

user's current location, the difference between the time when the reference click is heard and when the corresponding echo is heard will be much larger.

We used the high-resolution acoustics modeling program Odeon, a room acoustic simulation and measurement software, in order to artificially create echoes of clicks based on the room shape. This program generates acoustic cues based on the shape of the maze: for example, the time delay between the emitted sound and the echo provides a reliable cue as to the distance of nearby obstacles. The spatial location of the nearby obstacles is further indicated by the acoustic effects of the head in the path of the echoes. (HRTF 58 was used from the CIPIC database [14] because it was measured from a head with anthropometric feature closest to the average of the CIPIC database and has been validated with naive users in favorable comparison to KEMAR [15]. However, combinations of HRTFs may have different acoustic properties than individual HRTFs and may be worth using in the future [16]).

After creating a set of echoes in Odeon to simulate hallways ending in all the possible junctions (deadend, elbow left, elbow right, t-junction, and stairwells) we boosted the level of all the echoes by the same amount to increase their audibility while keeping the outgoing reference click at its original level. This helped users distinguish between the subtle changes in echoes as they moved through a hallway. After pilot testing, the echoes were boosted 21 dB at the start of the game. Every 15 levels in the game, the echo boost was decreased by 2 dB until users indicated that the game was too difficult to play. This lower limit occurred at or above an echo boost of 7 dB or all users.

Auditory cues were given after each move gesture just to indicate that the motions had been accomplished, but no correct/incorrect feedback was given for individual motions unless that motion caused the avatar to crash into a virtual wall of the maze, in which case, a "crash" sound was played. Implicitly, the absence of a crash sound after a motion indicated that it was a safe move, but nothing indicated whether or not the avatar was heading in the correct direction. However, if the avatar returned all the way to the start of the maze, an auditory warning was played.

3.4 Traditional Psychoacoustic Training

3.4.1 Participants

After pretesting with echolocation, we trained 5 sighted college students in the laboratory for 15 hours with

artificial sounds. All participants were verified to have normal hearing except for one participant who had a mild low-frequency hearing loss in one ear (25 dB HL) from 250 to 1000 Hz; who exhibited normal sensation levels and normal interaural discrimination thresholds using our broadband click stimuli.

3.4.2 Stimuli

Palatal mouth *referent clicks*, recorded by an undergraduate research assistant, were used as a foundation for the stimuli. These clicks were recorded using Roland CS-10EM binaural microphones in an empty IAC sound-attenuating chamber in which the walls and ceiling were covered with 4-in. Auralex foam wedges to reduce echoes. Over 50 clicks were recorded, but only the 17 clicks with waveforms similar to the ideal palatal clicks described by Rojas were used [17]. A custom echo generation program, written in Matlab, was applied to each of these clicks to generate realistic echoes. The echo was generated by adding a copy of the referent click at the appropriate delay corresponding to the distance of the reflecting object. The referent sound travels to the reflecting object and back to the listener as an echo, so it travels twice the distance between the listener and the object. Then, using a speed of sound of 343 m/s, an echo from an object 5 m away, for example, would occur 29 milliseconds ($(2 * 5) \text{ m} / 343 \text{ m/s}$) after the onset of the referent.

When tracking on lateral position, ILDs were implemented in the left and right channels by attenuating one channel according to the angle and distance of the reflecting object. The maximum ILD, used when an object was 90° to the left or right, was 10 dB. ILDs for angles between 0° and 90° were $((1/9) * \text{angle}) \text{ dB}$. The overall level of the echo decreased in both channels by an additional 6 dB each time distance was doubled relative to 1m. During some tracks interaural time delay (ITD) was manipulated instead of ILD. (Those data obtained with ITDs are not presented here due to incomplete data sets.) Determination of the echo levels at which the tasks could be performed at a medium difficulty level for the average person occurred during pretesting. In subsequent staircases, echo levels were reduced as needed to keep thresholds similar over time. Echoes were generated at 10-degree intervals between -90° (90° to the left) and +90° (90° to the right). In addition, echoes were generated at distances between 1 and 5 meters at 0.5-meter increments. The referent click was always presented at the same level, 30 dB SL (sensation level). Within one trial, the reference click was the same for both intervals; the only difference

between the intervals was the timing and level of the echo, calculated relative to the referent click, that accompanied each click. Between trials, clicks were chosen randomly from the 17 clicks described previously.

3.4.3 Procedure

After providing written consent, participants were informed about the structure of the tasks and were given the opportunity to ask questions. They were also instructed not to focus on any one cue in the stimuli and to close their eyes during the experiment. They performed the experiment while seated at a computer in the aforementioned listening chamber. Participants listened to the stimuli through Sennheiser HD 600 headphones.

There were two main tasks: distance and left/right discrimination. All participants performed the distance task first within each of the 15 hour-long test sessions. Each of these tasks was a 2-interval forced choice (2IFC) task in which the order of the intervals was randomized. During the distance task, each interval of a trial contained a referent click, which was centered (no ILDs were applied) and whose channels were normalized relative to each other, followed by an echo generated according to the aforementioned parameters. The angle of echoes in the distance task was kept constant at 0°. Participants discriminated a click with a close echo from a click with a far echo and reported which interval, 1 or 2, contained the closer echo by pressing the corresponding key. During the left/right task, the echo always had a simulated distance of 1 m whereas the angle varied between -90° and +90° at 10-degree increments. In the left/right condition, participants reported whether the echoes in the two intervals moved from right to left or from left to right by pressing the 1 or 2 key, respectively.

During each trial of the 2IFC tasks, participants were shown a sentence reminding them of the correct key presses. (For example, in the distance condition: “Which click contained the closer echo? (1 2)”). A trial consisted of two clicks separated by 500 ms of silence. For example, a trial in the distance condition could contain a click with an echo from 1 m away, followed by 500 ms of silence, followed by a click with an echo from 4.5 m away. Pre-recorded verbal feedback was given (e.g. “correct!”). The next trial was determined using a three-down-one-up staircase paradigm [18]. The staircase paradigm allowed for the determination of a threshold at which participants responded accurately to about 78% of the trials. At the beginning of all three

conditions, the staircases started at the easiest level (distance: 5m, lateralization and left/right: 90 degrees). If participants correctly answered three trials in a row, the subsequent trial increased in difficulty by one level. If participants incorrectly answered a single trial, the subsequent trial decreased in difficulty by one level. The track ended after 11 reversals were observed or if the track lasted for over 70 trials without 11 reversals. Here, reversals are defined to be points during the track where participants answered correctly three times after answering incorrectly on the previous trial, or points where participants answered incorrectly once after answering correctly on the previous trial. Participants performed 2-5 tracks per condition per test session. A condition average for each participant was calculated by computing the average of each of that participant’s tracks. If the track did not contain 6 reversals or if the participant performed more than 3 trials at the easiest level during any given reversal in the last 6 reversals, that track was not included in the participant’s threshold calculation. The staircase adjusted the distance (in depth, or angular distance) between the two intervals of a trial and determined a threshold. After each track, performance was reviewed by an experimenter. If the participant’s performance was good, the level of the echo was decreased in the next track. In this way, the experimenter aimed to keep the threshold relatively constant while gradually decreasing the echo level over time.

4. RESULTS OF GAME PLAY VS. PSYCHOACOUSTIC TRAINING

4.1 Learning during training

Using the EchoExplorer™ game, we measured number of crashes into walls per level, number of echoes requested per level, number of steps taken per level, and active time per level. Because echo level did decrease by 2 dB every 15 levels after the tutorial, in order to look for learning effects, we pulled out performance at a few echo levels to compare the number of crashes per level as training went on. The maze level was cycled through a few times so that we were able to compare performance at similar echo/maze levels over time. Figure 1 shows log fits to the average number of crashes per level as a function of the sequential time of play of each level. Crashes were higher initially for the earliest level with the 21 dB boost because it was the first level encountered but, as expected, the asymptotic

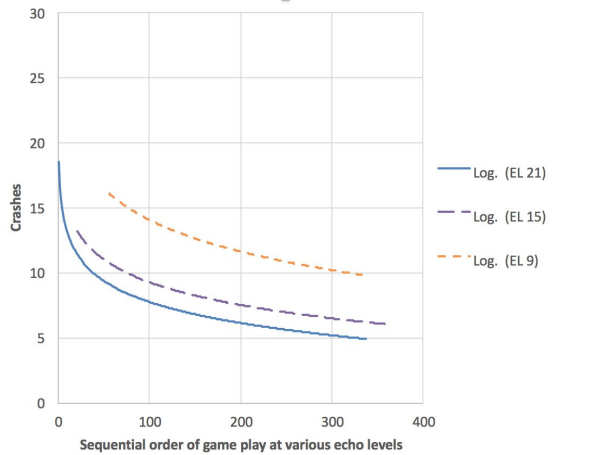


Figure 1. Number of times crashing into maze walls as a function of the sequential order in which the level was completed. The parameter is the level boost of the echo in dB, either 9, 15 or 21 dB. The average data across all App participants are fitted with Log functions.

performance level was best for this condition. Even at the most difficult levels, the echo was still boosted beyond what would typically occur in a real hallway.

There was evidence of learning during the lab training. Across participants in the Lab group, the average simulated distance that supported 78% discrimination from a 1m distance was 4.45 m (SD 0.9). Figure 2 shows the echo level required for discrimination of distance as a function of training hour for four participants. (Varying echo level data were not available for the fifth participant due to a procedural error). Improvement (measured as a decrease in echo level) ranged from 4 to 12 dB over time. Across participants, the average simulated angle difference that supported 78% correct discrimination between right and left lateral positions was 38.4 degrees (SD 9.9). Figure 3 shows the echo level required for discrimination of lateral position (in the left/right task) for all 5 participants in the Lab group. In both graphs, the echo level required decreased over time. The echo was still boosted beyond what would typically occur in natural conditions, but by the end, all participants could reliably discriminate the echo when its level was lower than the source level.

4.2 Improvement in echolocation

Discrimination of lateral position and distance was evaluated in both a pretest and posttest. Dprime was the measure of sensitivity to the different possible locations of the board [19]. The four lateral positions (-90, -45, 45,

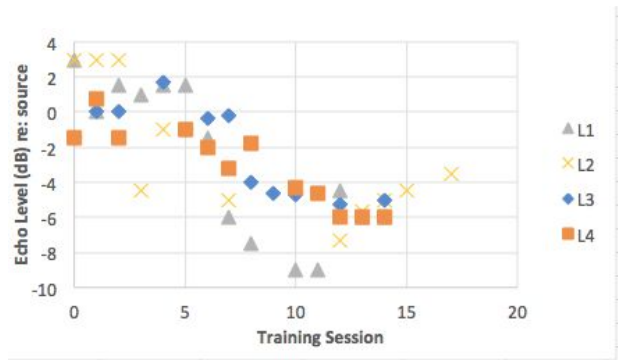


Figure 2. Echo level used relative to the initial outgoing click in order to support average threshold performance on distance discrimination as a function of hours of training in the lab (for individuals in the Lab group).

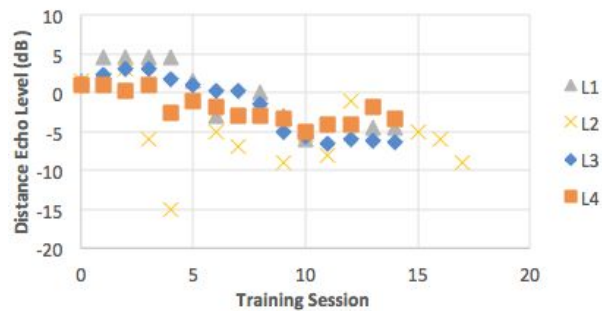
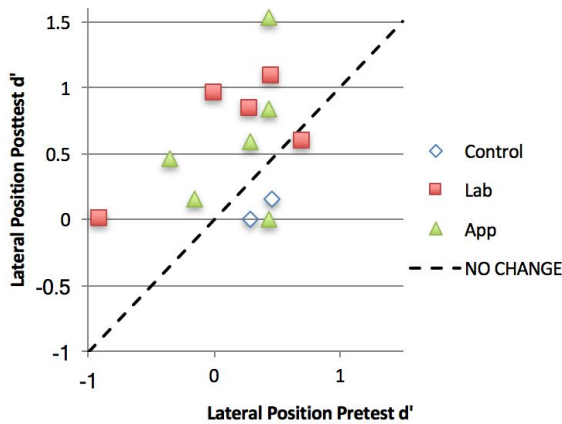


Figure 3. Echo level used relative to the initial outgoing click in order to support average threshold discrimination of left/right lateral position as a function of hours of training in the lab (for the Lab group).

and 90) yielded a chance level of 25% which would be equivalent to a d' of 0 for a four-alternative forced-choice task. The three possible distances (0.9, 1.8 and 2.7 meters) yielded a chance level of 33% with a d' of 0 for a three-alternative forced-choice task.

When head position was fixed, there was modest sensitivity to lateral position with an average d' of 0.10 at pretest and 0.43 at posttest, with 9 of 11 trained participants showing improvement. When the head was moved freely, the average lateral position discrimination was 0.15 at pretest and 0.64 at posttest, with 10 of 11 participants showing improvement. For the head moving condition, post test d' versus pretest d' is shown for all observers in Figure 4. No change in performance would be implied by the dashed line, whereas improvement is indicated by all data points above that line. The average sensitivity was relatively low given that a d' value of 1 is typically considered threshold discrimination, similar to the 78% correct



participants. We saw that learning in the lab using a standard psychoacoustic method was not substantially superior to learning with the app. However, a key advantage to training with the game is that it is under the user's control and can be used at their convenience; it therefore is more accessible and practical than the customized psychoacoustic training method in the lab. It should also be pointed out that the EchoExplorer™ game was tested as a beta version and is not yet optimized. Therefore, we find these experimental results encouraging for the future of using games to learn new ways to use sound for navigation.

6. ACKNOWLEDGMENT

We thank Sarah Kwan, Tejal Kudav, Kiran Matharu, Jacqueline Hon, Jessica Kwon, Aaron Schwartz, Chieko Asakawa, Catherine Getchell, participants at the Blind and Vision Rehabilitation Services of Pittsburgh, and Art Rizzino for helpful discussions, feedback and pointers. We also thank Google Inc., the NSF Center for Science of Information (CSoI), CMU's SURG program, and the NSF REU program for their generous support.

7. REFERENCES

- [1] A. J. Kolarik, S. Cirstea, S. Pardhan, and B. C. J. Moore, "A summary of research investigating echolocation abilities of blind and sighted humans," *Hearing Research*, vol. 310, pp. 60-68, 2014.
- [2] B. F. G. Katz and L. Picinali, "Spatial audio applied to research with the Blind.," in *Advances in Sound Localization*, 2011, pp. 225-250.
- [3] D. Pelegrin-Garcia, M. Rychtáriková, C. Glorieux, and B. F. G. Katz, "Interactive auralization of self-generated oral sounds in virtual acoustic environments for research in human echolocation," in *Proceedings of Forum Acusticum 2014*, 2014.
- [4] H. Wallach, E.B. Newman, and M. R. Rosenweig, "The precedence effect in sound localization," *The American Journal of Psychology*, vol. 62(3), 315-336, 1949.
- [5] K. Saberi, and J. V. Antonio, "Precedence-effect thresholds for a population of untrained listeners as a function of stimulus intensity and interclick interval," *J. Acoustical Soc. Am.*, vol. 114, pp. 420-429, 2003.
- [6] L. Wallmeier, N. Geßele, and L. Wiegrebe, "Echolocation versus echo suppression in humans," *Proceedings of the Royal Society B.*, vol. 280, 20131428, 2013.
- [7] B. C. J. Moore, *An Introduction to the Psychology of Hearing, Sixth Edition*. Bingley, UK: Emerald Group Publishing Limited, 2012.
- [8] S. Teng, and D. Whitney, "The acuity of echolocation: Spatial resolution in the sighted compared to expert performance," *Journal of Visual Impairment and Blindness*, vol. 105(1), pp. 20-32, 2011.
- [9] B. N. Schenkman, and M. E. Nilsson, "Human echolocation: Blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object," *Perception*, vol. 39, pp. 483-501, 2010.
- [10] A. Tonelli, L. Brayda, and M. Gori, "Depth echolocation learnt by novice sighted people," *PLoS One*, vol. 11, no. 6, pp. 1-14, 2016.
- [11] L. Thaler, "Echolocation may have real-life advantages for blind people: an analysis of survey data," *Frontiers in Psychology*, 2013, vol. 4(98), eCollection.
- [12] W. Wu, R. Morina, A. Schenker, A. Gotsis, H. Chivukula, M. Gardner, F. Liu, S. Barton, S. Woyach, B. Sinopoli, P. Grover, and L. M. Heller, "EchoExplorer™: A game app for understanding echolocation and learning to navigate using echo cues," in *ICAD Proceedings 2017*, 2017.
- [13] C. S. Green, D. Bavelier, "Action video game training for cognitive enhancement," *Current Opinion in Behavioral Sciences*, vol. 4, pp. 103-108, 2015.
- [14] R.V. Algazi, R.O. Duda, D.M. Thompson, C. Avendano, The CIPIC HRTF Database, *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics In WASSAP '01* (2001).
- [15] E. De Sena, N. Kaplanis, P.A. Naylor, T. van Waterschoot, "Large-scale auralised sound localisation experiment," *AES 60th International Conference*, 2016.
- [16] L. S. R. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual attributes for the comparison of head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3623-3632, 2016.
- [17] J. A. M. Rojas, J. A. Hermosilla, R. S. Montero, and P. L. L. Espi, "Physical Analysis of Several Organic Signals for Human Echolocation: Oral Vacuum Pulses," *Acta Acustica United with Acustica*, vol. 95, pp. 325-330, 2009.
- [18] M. R. Leek, "Adaptive procedures in psychophysical research," *Perception & Psychophysics*, vol. 63(8), pp. 1279-1292, 2001.
- [19] N.A. Macmillan and C.D. Creelman, *Detection Theory: A Users Guide*, 2nd Ed. Psychology Press, 2004.

