p = 0.193)$ and the pinna/task interaction $(p = 0.125)$. Population and single-subject results are shown in Figure 4b-c.

B. Results, Experiment 2

Results for laterality and elevation judgments replicated those from Experiment 1: strong above-chance performance for laterality in both pinna (76.7% correct, $p < 0.001$, $n = 120$, two-tailed binomial test, Bonferroni-Holm multiple comparison correction over 6 tests) and no-pinna (83.3% correct, $p < 0.001$, $n = 120$, two-tailed binomial test, Bonferroni-Holm multiple comparison correction over 6 tests) conditions. Because there appeared to be little benefit from feedback for these judgments, we conclude that it may be unnecessary for laterality judgments. Performance was still at chance for elevation, indicating that feedback over the course of a single experimental session was insufficient for this task.

However, performance on depth judgments improved markedly over Experiment 1, with group performance above chance for both pinna (70% correct, $p < 0.001$, $n = 120$, two-tailed binomial test, Bonferroni-Holm multiple comparison correction over 6 tests) and no-pinna (68.3% correct, $p < 0.001$, $n = 120$, two-tailed binomial test, Bonferroni-Holm multiple comparison correction over 6 tests) conditions. Performance ranges were also lower (smaller variance) for depth judgments compared to Experiment 1, suggesting that feedback aided a more consistent interpretation of depth cues. As in Experiment 1, the presence or absence of the artificial pinnae did not significantly affect performance in any of the three tasks: logistic regression results were nonsignificant for the effect of pinnae $(p = 0.538)$ and the pinna/task interaction $(p = 0.303)$. Population and single subject results are shown in Figure 4d-e.

C. Results, Experiment 3

The subjects typically performed well above chance in determining the exact position of the plate from 10 positions, the plate's absence/presence, its horizontal position (laterality), and its vertical position (elevation). This is illustrated in Figure 5b, the dotted line indicating chance performance, and a ringed gray dot indicating the subject performed significantly better than chance by the binomial test.

For Subject 1, the spatially arranged confusion matrix of Figure 5c indicates that the subject reported the exact correct position from the 10 positions with high probability. Overall performance was 48% correct, significantly greater than a chance performance of 10% ($p \ll 0.001$, $n = 100$, two-tailed binomial test, Bonferroni-Holm multiple comparison correction over all tests and subjects of Experiment 3). For all non-absent trials, 72% of localization judgments were within one horizontal or one vertical position of the true target position. Figure 5d shows the confusion matrix collapsed over spatial position to show only the absence or presence of the plate. The present/absent state was reported with 98% accuracy, significantly better than chance ($p < 0.001$, $n = 100$, two-tailed binomial test, Bonferroni-Holm corrected). Figure 5e shows the confusion matrix collapsed over the vertical dimension (for the 93 cases where the plate was present), thus showing how well the subject estimated horizontal position in the horizontal dimension. The horizontal position of the plate was correctly reported 56% of the time, significantly above chance performance ($p \ll 0.001$, $n = 93$, two-tailed binomial test, Bonferroni-Holm corrected). Figure 5f shows the confusion matrix collapsed over the horizontal dimension, thus showing how well the subject estimated position in the vertical dimension. The vertical position of the plate was correctly reported 68% of the time, significantly above chance performance ($p \ll 0.001$, $n = 93$, two-tailed binomial test, Bonferroni-Holm corrected).

The remaining two subjects showed similar results to Subject 1. Subject 2 was significantly above chance for exact position, for absent vs. present, for horizontal localization, and for vertical localization (respectively 44%, $p \ll 0.001$, $n = 100$; 94%, $p = 0.0084$, $n = 100$; 69%, $p \ll 0.001$,

$n = 90$; 49%, $p \ll 0.001$, $n = 90$, two-tailed binomial test, Bonferonni-Holm corrected). Subject 3 was significantly above chance for exact position, for absent vs. present, and for horizontal localization, but not for vertical localization (respectively 26%, $p \ll 0.001$, $n = 100$; 95%, $p = 0.0084$, $n = 100$; 61%, $p \ll 0.001$, $n = 93$; 34%, $p = 0.25$, $n = 93$, two-tailed binomial test, Bonferroni-Holm corrected). Figure 5g-n shows the confusion matrices for Subjects 2 and 3.

V. DISCUSSION

In Experiments 1 and 2, we found that relatively precise spatial discrimination based on echolocation is possible with little or no practice in at least two of three spatial dimensions. Echoic laterality cues were clear and intuitive regardless of feedback, and likely made use of interaural level and/or time differences of the individual echoes. Echoic distance cues were also readily discriminable with feedback. Depth judgments without feedback were characterized by very large variability compared to the other tasks: performance ranged from 0 - 100% (pinna) and 10 - 100% (no-pinna) across subjects. This suggests the presence of a cue that was discriminable but nonintuitive without trial-by-trial feedback.

While we did not vary distances parametrically, as would be necessary to estimate psychophysical thresholds, our results permit some tentative observations about the three-dimensional spatial resolution achieved with artificial echolocation. Direct comparison with previous work on non-ultrasonic human echolocation is difficult; e.g., [30] tested absolute rather than relative laterality, did not alter the echo stimuli, and included a third "center" condition. However, the reflecting object in that experiment was a large rectangular board subtending $29° \times 30°$, such that lateral center-to-center separation was $29°$, whereas disc positions in the present study were separated by a smaller amount, $19.9°$, and presented less than 10% of the reflecting surface. [30] reported ~ 60-65% correct laterality judgments for sighted subjects which is somewhat less than our measures (they reported $\sim 75\%$ for blind subjects). Another study [31] reported a threshold of $6.7°$ azimuth in a virtual echo discrimination task; the reflecting surfaces in that study were virtualized and presented at radially normal angles to the listener. Using flat circular stimuli similar to those reported in the current study, [10] reported horizontal discrimination thresholds averaging $\sim 3.5°$ in a relative localization task among blind trained echolocation experts, but sighted subjects varied widely in performance and were, as a group, unable to perform the task. Prior results such as these suggest an increase in effective sensitivity when using artificial ultrasonic echo cues, but also hint at considerable potential for threshold improvement with larger surfaces, optimized reflection angles, or subject expertise.

Depth judgments were reliably made at a depth difference of 47 cm in Experiment 2, corresponding to an unadjusted echo-delay difference of ~ 2.8 ms, or ~ 69 ms with a dilation factor of 25. A 69-ms time delay is discriminable by humans but was only interpreted correctly with feedback, suggesting that the distance information in the echo recordings, although initially nonintuitive, became readily interpretable with practice.

Our signal recordings included complex reverberations inherent in an ecological, naturalistic environment. Thus the discrimination task was more complex than a simple delay between two isolated sounds. The cues indexing auditory depth include not only variation in pulse-echo timing delays, but also differences in overall reflected energy and reverberance which are strongly distance-dependent. In fact, as cues produced by active echoes, discrete pulse-echo delays are not typically encountered by the human auditory system. Single-subject echoic distance discrimination thresholds as low as ~ 11 cm [8] (~ 30 cm in extensively trained sighted subjects [32]) have been reported for natural human echolocation. Thus, it is likely that training would improve depth discrimination considerably, especially with time-dilated echo information, in theory down to ~ 0.5 cm with 25-fold dilation.

Performance was low on the elevation task in both pinna and no-pinna conditions. It is possible that the echo recordings do not contain the elevation information necessary for judgments of $16°$ precision. However, our tasks were expressly designed to assess rapid, intuitive use of the echo cues provided, while the spectral cues from new pinnae take time to learn; elevation judgments in humans depend strongly on pinna shape [19], [33], and recovering performance after modifying the pinna can take weeks [23]. Vertical localization behavior in bats depends on direction dependent filtering due details of pinna and tragus shape and position [34], [35], and also on active outer ear position adjustments [36]. Thus, the design and construction of the artificial pinnae used in the present experiment may not provide the full benefits of their bat counterparts, and would likely benefit from refinement to optimize their filtering properties. For instance, pinnae could be optimized to maximize the learnability of the new pinna transform by humans. Considering left-right localization, the time dilation employed by the Sonic Eye, by expanding the interaural time differences, may improve interaural time discrimination in some cases, possibly allowing for supernormal laterality localization with practice, especially near the midline. For peripheral sound sources the time dilation will cause unecologically large interaural time differences which although discriminable, tend to be harder to discriminate by a degree that approximately counteracts the advantage of the time-dilation [37]. We do not expect time-dilation to strongly influence vertical localization capacities in our setup.

In line with these observations, Experiment 3 suggests that both laterality and elevation localization cues were available to a user with a moderate amount of training. This is qualitatively consistent with previous measures of spatial resolution in blind and sighted subjects performing unaided spatial echolocation tasks [9], [38]. While further research is needed to validate such comparisons and, more generally, characterize the behavioral envelope of Sonic Eye-aided echolocation, we consider the results presented here as encouraging. Specifically, they suggest that performance on behaviorally relevant tasks is amenable to training. Informal observations with two further participants suggest an ability to navigate through hallways, detecting walls and stairs, while using the Sonic Eye blindfolded. A degree of shape discrimination may also be present (for example an open vs. closed hand), consistent with

[39], who demonstrated human object discrimination using downsampled ultrasonic recordings of dolphins' reflected click trains, and with [12], in which blind humans discriminated echoically between 3-D shapes.

Any practical configuration such as that tested in Experiment 3 should minimize interference between echolocation signals and environmental sounds (e.g., speech or approaching vehicles). To this end, open-ear headphones ensure that the ear remains unobstructed, as described in Section 2. However, future testing should include evaluations of auditory performance with and without the device, and training designed to assess and improve artificial echolocation in a naturalistic, acoustically noisy environment.

We note that performance on the experiments reported here likely underestimates the sensitivity achievable by using the device for several reasons. First, in Experiments 1 and 2, the head was virtually fixed relative to the target object (due to the headphone presentation of recorded echoes). This would not apply to a user in a more naturalistic context. Second, we assessed the intuitive and immediately usable perceptual information in the echoes, while extensive training would only build on that baseline. Third, the participants tested were not just untrained, but normally sighted. Blind and visually impaired users may differ in performance from sighted users due to some combination of superior auditory capabilities [40], [41], [42] and reported deficits, e.g. [43]. Testing this device with blind subjects will be an important direction for future work. Finally, ongoing development of the prototype continues to improve the quality of the emitted, received, and processed signal and its interface.

VI. SUMMARY AND CONCLUSION

Here we present a prototype assistive device to aid in navigation and object perception via ultrasonic echolocation. The ultrasonic signals exploit the advantages of high-frequency sonar signals and time-stretch them into human-audible frequencies. Depth information is encoded in pulse-echo time delays, made available through the time-stretching process. Azimuthal location information is encoded as inter-aural time and intensity differences between echoes recorded by the stereo microphones. Finally, elevation information is captured by artificial pinnae mounted to the microphones as direction-dependent spectral filters. Thus, the device presents a three-dimensional auditory scene to the user with high theoretical spatial resolution, in a form consistent with natural spatial hearing. Behavioral results from two experiments with naive sighted volunteers demonstrated that two of three spatial dimensions (depth and laterality) were readily available with no more than one session of feedback/training. Elevation information proved more difficult to judge, but a third experiment with moderately trained users indicated successful use of elevation information as well. Taken together, we interpret these results to suggest that while some echoic cues provided by the device are immediately and intuitively available to users, perceptual acuity is potentially highly amenable to training. Thus, the Sonic Eye may prove to be a useful assistive device for persons who are blind or visually impaired.

REFERENCES

[1] S. Teng *et al.*, "A device for human ultrasonic echolocation," *IEEE Workshop on Multimodal and Alternative Perception for Visually Impaired People, ICME*, 2013.

[2] G. Jones, "Echolocation." *Curr. Biol.*, vol. 15, no. 13, pp. R484–R488, Jul. 2005.

[3] C. M. DeLong *et al.*, "Evidence for spatial representation of object shape by echolocating bats (eptesicus fuscus)," *J. Acoust. Soc. Am.*, vol. 123, no. 6, pp. 4582–4598, Jun 2008.

[4] A. Grinnell, "Hearing in bats: an overview," in *Hearing by Bats*, A. Popper and R. Fay, Eds. New York: Springer, 1995, vol. 3, pp. 1–36.

[5] "Sound emission and detection by delphinids," in *Cetacean behavior: Mechanisms and functions*, L. Herman, Ed. New York: J. Wiley and Sons, 1980, pp. 1–51.

[6] W. Au and D. Pawloski, "Cylinder wall thickness difference discrimination by an echolocating atlantic bottlenose dolphin," *J. Comp. Physiol. A*, vol. 170, pp. 41–47, Jan 1992.

[7] M. Supa *et al.*, "'facial vision': The perception of obstacles by the blind," *Am. J. Psychol.*, vol. 57, no. 2, pp. 133–183, Apr. 1944.

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

[9] C. Rice, "Human echo perception," *Science*, vol. 155, no. 3763, pp. 656–664, Feb. 1967.

[10] S. Teng *et al.*, "Ultrafine spatial acuity of blind expert human echolocators," *Exp Brain Res*, vol. 216, no. 4, pp. 483–488, Feb. 2012.

[11] T. Papadopoulos *et al.*, "Identification of auditory cues utilized in human echolocation – objective measurement results," *Biomed. Signal Process. Control*, vol. 6, Jul. 2011.

[12] L. Thaler *et al.*, "Neural correlates of natural human echolocation in early and late blind echolocation experts," *PLoS ONE*, vol. 6, no. 5, p. e20162, Jan. 2011.

[13] ——, "Neural correlates of motion processing through echolocation, source hearing, and vision in blind echolocation experts and sighted echolocation novices," *J. Neurophysiol.*, vol. 111, no. 1, pp. 112–127, 2014.

[14] J. L. Milne *et al.*, "Parahippocampal cortex is involved in material processing via echoes in blind echolocation experts," *Vision Res*, Jul. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.visres.2014.07.004

[15] O. Hoshino and K. Kuroiwa, "A neural network model of the inferior colliculus with modifiable lateral inhibitory synapses for human echolocation," *Biol. Cybern.*, vol. 86, no. 3, pp. 231–240, Mar. 2002.

[16] B. N. Schenkman and G. Jansson, "The detection and localization of objects by the blind with the aid of long-cane tapping sounds," *Hum Factors*, vol. 28, no. 5, pp. 607–618, 1986. [Online]. Available: http://hfs.sagepub.com/content/28/5/607.abstract

[17] J. Brazier, "The benefits of using echolocation to safely navigate through the environment," *Int J Orientat Mobil*, vol. 1, pp. 46–51, 2008.

[18] J. D. Pye, "Why ultrasound?" *Endeavour*, vol. 3, no. 2, pp. 57–62, 1979.

[19] J. Middlebrooks and D. Green, "Sound localization by human listeners," *Ann Rev Psychol*, vol. 42, no. 1, pp. 135–159, 1991.

[20] P. Zahorik *et al.*, "Auditory distance perception in humans: A summary of past and present research," *Acta Acust united with Acust*, vol. 91, pp. 409–420, 2005.

[21] H. Riquimaroux *et al.*, "Cortical computational maps control auditory perception," *Science*, vol. 251, no. 4993, pp. 565–568, Feb 1991.

[22] B. G. Shinn-Cunningham *et al.*, "Adapting to supernormal auditory localization cues. i. bias and resolution," *J. Acoust. Soc. Am.*, vol. 103, no. 6, pp. 3656–66, Jun 1998.

[23] P. Hofman *et al.*, "Relearning sound localization with new ears," *Nat Neurosci*, vol. 1, no. 5, Sept 1998.

[24] A. J. King *et al.*, "The shape of ears to come: dynamic coding of auditory space," *Trends Cogn Sci*, vol. 5, no. 6, pp. 261–270, Jun 2001.

[25] O. Collignon *et al.*, "Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects," *Exp. Brain. Res.*, vol. 192, no. 3, pp. 343–358, 2008.

[26] T. Ifukube *et al.*, "A blind mobility aid modeled after echolocation of bats," *IEEE Trans on Biomed Eng*, vol. 38, no. 5, May 1991.

[27] L. Kay, "Auditory perception of objects by blind persons, using a bioacoustic high resolution air sonar," *J. Acoust. Soc. Am.*, vol. 107, no. 6, Jun. 2000.

[28] P. Mihajlik and M. Guttermuth, "Dsp-based ultrasonic navigation aid for the blind," in *Proc 18th IEEE Instrum Meas Technol Conf*, 2001, pp. 1535–1540.

[29] D. Waters and H. Abulula, "Using bat-modelled sonar as a navigational tool in virtual environments," *Int J Hum-Comp Stud*, vol. 65, pp. 873–886, 2007.

[30] A. Dufour *et al.*, "Enhanced sensitivity to echo cues in blind subjects," *Exp Brain Res*, vol. 165, pp. 515–519, 2005.

[31] L. Wallmeier *et al.*, "Echolocation versus echo suppression in humans," *Proc R Soc B*, vol. 280, no. 1769, p. 20131428, Aug 2013.

[32] S. Schörnich *et al.*, "Discovering your inner bat: Echo–acoustic target ranging in humans," *J. Assoc Res Otolaryngol*, vol. 13, no. 5, pp. 673–682, 2012.

[33] G. Recanzone and M. Sutter, "The biological basis of audition," *Ann Rev Psychol*, vol. 59, pp. 119–42, 2008.

[34] B. D. Lawrence and J. A. Simmons, "Echolocation in bats: the external ear and perception of the vertical positions of targets," *Science*, vol. 218, no. 4571, pp. 481–483, Oct 1982.

[35] C. Chiu and C. F. Moss, "The role of the external ear in vertical sound localization in the free flying bat, eptesicus fuscus," *J. Acoust. Soc. Am.*, vol. 121, no. 4, pp. 2227–2235, Apr 2007.

[36] J. Mogdans *et al.*, "The role of pinna movement for the localization of vertical and horizontal wire obstacles in the greater horseshoe bat, rhinolopus ferrumequinum," *J. Acoust. Soc. Am.*, vol. 84, no. 5, pp. 1676–1679, Nov 1988.

[37] J. E. Mossop and J. F. Culling, "Lateralization of large interaural delays," *J. Acoust. Soc. Am.*, vol. 104, no. 3 Pt 1, pp. 1574–9, Sept 1998.

[38] S. Teng and D. Whitney, "The acuity of echolocation: Spatial resolution in sighted persons compared to the performance of an expert who is blind," *J Vis Impair Blind*, vol. 105, pp. 20–32, Jan 2011.

[39] C. M. DeLong *et al.*, "Human listeners provide insights into echo features used by dolphins (¡ em¿ tursiops truncatus¡/em¿) to discriminate among objects." *J Comp Psychol*, vol. 121, no. 3, p. 306, 2007.

[40] N. Lessard *et al.*, "Early-blind human subjects localize sound sources better than sighted subjects." *Nature*, vol. 395, no. 6699, pp. 278–80, Sept 1998.

[41] B. Röder *et al.*, "Improved auditory spatial tuning in blind humans," *Nature*, vol. 400, pp. 162–166, Jul 1999.

[42] P. Voss *et al.*, "Relevance of spectral cues for auditory spatial processing in the occipital cortex of the blind," *Front Psychol*, vol. 2, Mar 2011.

[43] M. P. Zwiers *et al.*, "A spatial hearing deficit in early-blind humans," *J. Neurosci.*, vol. 21, no. 9, pp. 141–145, 2001.
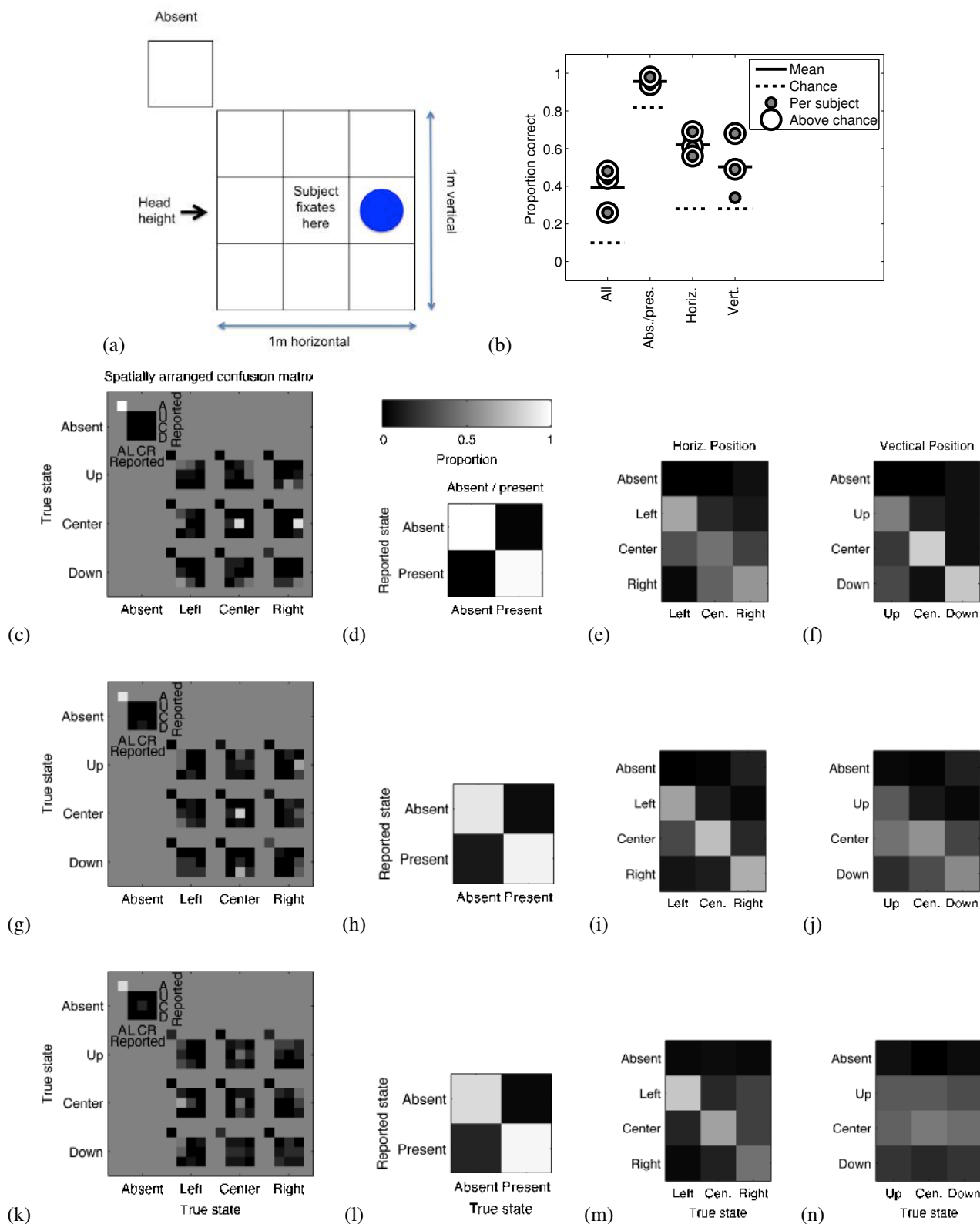
Fig. 5. Ten-position localization in three trained participants. A subject was asked to identify the position of an ∼ 30 cm plastic plate held at 1 m distance. *(a)* Schematic illustration of the 10 possible configurations of the plate, including nine spatial locations and a tenth 'absent' condition. *(b)* Summary of fraction correct for the 3 subjects, for exact identification, and for identification of absent/present, horizontal position, and vertical position. *(c)* Spatially arranged confusion matrix of behavioral results for Subject 1. Each sub-figure corresponds to a location of the plate, and the intensity map within each sub-figure indicates the fraction of trials the subject reported each position for each plate location. Black corresponds to a subject never indicating a location, and white corresponds to a location always being indicated. Each sub-figure sums to 1. *(d)* Confusion matrix grouped into plate absent and present conditions for Subject 1. *(e)* Confusion matrix grouped by horizontal position of the plate for Subject 1. *(f)* Confusion matrix grouped by vertical position of the plate for Subject 1. *(g-j)* Same as in c-f, but for Subject 2. *(k-n)* Same as in c-f, but for Subject 3.