Modeling echo-target acquisition in blind humans

Santani Teng (santani@ski.org)
Smith-Kettlewell Eye Research Institute, 2318 Fillmore St.
San Francisco, CA 94115, USA

Giovanni Fusco (giofusco@ski.org)
Smith-Kettlewell Eye Research Institute, 2318 Fillmore St.
San Francisco, CA 94115, USA

Abstract

Echolocating organisms ensonify their surroundings, then extract object and spatial information from the echoes. This behavior has been observed in some blind humans, but the computations underlying this strategy remain extremely poorly understood. Here we tracked the movements and echo emissions of an expert blind echolocator performing a target detection and localization task. We found that the precision of responses as well as target acquisition movements depended significantly on the size of the target and availability of active echo cues. The distribution of click directions suggested that the maximal energy of each click was always directed at the target. Our results pave the way toward characterizing human echolocation in the context of other active sensing behaviors, constraining the types of perceptual mechanisms mediating its behavior, and at a practical level, building a quantitative evidence base for optimizing therapeutic training interventions.

Keywords: Echolocation; blindness; localization; active sensing

Introduction

In the absence or insufficiency of vision, the critical function of perceiving and interacting with the environment falls to nonvisual sensory modalities. In the auditory domain, many species have developed *active echolocation*: they ensonify their surroundings, then extract object and spatial information from the echoes. In addition to well-known examples such as bats and dolphins, some blind humans use active echolocation as a perceptual method. While ensonification methods vary, the typical oral signal is a sharp palatal or alveolar "click" produced with the tongue.

A growing body of recent as well as historical work on human echolocation comprises reports of psychophysical performance, acoustic signal properties, and neural correlates (Kolarik, Cirstea, Pardhan, & Moore, 2014). However, virtually the entire literature reports performance measures such as spatial acuity (Teng, Puri, & Whitney, 2012) but does not characterize the behavioral (or neural) process by which that performance is achieved. Thus, with few exceptions, it remains almost totally unknown exactly how an echolocator goes about, e.g., detecting and localizing a nearby object. By contrast, extensive scientific attention has been directed to humans' *visual* exploration of a scene via eye movements (Yang, Lengyel, & Wolpert, 2016; Yang, Wolpert, & Lengyel, 2016), and echolocation behavior in nonhuman species such as bats (Yovel, Falk, Moss, & Ulanovsky, 2010). Characterizing human echolocation in this way would place it in the context of other active sensing behaviors, constrain the types of perceptual mechanisms mediating its behavior, and at a practical level, could serve as a basis for optimizing therapeutic training interventions.

To this end, here we measured the time course of motor behavior (head movements and click emissions), the subsequent sensory input (auditory echo returns), and the resultant performance (azimuthal response accuracy and precision) as an experienced blind echolocation practitioner performed an echoic target acquisition and localization task. We show that both performance accuracy as well as target acquisition (head movement) behavior are affected by the size of the target and the availability of actively generated click information, and that human echo-acquisition strategy is likely different from that of clicking bats.

Methods

Participant

The participant, EB, is a male, highly proficient daily echolocation practitioner, age 53 y, completely blind since infancy. EB provided informed consent in accordance with protocols approved by the Smith-Kettlewell Institutional Review Board.

Apparatus, Stimuli, and Task

The experiment was conducted in a soundproof, double-walled sound-attenuating booth (IAC Acoustics, see Fig. 1). Using a ceiling-mounted rod revolving around an axis directly above the participant's chair, the experimenter presented a reflecting target at a range of azimuths between -100° and +100° relative to EB's heading (i.e., a region slightly exceeding the subject's frontal hemifield). The target was a rectangle of 1-cm-thick cardboard, covered with aluminum foil, and mea-
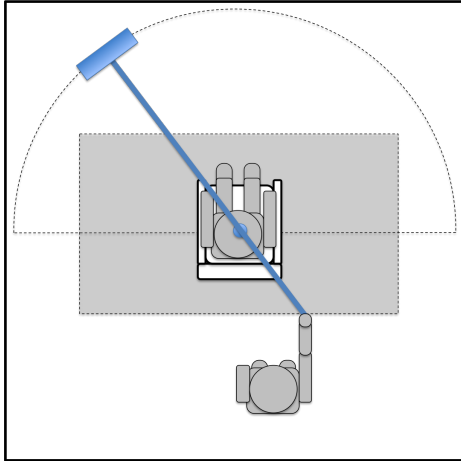
Figure 1: Experimental room and setup. Participant is seated in the center, experimenter standing behind to manipulate stimuli. Rod and attached echo reflector (blue) are mounted to ceiling. Dashed semicircle delineates approximate range of stimulus azimuths. Gray box indicates approximate field of view of video camera (see Methods, Fig. 2).

suring 29 cm wide by 36 cm tall ("Big Target"). Mounted 1 m away from the center of rotation, it subtending 17°azimuthally from the rod's center of rotation. In a more difficult block the target was much smaller, a strip measuring only 2.5x17 cm ("Small Target"), subtending 1.4°azimuthally. Each trial began with the subject pointed straight ahead (0°) with ears covered. Upon trial initiation via shoulder tap from the experimenter, EB began the task of detecting and localizing the reflector as accurately as possible, indicating its position via head orientation and button press. Additionally, we included a control block of trials in which EB had to perform the task without the benefit of any active clicks. To balance experimental control with ecological validity, we did not fix EB's head, but tracked its position and orientation while he remained seated in his chair. In total, EB performed 108 large-target trials, 45 small-target trials, and 27 no-click control trials over three sessions spanning two days.

Data Acquisition and Processing

Sessions were recorded with a video camera (Hero5 Session, GoPro Inc.) at 60 Hz and 1920x1080 resolution, with audio recorded at 48 kHz for click extraction. Additionally, we recorded audio via binaural in-ear microphones and a digital recorder (SP-TFB-2, Sound Professionals; DR-2D, Tascam) at 96 kHz and 24 bits to capture subject-centered click and echo acoustic properties for further analysis.

Head and Target Tracking To monitor head pose over time, we fitted the participant with a cap displaying an ArUco marker, a binary matrix image used for fast detection and pose estimation within the OpenCV framework (Garrido-Jurado, Muñoz-Salinas, Madrid-Cuevas, & Marín-Jiménez, 2014; Bradski, 2000). Additionally, the rod holding the echo reflector was affixed with a red direction indicator that was always within the camera's field of view. To minimize distortions arising from misaligned camera position and rod center of rotation (CoR), we used video from each block to estimate the rod CoR empirically, as shown in 3. Finally, when the subject indicated a response via button press, a green LED was illuminated in a consistent potion of the camera view. Thus, after calibrating video and measuring physical dimensions, each frame of video provided enough information to extract the subject's head position and heading; the ground truth azimuth of the target; the subject-relative azimuth of the target; the azimuth error (how far from target center the subject was pointing); and whether a response had been made (see Fig. 2).

Echoacoustic Tracking To align the subject's clicking and self-motion behavior, we extracted the audio recordings from each trial and analyzed amplitudes within a 1001-sample (48-ms) sliding window. Using the center value of the window as reference, we devised a simple heuristic that flags the signal as a click if less than 5% of the samples within the window are close in amplitude to the reference value. This method allowed to isolate clicks, that are characterized by a sharp peak, from background and environmental noise. Candidate clicks were then manually inspected and verified, and aligned with the timestamps of the head motion data.
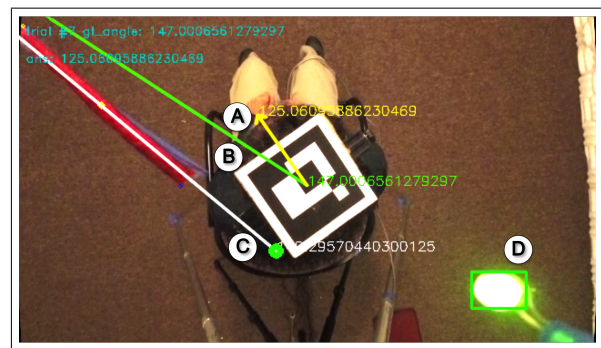

Figure 2: Framewise extraction of experiment parameters from video. A. Yellow arrow: Subject head orientation from head-mounted ArUco marker. B. Green line: Vector from subject head to target center. C. Green dot, white line: Computed center of rotation and radial vector to target center. D. LED indicating subject response.
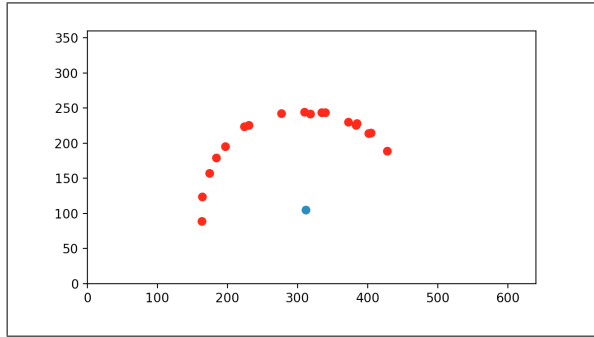
Figure 3: Estimating the Center of Rotation. We extracted the centroid of the rod (red dots) by segmenting it using several frames in which the target was positioned at different orientations. We used these points to fit an ellipse, which allows us to the center of rotation of the rod. A manual rectification of the center then followed to correct noisy segmentation.

Behavioral Data Analysis For each trial, we collected EB's heading and the target heading at the time of button press and then computed an azimuth error. For each condition, we fit a Gaussian curve to the histogram of azimuth response errors using the histfit function in Matlab (The Mathworks, Inc.), extracting average error (μ) and standard deviation (σ) parameters and 95% confidence intervals (CIs) from the curve parameters.

Results

Behavioral accuracy and precision

Fig. 4 shows azimuth response distributions by condition. As expected, EB performed more precisely when actively ensonifying a large target vs a smaller target. Interestingly, the average error in the Big Target condition (6.0°, CI 3.4-8.6) did not differ significantly from that in the Small Target or No-Click conditions (4.5°and 4.1°respectively). However, the precision of responses varied significantly, with standard deviations of 13.6 °(CI 12.0°-15.7°) for the Big Target condition, 32.3°(26.7-40.8) for the Small Target condition, and 45.9°(36.1-62.9) in the No-Click condition.

Head movement time courses

Preliminary analysis suggests that when ensonifying targets with clicks, EB 'homed in' on the target more precisely and quickly, taking less time per trial and traversing a smaller range of azimuths in the process. Fig. 5 displays representative Big-Target and No-Click trials, at the end of which EB indicated a response. The range of relative azimuths after the first direction reversal (to correct for the randomized initial relative position) spans approximately 50°in the Big-Target trial and 116°in the

No-Click trial, and the No-Click trial lasted about twice as long as the Big-Target click trial.

Summary and Conclusions

In this study we recorded and characterized, for the first time, the pattern of target acquisition and localization in a proficient blind human echolocation practitioner. We show that precision of responses was sensitive to the size of both passive (via the available target surface area) and active (via the restriction of clicks) echolocation information, and that these manipulations did not introduce a systematic overall bias into the response distributions.

The term *echolocation* is not monolithic, but rather comprises a family of behaviors and percepts that are observer- and task-dependent. This work represents an initial step toward characterizing not just observable strategies, but likely acoustic and perceptual mechanisms mediating this behavior. For example, Egyptian fruit bats are among the few echolocating bat species to click rather than 'chirp' their echolocation calls, and during a similar localization task, were found to ensonify their target bimodally — slightly off-axis to either side such that the target fell within the region of steepest slope, rather than maximum intensity, of the click energy distribution (Yovel et al., 2010). Human echolocation clicks have been shown to be far less directional, essentially isotropic within a roughly 60°cone (Thaler et al., 2017). In the present study, almost all EB's clicks to the large target were directed within 28°to either side of target center. Thus, our results are not consistent with a maximum-slope localization strategy of the type described by Yovel et al (2010).

Head position and click time courses suggest the multifold accuracy and speed benefit of active ensonification and a robust echo signal, not only improving precision (4), but also reducing time to reported target acquisition, the number of directional reversals, and the range of azimuths traversed in the process (5).

Further analysis of our present data will facilitate a more complete picture of echo-target acquisition in blind expert humans, including an analysis of the loop between head motion, click emission, and the returning binaural echo signal. Additionally, more data collection with other experts, blind non-experts, and sighted control participants will allow us to characterize more fully how echolocation expertise manifests under different conditions and in different stages of training.
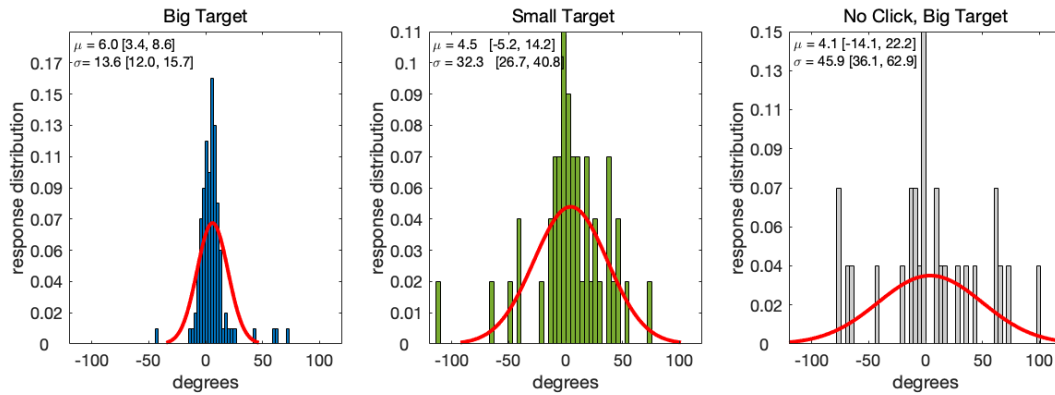
Acknowledgments

Figure 4: Azimuth response distributions by condition. X axes represent headings relative to target azimuth (0°is an exact line to the target center). Parameters μ and σ denote means and standard deviations (in degree units), respectively, of Gaussian fits to response histograms, with 95% confidence intervals in brackets.
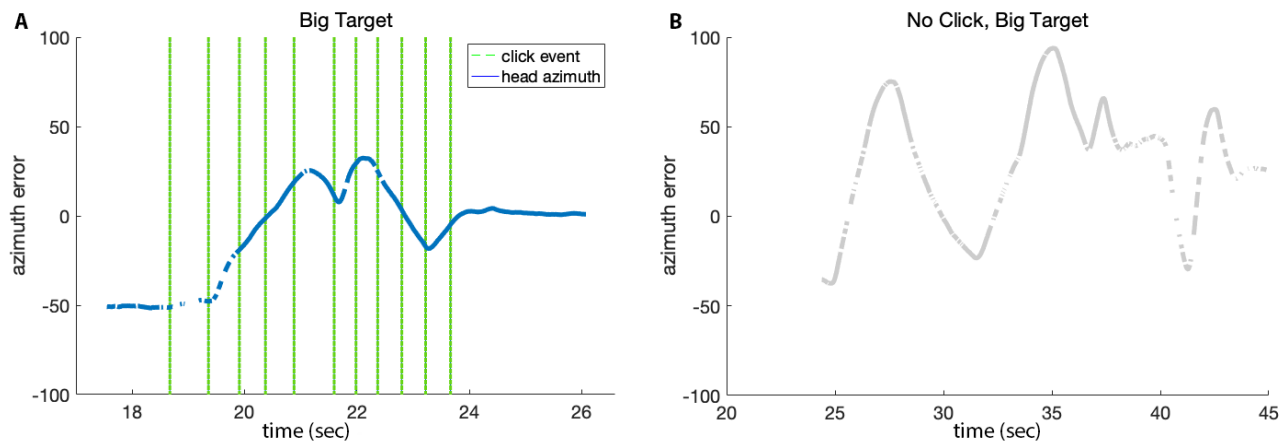

Figure 5: Within-trial head movement time courses A. Representative Big-Target trial. Vertical green lines denote click events. B. Representative No-Click trial. Y axes denote heading in degrees relative to target heading. X axes indicate elapsed time (sec) in current trial block. Discontinuities in the trace are due to the inability to detect the head marker from the video in the presence of partial occlusions.

References

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*.

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

Teng, S., Puri, A., & Whitney, D. (2012). Ultrafine spa-tial acuity of blind expert human echolocators. *Experimental Brain Research*, *216*(4), 483–488.

Thaler, L., Reich, G. M., Zhang, X., Wang, D., Smith, G. E., Tao, Z., . . . Antoniou, M. (2017). Mouth-clicks used by blind expert human echolocators signal description and model based signal synthesis. *PLoS Computational Biology*, 1–17.

Yang, S. C.-H., Lengyel, M., & Wolpert, D. M. (2016). Active sensing in the categorization of visual patterns. *eLife*, *5*, 1–22.

Yang, S. C. H., Wolpert, D. M., & Lengyel, M. (2016). *Theoretical perspectives on active sensing*.

Yovel, Y., Falk, B., Moss, C. F., & Ulanovsky, N. (2010). Optimal localization by pointing off axis. *Science*, *327*, 701–704.