Figures and figure supplements

Augmented reality powers a cognitive assistant for the blind

Yang Liu et al
Figure 1. Hardware platform and object localization task. (A) The Microsoft HoloLens wearable augmented reality device. Arrow points to one of its stereo speakers. (B) In each trial of the object localization task, the target (green box) is randomly placed on a circle (red). The subject localizes and turns to aim at the target. (C) Object localization relative to the true azimuth angle (dashed line). Box denotes s.e.m., whiskers s.d. (D) Characteristics of the seven blind subjects.

DOI: https://doi.org/10.7554/eLife.37841.002
**Figure 1—figure supplement 1.** Obstacle avoidance utility and active scene exploration modes. (A) to (C) An object avoidance system is active in the background at all times. Whenever a real scanned surface or a virtual object enters a danger volume around the user (red in A), a spatialized warning sound is emitted from the point of contact (B). The danger volume expands automatically as the user moves (C), so as to deliver warnings in time. (D) to (E) Active exploration modes. In Scan mode (D) objects whose azimuthal angles fall in a certain range (e.g. between $-60$ and $+60$ deg) call themselves out from left to right. In Spotlight mode (E) only objects within a narrow cone are activated, and the object closest to the forward-facing vector calls out.

DOI: https://doi.org/10.7554/eLife.37841.003
Figure 1—figure supplement 2. Process of scene sonification. The acquisition system should parse the scene (A) into objects and assign each object a name and a voice (B). In our study, this was accomplished by a combination of the HoloLens and the experimenter. The HoloLens scans the physical space (C) and generates a 3D mesh of all surfaces (D). In this digitized space (E) the experimenter can perform manipulations such as placing and labeling virtual objects, computing paths for navigation, and animating virtual guides (F). Because of the correspondence established in D, these virtual labels are tied to the physical objects in real space.

DOI: https://doi.org/10.7554/eLife.37841.004
Figure 2. Spatial memory task. (A) Five objects are arranged on a half-circle; the subject explores the scene, then reports the recalled object identities and locations. (B) Recall performance during blocks 1 (left) and 2 (right). Recalled target angle plotted against true angle. Shaded bar along the diagonal shows the 30 deg width of each object; data points within the bar indicate perfect recall. Dotted lines are linear regressions. (C) Slope and (D) correlation coefficient for the regressions in panel (B).
DOI: https://doi.org/10.7554/eLife.37841.005
Figure 2—figure supplement 1. Mental imagery task supplementary data. Spatial memory data (Figure 2) from blocks 1 (left) and 2 (right) by subject. Shaded areas indicate the true azimuthal extent of each object. Markers indicate recalled location. Most recalled locations overlap with the true extent of the object. Subjects 8–10 were normally sighted and performed the exploration phase using vision.

DOI: https://doi.org/10.7554/eLife.37841.006
Figure 3. Direct navigation task. (A) For each trial, a target chair is randomly placed at one of four locations. The subject begins in the starting zone (red shaded circle), follows the voice of the chair, and navigates to the target zone (green shaded circle). (B) All raw trajectories from one subject (#6) including 1 s time markers. Oscillations from head movement are filtered out in subsequent analysis. (C) Filtered and aligned trajectories from all trials of 3 subjects (#3, 4, 7). Arrow highlights a trial where the subject started in the wrong direction. (D) Trajectories of subjects performing the task with only a cane and no HoloLens. (E) Deviation index, namely the excess length of the walking trajectory relative to the shortest distance between start and target. Note logarithmic axis and dramatic difference between HoloLens and Cane conditions. (F) Speed of each subject normalized to the free-walking speed.

DOI: https://doi.org/10.7554/eLife.37841.007
Figure 3—figure supplement 1. Direct navigation task extended data. Trial distance (A) and trial duration (B) for the first 20 trials of all subjects. A modest effect of practice on task duration can be observed across all subjects (B). (C) Low-pass filtered, aligned trajectories of all subjects. In most trials, subjects reach the target with little deviation. (D) Dynamics of navigation, showing the distance to target as a function of trial time for one subject. (E) Head orientation vs distance to target for two subjects. Note subject six begins by orienting without walking, then walks to the target. Subject two orients and walks at the same time, especially during early trials.

DOI: https://doi.org/10.7554/eLife.37841.008
Figure 3—figure supplement 2. Additional experimental functions. (A) to (B) Automated sign recognition using computer vision. Using Vuforia software (https://www.vuforia.com/) the HoloLens recognizes a men's room sign (A, image viewed through HoloLens) and installs a virtual object (cube, arrow) next to the sign. (B) This object persists in the space even when the sign is no longer visible. (C) Automated wayfinding. The HoloLens generates a path to the target (door) that avoids the obstacle (white box). Then a virtual guide (orange balloon) can lead the user along the path. See Videos 2–3. (D) Navigation in the presence of obstacles. The subject navigates from the starting zone (red circle) to an object in the target zone (green circle) using calls emitted by the object. Three vertical columns block the path (black circles), and the subject must weave between them using the obstacle warning system. Raw trajectories (no filtering) of a blind subject (#5) are shown during outbound (left) and return trips (right), illustrating effective avoidance of the columns. This experiment was performed with a version of the apparatus built around the HTC Vive headset. (E) Orienting functions of the virtual guide. In addition to spatialized voice calls, the virtual guide may also offer turning commands toward the next waypoint. In the illustrated example, the instruction is ‘in x meters, turn right.’ (F) Real-time object detection using YOLO (Redmon and Farhadi, 2018). Left: A real scene. Note even small objects on a textured background are identified efficiently based on a single video frame. Right: A virtual scene from the benchmarking environment, rendered by Unity software.

DOI: https://doi.org/10.7554/eLife.37841.009
**Figure 4.** Long-range guided navigation task. (A) 3D reconstruction of the experimental space with trajectories from all subjects overlaid. (B and C) 2D floor plans with all first trial trajectories overlaid. Trajectories are divided into three segments: lobby (Start – Start 2), stairwell (Start 2 – Start 3), and hallway (Start 3 – Destination). Red arrows indicate significant deviations from the planned path. (D) Deviation index (as in Figure 3E) for all segments by subject. Outlier corresponds to initial error by subject 7. Negative values indicate that the subject cut corners relative to the virtual guide. (E) Duration and (F) normalized speed of all the segments by subject.

DOI: https://doi.org/10.7554/eLife.37841.010
Figure 4—figure supplement 1. Guided navigation trajectories. (A) 3D model of the experimental space as scanned by the HoloLens. (B) Subject and guide trajectories from the long-range guided navigation task. Note small differences between guide trajectories across experimental days, owing to variations in detailed waypoint placement.

DOI: https://doi.org/10.7554/eLife.37841.011
Figure 5. Benchmark testing environment. (A) A virtual living room including 16 pieces of furniture and other objects. (B) Localization of a randomly chosen object relative to the true object location (0 deg, dashed line) for four subjects using CARA (C) or vOICe (V). Box denotes s.e.m., whiskers s.d. For all subjects the locations obtained with vOICe are consistent with a uniform circular distribution (Rayleigh z test, p>0.05). (C) Navigation toward a randomly placed chair. Trajectories from one subject using CARA (left) and vOICe (middle), displayed as in Figure 3C. Right: Number of trials completed and time per trial (mean ± s.d.). (D) Navigation toward a randomly placed key on the floor (small green circle). Trajectories and trial statistics displayed as in panel C.

DOI: https://doi.org/10.7554/eLife.37841.015
Figure 5—figure supplement 1. Benchmark tests in a virtual environment. Trajectories of three additional subjects. (A) Navigation to a randomly placed chair, using either CARA or vOICe, displayed as in Figure 5C. Subject #4 exhibited some directed navigation using vOICe. (B) Finding a dropped key, as in Figure 5D.

DOI: https://doi.org/10.7554/eLife.37841.016