

# Human-centred design and fabrication of a wearable multimodal visual assistance system

Received: 2 July 2024

Accepted: 20 February 2025

Published online: 14 April 2025

 Check for updates

Jian Tang<sup>1,2</sup>, Yi Zhu<sup>1,2</sup>, Gai Jiang<sup>1</sup>, Lin Xiao<sup>1</sup>, Wei Ren<sup>1</sup>, Yu Zhou<sup>3</sup>, Qinying Gu<sup>2</sup>, Biao Yan<sup>4</sup>, Jiayi Zhang<sup>4</sup>, Hengchang Bi<sup>5</sup>, Xing Wu<sup>5</sup>, Zhiyong Fan<sup>2,3</sup> & Leilei Gu<sup>1,2,5,6</sup>✉

Artificial intelligence-powered wearable electronic systems offer promising solutions for non-invasive visual assistance. However, state-of-the-art systems have not sufficiently considered human adaptation, resulting in a low adoption rate among blind people. Here we present a human-centred, multimodal wearable system that advances usability by blending software and hardware innovations. For software, we customize the artificial intelligence algorithm to match the requirements of application scenario and human behaviours. For hardware, we improve the wearability by developing stretchable sensory-motor artificial skins to complement the audio feedback and visual tasks. Self-powered triboelectric smart insoles align real users with virtual avatars, supporting effective training in carefully designed scenarios. The harmonious corporation of visual, audio and haptic senses enables significant improvements in navigation and postnavigation tasks, which are experimentally evidenced by humanoid robots and participants with visual impairment in both virtual and real environments. Postexperiment surveys highlight the system's reliable functionality and high usability. This research paves the way for user-friendly visual assistance systems, offering alternative avenues to enhance the quality of life for people with visual impairment.

Developing artificial vision for blind people has been a long-standing human pursuit<sup>1</sup>. Wearable electronic visual assistance systems offer a promising solution for people with visual impairment or visually impaired people (VIPs), providing an alternative to medical treatments and implanted prostheses. Instead of restoring natural vision, these devices compensate for vision loss by converting environmental information into other sensory modalities and feeding it into the brain to assist with daily tasks. Cutting-edge systems have enabled VIPs to

perform essential visual functions such as navigation, recognition and intricate tasks such as locating misplaced keys<sup>2,3</sup>. Despite their evolving capabilities, these technologies have not seen widespread adoption within the VIP community<sup>4</sup>. The limited uptake can be largely attributed to usability challenges, including the cognitive and physical load in use, as well as the intricate training process required before usage.

The usability of a device hinges primarily on its performance. For visual assistance, this denotes how efficiently it delivers the

<sup>1</sup>Qingyuan Research Institute, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. <sup>2</sup>Shanghai Artificial Intelligence Laboratory, Shanghai, China. <sup>3</sup>Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. <sup>4</sup>State Key Laboratory of Medical Neurobiology, MOE Frontiers Center for Brain Science, Institutes of Brain Science, Department of Ophthalmology and Vision Science, Eye & ENT Hospital, Fudan University, Shanghai, China. <sup>5</sup>In Situ Devices Center, School of Integrated Circuits, East China Normal University, Shanghai, China. <sup>6</sup>National Key Laboratory of Advanced Micro and Nano Manufacture Technology, Shanghai Jiao Tong University, Shanghai, China. ✉e-mail: [leilei.gu@sjtu.edu.cn](mailto:leilei.gu@sjtu.edu.cn)

intended information to the user. This process can be further divided into two components: capturing information and relaying data into brain. Advanced artificial intelligence (AI) algorithms, such as vision foundation models and large language models, have shown promise in addressing the first issue by demonstrating a strong understanding of human perception and the environment<sup>5,6</sup>. The challenge lies in crafting powerful yet compact algorithms that can operate on low-power computing units such as field-programmable gate arrays and microcontroller units (MCUs). The feedback strategy serves as the human-machine interface. Aligning encoding methodologies with the brain's natural information processing mechanisms can streamline training processes and reduce cognitive burden. Particularly for visual assistance, the non-visual modalities have lower spatial bandwidth than vision, hindering information capacity and delivery speed<sup>7,8</sup>. Despite achievements in feedback strategies<sup>9-12</sup>, efficient feedback technology with high biocompatibility is still the bottleneck challenge. Harmonious integration of multiple modalities is crucial for effective information delivery through neuro-interfaces, requiring collaborative efforts in human-centred AI and neuroscience-based encoding methodologies.

Moreover, device usability is heavily influenced by its wearability. Wearable visual assistance systems are intricate systems that amalgamate diverse hardware components. Conventional wearable visual assistance systems frequently encounter challenges related to bulkiness, weight and limited wearability<sup>13,14</sup>. Fortunately, recent strides in sensing materials, device architectures and integration technologies offer hope for enhancing wearability. An array of bio-integrated devices have been devised, including ultrathin artificial skins (A-skins) capable of continuous physiological sensing<sup>15</sup>, artificial throats for speech modulation<sup>16</sup> and artificial eyes mimicking biological vision<sup>17</sup>. These devices are designed to seamlessly interface with the body, ensuring a conformal and stable contact that minimizes discomfort during wear. Moreover, they have exhibited natural and personalized human-machine interaction by providing immersive feedback on the basis of individual threshold maps<sup>18</sup>. These advancements in skin electronics provide sophisticated sensing capabilities, improve wearability and bolster the privacy of VIPs through miniaturization.

In this work, we present a wearable visual assistance system aiming at enhancing the independent life of VIPs. Our focus is on enhancing human-machine fusion by aligning software and hardware innovations. The system customizes visual input into audio and haptic feedback, ensuring high accuracy, low latency and adaptability to diverse environments. It boasts an intuitive spatial cue auditory output and navigation algorithms crafted to reduce cognitive burden. The strategic design and placement of A-skins assists with navigation and postnavigation tasks in harsh environments. Virtual reality (VR) training facilitates rapid and safe training. Triboelectric nanogenerator-based smart insoles synchronize movements across virtual and real worlds with minimal power consumption. Real-world testing by VIPs validates its efficacy in enhancing visual function and usability. This work represents a notable advancement in wearable visual assistance technology, promising broader adoption and an improved quality of life for VIPs.

## Results

### Overall system design

The wearable visual assistance system features an RGB-D (red, green, blue and depth) camera, artificial skins and a VR training platform interfaced via triboelectric smart insoles. It uses intuitive multimodal feedback to assist VIPs with daily tasks and provides VR training (Fig. 1a). As suggested by Fig. 1b, VIPs set a target via voice commands. AI algorithms process RGB-D data to estimate the target's orientation and determine an obstacle-free direction in real time. Bone conduction earphones deliver spatialized cues to guide users. The A-skins, worn on the wrists, enhance navigation by detecting lateral obstacles, effectively

reducing collision risks in cluttered environment. The system updates the three-dimensional (3D) scene in real time as the VIP approaches to the target step-by-step while avoiding obstacles in path. The A-skins also assist with postnavigation tasks by controlling the hand-target distance, improving eye-hand coordination and thus task execution accuracy. Before moving to the real world, the VIPs are trained in a virtual environment to master the system. Under the guidance of this system, the VIPs can navigate through hallways with obstacles to complete daily tasks. Detailed discussion can be found in the following sections.

### Customizing artificial vision

Accurate and quick extraction of key environmental information is pivotal in this system. The RGB-D sensors are integrated into 3D-printed glasses, and mounted on the heads of VIPs (Fig. 2a,b), allowing VIPs to perform tasks unencumbered. The RGB-D camera is selected as the main visual input for its powerful information-capturing ability and high portability (Supplementary Fig. 1). To achieve high wearability, we have used a compact Raspberry Pi microcontroller for resource-limited computing. The mobile object detection algorithm YoloV8n (ref. 19) stands out for its balanced high recognition performance and fast inference speed within this computational constraint (Supplementary Fig. 2).

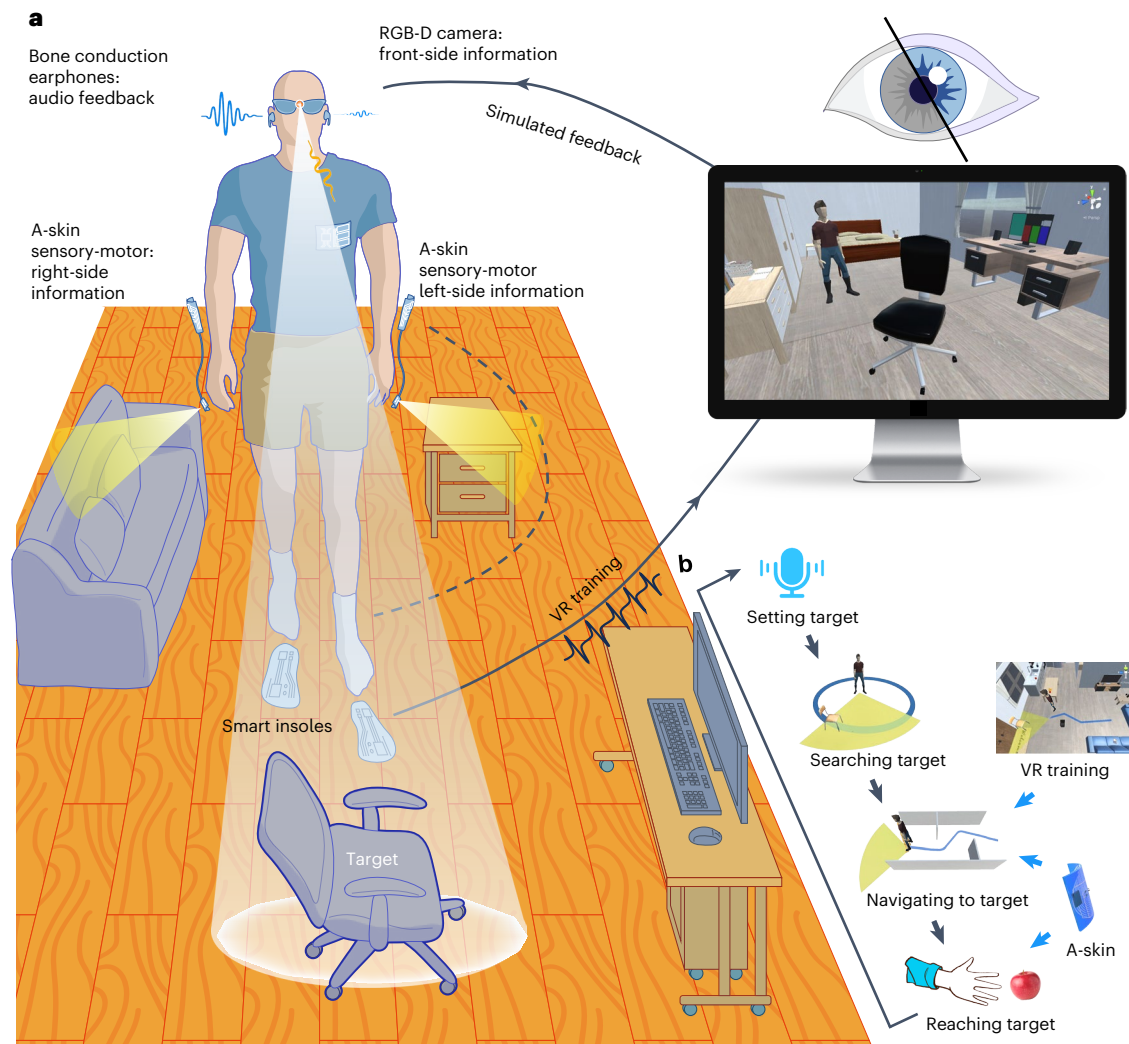
Our foundation dataset is the extensive COCO dataset, and we have finetuned it to match the scenario of visual assistance and our environment (Supplementary Fig. 3a-c). The visual performance is noticeably improved. The revised model demonstrated superior mean average precision on our custom test dataset (revised 0.82, pretrained 0.59), and maintained comparable mean average precision on the COCO test dataset (revised 0.41, pretrained 0.46) (Fig. 2c and Supplementary Fig. 3d,e). This indicates that in addition to the 21 key objects, the system's output capabilities can be expanded to the full range of 82 categories from the whole dataset. However, the increase in categories may reduce the accuracy. Ultimately, the VIPs are flexible to choose between visual breadth and accuracy on the basis of their individual needs.

In addition to the overall accuracy, visual assistance necessitates consistent target visibility despite changes in distance and perspective. To achieve omnidirectional detection, we further augmented the training dataset with images depicting objects from diverse angles and distances. Figure 2d showcases the notable enhancement in our model's capability to identify a typical chair from various angles and distances, both before and after this modification. More instances are presented in Supplementary Fig. 4. Combined with the depth map, the position of the object relative to the camera can be extracted for further analysis (Supplementary Fig. 5).

Besides, we also addressed the real-life challenges of low ambient light and motion blur. The combination of RGB and infrared images ensures the system's robustness to light changes (Supplementary Fig. 6). The embedded inertial measurement unit (IMU) is used to evaluate the motion turbulence and select stable RGB-D frames for computation, avoiding inaccurate analysis from blurred images (Supplementary Fig. 7). Afterwards, we demonstrated the system's visual capabilities using a humanoid robot. The task is to approach a target (apple) under guidance of artificial vision. The sequence highlights the system's real-time visual recognition, setting a foundation for user-centred assistive technologies (Fig. 2e and Supplementary Video 1a).

### Personalizing mobility algorithm

The phase after orientation is mobility, which refers to the ability to approach designated target while avoiding obstacles in the path. The effectiveness of this process hinges on two critical factors: AI's computation of obstacle avoidance and the relay of computed information into human. The obstacle avoidance rules need to be robust, aligned with human behaviours. The feedback cues must be intuitive to minimize the cognitive load.



**Fig. 1 | Overview of the wearable multimodal visual assistance system.** **a**, Structure of the wearable system. The wearable visual assistance system includes an RGB-D camera, A-skins and a VR training platform. It uses intuitive multimodal feedback to assist VIPs with daily life tasks and provides VR rehabilitation training. **b**, Workflow of the system. It demonstrates the process of

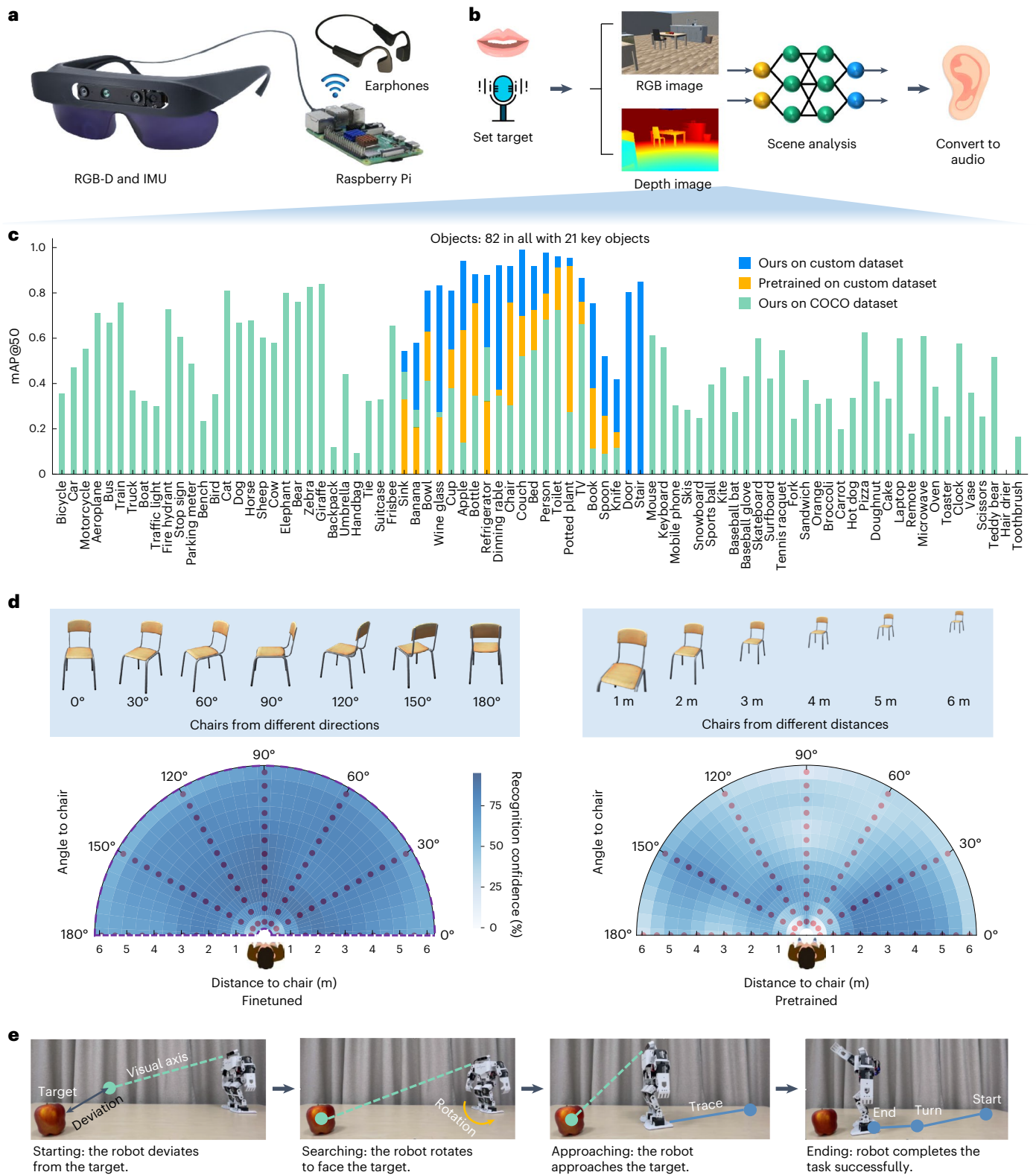
a representative visual assistance task. The system receives instructions from the VIP and then assists in navigating to the target of interest. The A-skin enhances navigation and provides postnavigation reach-and-grasp assistance. VR training helps the VIP understand and adapt to the system.

**Obstacle avoidance.** This system achieves the robust obstacle avoidance by analysing the depth data in front. To enable accurate detection of various obstacles, such as hanging, ground-level and sunken obstacles, we use a dual-method approach: a global threshold combined with a ground interval. The global threshold method checks all pixels in the depth image, and pixels with depth within this threshold are considered as obstacles. However, to avoid misidentifying the ground as obstacle, the threshold value is limited. Particularly in visual assistance, large pitch occurs when users look down, further reducing the threshold value. Figure 3a(i) shows the relationship between field of view (FoV), user height and camera pitch. As a result, this method effectively detects hanging and tall obstacles, but struggles to detect obstacles that are low and far away.

As a complement, we introduced the ground interval method. This technique models the unobstructed ground as a monotonically increasing function along the vertical index of the depth image<sup>20</sup>. The presence of obstacles causes deviations in depth from the fitted ground curve. Supplementary Fig. 8 compares the effects of different fitting functions, ultimately selecting an exponential function due to its superior fit over a wide range. Supplementary Fig. 9a–f introduces its detailed process of obstacle detection. This method performs

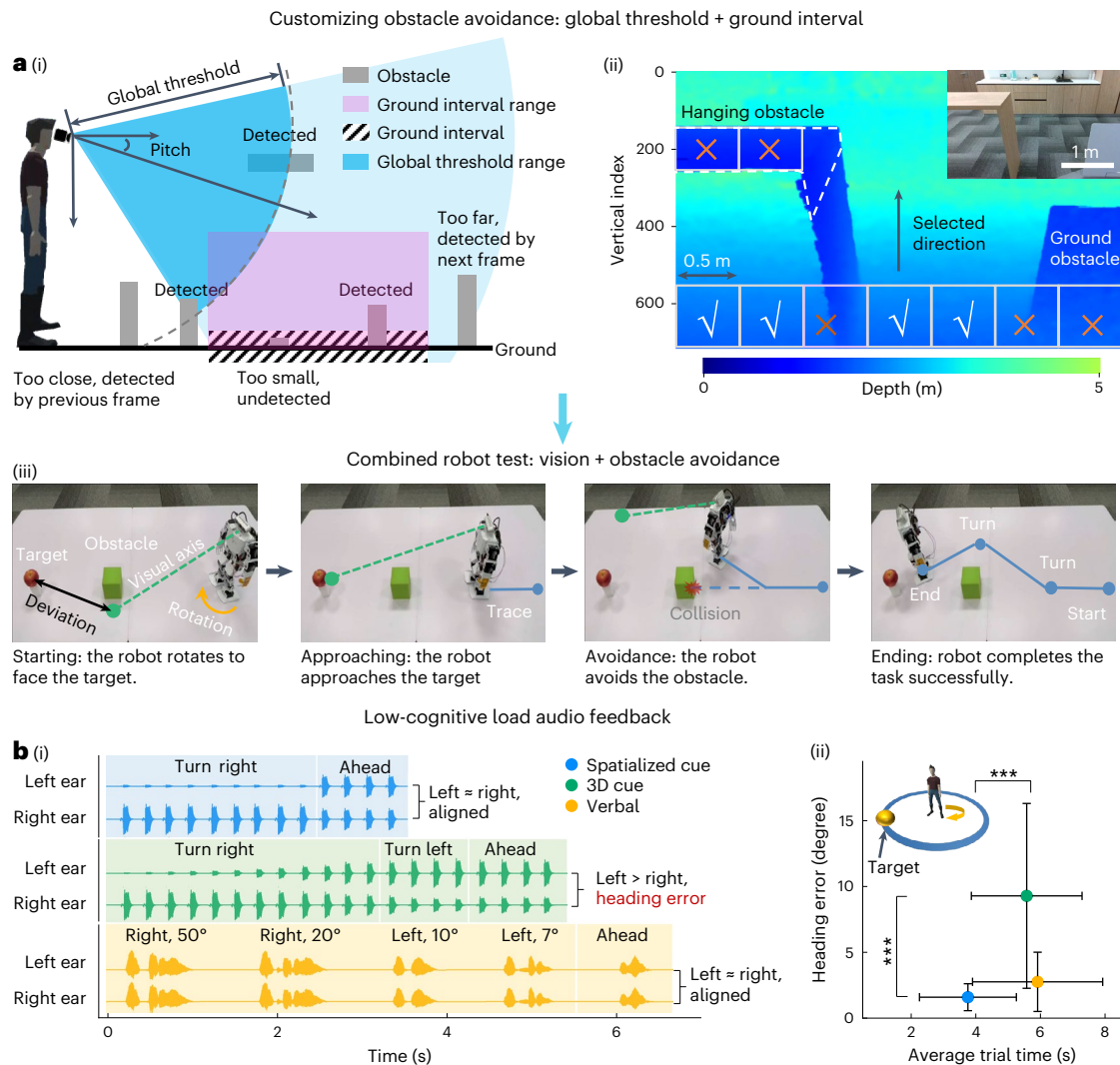
effectively under various conditions, including sensor angle, user height and ground slopes, with only slight variations in the working range (Supplementary Fig. 9g–i). To account for real-world variations in ground surfaces and sensor errors, this method requires careful calibration of the threshold height. To achieve this, we built a dataset from our own environment and selected the ground interval that can generate the highest detection rate, with details shown in Supplementary Fig. 10. With this method, we observed the highest detection rate of 95% at a threshold of 0.18 m.

On the basis of the results of obstacle detection, the path selection can be realized. The depth image is empirically divided into seven segments (in columns). Each segment is assessed for the presence of obstacles and the passable width. The segment without obstacles and requiring the smallest rotation is selected as the optimal direction (Fig. 3a(ii)). Supplementary Fig. 11 shows representative examples of estimating the optimal direction in real scenes. We further demonstrated a complete navigation task, which has an obstacle in the path, by a humanoid to verify the orientation and mobility ability of this system (Fig. 3a(iii) and Supplementary Video 1b). The whole algorithm is based on infrared depth image, therefore, the navigation can be performed well, even in a completely dark environment.



**Fig. 2 | Personalizing artificial vision. a**, Hardware components of the artificial visual components. **b**, Schematic of workflow of the artificial vision. **c**, Tuning dataset to improve the overall detection rate. The model achieves high mAP specifically for the key 21 objects and high flexibility to expand to cover all the 82 categories. The VIPs can balance between accuracy and object categories.

**d**, Tuning dataset to improve omnidirectional detection. The area circled by the purple dashed line indicates its working range (confidence > 0.5). The model exhibits an improved recognition rate across visual angles and distances, enabling a dynamic navigation process. **e**, A humanoid demonstrates dynamic processes of visually locating and approaching a target.



**Fig. 3 | Customization to improve mobility.** **a**, Crafting obstacle avoidance algorithm. (i) Schematic of the complementary obstacle detection algorithm. Global threshold detects high and close obstacles (blue), while ground interval detects low and far objects (pink). (ii) Schematic of path selecting on the depth map. The representative depth image (height 1.7 m, pitch 25°) is divided into seven segments. The segment without obstacles and requiring the smallest rotation is selected as the optimal direction. The dashed region on top indicates obstacles detected by the global threshold. (iii) Humanoid demonstrates

combined abilities of visual guidance and obstacle avoidance. **b**, Comparison of audio feedback strategies. VIPs ( $n = 12$ ) turned toward a randomly positioned target under the instruction of three audio feedbacks. (i) A representative sequence of one task and (ii) the statistics of all 12 VIPs. The spatialized cue outperforms its counterparts by the short trial time and small heading error (\*\* $P < 5 \times 10^{-3}$ , \*\*\* $P < 5 \times 10^{-4}$ , two-way analysis of variance (ANOVA), random effect, participant; fixed effect, feedback method). The dots represent the mean, and the error bars indicate one standard deviation from the mean.

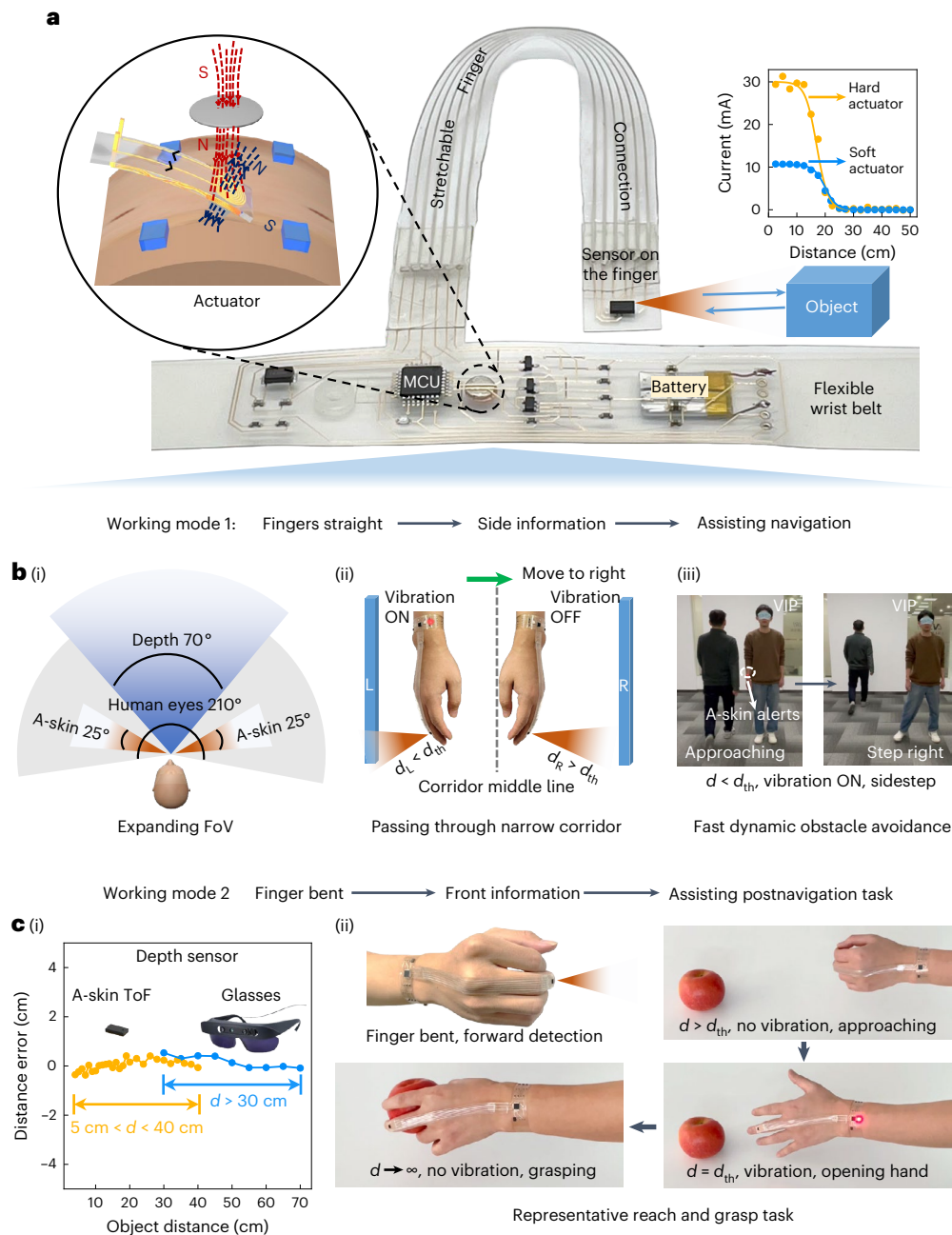
**Low-cognitive load audio feedback.** Clear and intuitive feedback signals facilitate easy interpretation by the brain, resulting in a lower cognitive load. This holds particular importance for visual assistance, where visual information capacity predominates over other sensory modalities. Preserving the functionality of these modalities is also essential, as it ensures their inherent capabilities remain largely intact, even as some bandwidth is allocated for visual assistance. By adhering to these principles, we have achieved a harmonious integration of the system’s functionality with the user’s natural sensory processing.

Auditory sense ranks as the second most dominant in humans and is often leveraged to convey high-density information in wearable visual aids<sup>21–24</sup>. Yet, auditory overload can cause discomfort during use<sup>23,24</sup>. The present design mitigates this challenge by using biocompatible encoding strategies to convey high-level scene information such as the result of path directions. We conducted a turn-to-target experiment to find the proper audio feedback with 12 VIPs. We compared three feedback approaches—spatialized cues, 3D sounds and verbal

instructions—using metrics such as final heading error and trial time to evaluate the performance. Figure 3b(i) and Supplementary Fig. 12a illustrate a typical task sequence, with the statistical results shown in Fig. 3b(ii) and Supplementary Fig. 12b,c. Spatialized cues exhibit the least heading error and minimal time cost, rendering them the preferred choice in our design. In contrast, 3D cues vary continuously with different azimuths, which can complicate distinction and lead to imprecise direction perception. While verbal instructions use simple and interpretable words, they require longer playback time, resulting in slower responses.

**Developing complementary artificial skins**

AI-powered artificial skins hold great promise in fields for human health<sup>25–27</sup>. Haptic feedback leverages the skin’s vast surface area and broad neuronal receptor distribution, making it an excellent medium for sensory rehabilitation<sup>28</sup>. However, its relatively low information density makes it challenging for visual assistance alone.



**Fig. 4 | Artificial skin sensory-motor for efficient haptic feedback.** **a**, Structure of the A-skins. The A-skin is a flexible system featuring a ToF sensor, an MCU and a haptic actuator. When the measured distance falls below a preset threshold, the actuator generates vibration feedback through electromagnetic force from a coil. **b**, A-skin function 1: navigation enhancement. (i) The A-skin expands the FoV

by detecting lateral obstacles. Representative scenarios of (ii) passing through a narrow corridor and (iii) avoiding dynamic obstacle are presented. **c**, A-skin function 2: postnavigation task assistance. (i) The A-skin expands the near-range distance of the RGB-D camera, enhancing hands-on task accuracy. (ii) Typical sequence of a reach-and-grasp task demonstrating its working process.

Here we developed a stretchable A-skin to complement audio feedback. The A-skin is an integrated sensory-motor device assisting the artificial vision by providing near-distance alerting (Fig. 4a and Supplementary Fig. 13). The system’s core component is a compact time-of-flight (ToF) sensor (Supplementary Fig. 14). When the distance is smaller than a predefined threshold, the actuator vibrates to stimulate the skin. The actuator is designed as a slim, lightweight polyethylene terephthalate (PET) cantilever for energy efficiency. Micropyramid arrays at the actuator’s tip can enhance stimulation of skin mechanoreceptors, further reducing power demand (Fig. 4a). Note that a deliberate gap is maintained between the driving circuit and the skin, which facilitates the actuator’s vibration while promoting

air circulation in between. This air circulation improves skin comfort, breathability and thus long-term wearability.

The A-skin is also strategically positioned to enhance usability. The placement is guided by the classical cortical homunculus, which are exaggerated human figures illustrating the proportion of the brain devoted to the sense of touch in each part of the body (Supplementary Fig. 15)<sup>29</sup>. The A-skins should be placed on regions with more sensory but fewer motor activities, aiming to achieve high sensitivity without interfering with normal motion tasks. On the basis of these considerations, the actuator, along with the driving circuit, is placed on the wrist. The sensor is positioned on the back of index or middle finger, with its function changing on the basis of the finger’s status (bent or straight).

Figure 4b illustrates the A-skin's first function of navigation assistance. In this mode, the VIP's hands hang naturally at their sides, with sensors on the extended fingers detecting lateral obstacles, thereby expanding the visual angle (Fig. 4b(i)). This setup is particularly useful for navigating complex environments. Figure 4b(ii) shows a scenario where the VIP passes through a narrow corridor. When the distance to the left wall is below the threshold, vibration from the left wrist prompts the participant to move right and vice versa for the right side. This mechanism helps keep the VIPs centrally aligned on their path, enhancing safety. The threshold ranges from 5 to 40 cm. A larger far-end value indicates a higher level of safety, while a smaller value allows for more flexibility. The VIP can personalize this threshold to suit their needs. The angular monitoring range of the ToF sensor can be slightly tuned. A subtle forward rotation enables the sensor to capture anterolateral information, improving the capacity to anticipate obstacles ahead. In addition, the time from obstacle detection to haptic feedback generation is only 18 ms (Supplementary Fig. 16), enabling quick response and dynamic obstacle avoidance. Figure 4b(iii) shows a user evading a person approaching to him rapidly from the left, with details shown in Supplementary Video 2a.

Figure 4c illustrates the system's second function of assisting with postnavigation tasks. The A-skin enhances the glasses' depth sensor by extending its effective range (Fig. 4c(i)). Figure 4c(ii) depicts a typical reach-and-grasp task. During reaching, the user's fingers bend forward while the sensor dynamically measures the distance to the object. On reaching the threshold distance, the vibrator signals the VIP to shift from reaching to grasping. The vibration ceases with this gesture change, providing a clear reminder. The threshold can be personalized to the VIPs' habits, enabling quick and safe task completion. Besides, the ToF sensor directly faces the object, circumventing the hand-object obstruction issue. Its small FoV and rapid response facilitate precise hand-object alignment (Supplementary Video 2b). With meticulous design in structure, placement and function, the A-skin enhances the artificial vision while maintaining high integration level and thereby better utility.

### Metaverse training and testing

Simple yet efficient training enables instinctive use of rehabilitation tools including visual assistance. The emerging metaverse assists training by providing immersive experience, enabling efficient fusion of the human-machine-environment at low cost and injury risk. We developed a VR training platform, whose workflow is schematically shown in Fig. 5a. Briefly, the avatar in the virtual environment, aligned with the RGB-D camera, identifies virtual obstacles and guides the VIP's movement in the real world. Their movements are captured by smart insoles, synchronizing the avatar's actions. This dynamic loop creates an interactive learning environment for the VIPs (Supplementary Video 3). The audio feedback can provide high level of immersion<sup>30</sup>, while the smart insoles offer a private and energy-efficient way to monitor the user's movement. Their detailed structure and training process are shown in Supplementary Fig. 17. The motion control approach was validated by a sighted participant, who wore the device and traversed designated paths (Fig. 5b and Supplementary Fig. 18).

On the basis of the above preparation, we proceeded to on-body testing. We recruited a group of VIPs ( $n = 12$ ) from diverse backgrounds (Supplementary Table 1), and meticulously designed three distinct scenarios for efficient training and validation. Scenes 1 and 2 are two fundamental scenarios for training. The first is a basic in-room task where a VIP can directly see their target, requiring VIPs to avoid obstacles on the basis of directional feedback (Fig. 5c(i)). The second represents a typical cross-room task scenario, where VIPs cannot directly see the target. They need to navigate to an intermediate target (the table) before approaching the final target (Fig. 5d(i)). These scenarios were randomly alternated during training, with each scenario trained seven times. The first (pretest) and last (posttest) sessions of each scenario were conducted in the real world using the wearable system.

Figure 5c(ii) shows a representative participant who walked slowly and collided with the environment during early sessions. This might be due to a lack of trust in the system or unfamiliarity with the feedback, leading to misinterpretation. As training progressed, the participant navigated faster and more smoothly, with fewer collisions. This improvement persisted when the scene layout was changed and when transitioning to real-world settings. Statistically, there was a significant improvement after training in various metrics such as walking speed, walking time, walking distance and collisions (Fig. 5c(iii)). Additional details can be found in the Supplementary Information (Supplementary Fig. 19). Figure 5d(i) shows the sequential navigation setup, which does not require global map-based path planning. Although the final target is initially obscured, VIPs can reach it through a series of intermediate targets. After training, a representative VIP showed fewer turns and walked faster and more smoothly (Fig. 5d(ii)). Statistical results for more VIPs are shown in Fig. 5d(iii) and Supplementary Figs. 20 and 21. These results proved that VR training improves mobility in the real world.

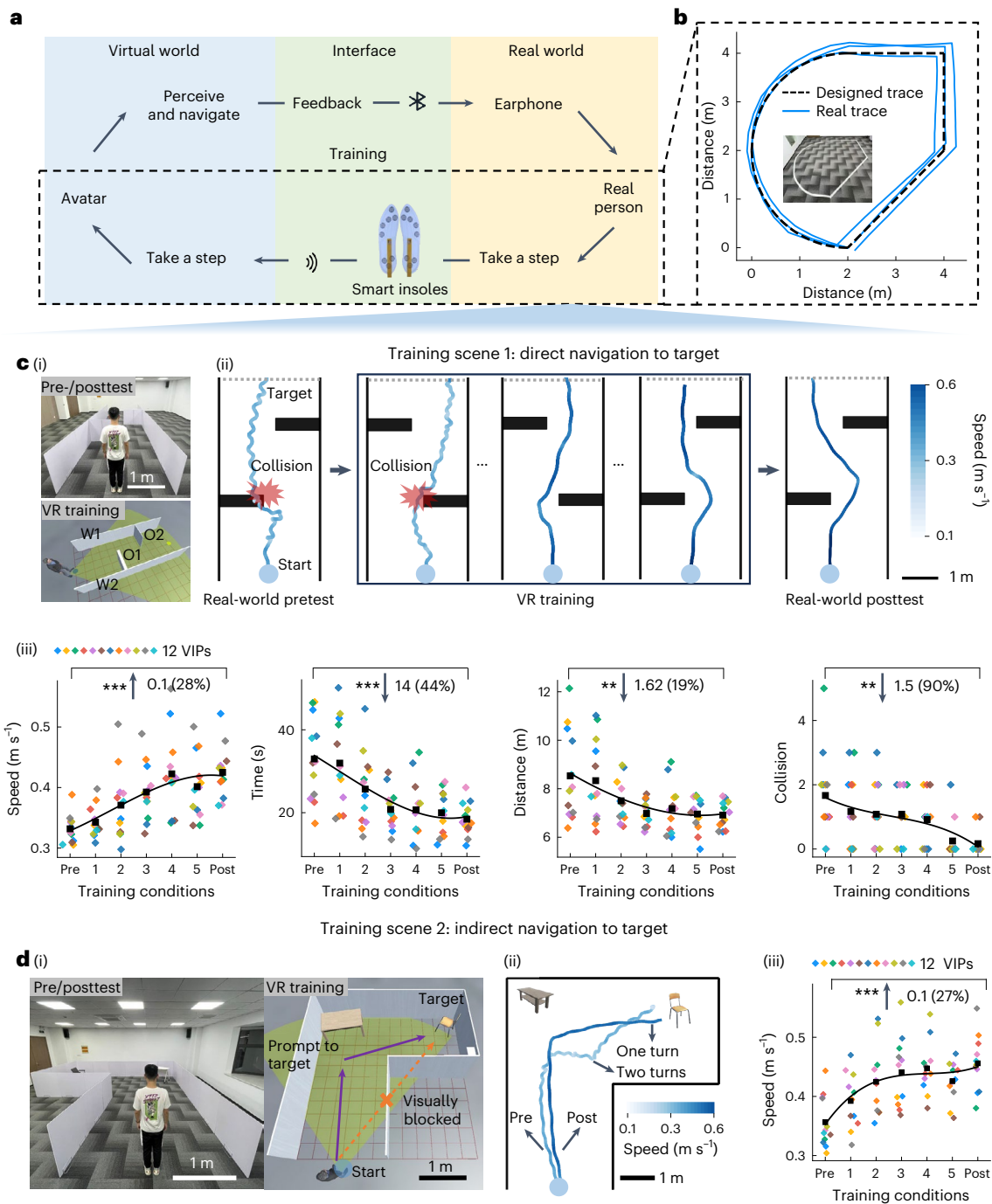
Our virtual environments, although basic, encompass crucial elements such as target search, linear movement and turning, with complex scenarios being combinations of these fundamentals. Through repeated training sessions, VIPs acquire proficiency in interpreting system feedback and making informed navigation choices guided by provided cues. To ensure that enhancements stem from skill development rather than mere scene memorization, we introduced random variations during training procedures, including altering the training sessions, modifying scene layouts and introducing rest intervals between sessions.

To further confirm the training benefits, we created scene 3—an untrained, intricate maze structure as the validation scene. After training in scenes 1 and 2, VIPs were promptly transitioned to scene 3 for evaluation. Participants were required to navigate the maze to locate a key object (a chair) at the exit (Fig. 6a(i) and Supplementary Fig. 22). Before training, the majority of novices (9 out of 12) experienced collisions or could not independently complete the task. However, all participants who underwent training successfully accomplished the task, as depicted in Supplementary Video 4 and Supplementary Fig. 23. Substantial improvements across various metrics are illustrated in Supplementary Fig. 24. These results demonstrate that the training in scenes 1 and 2 laid the groundwork for navigating complex scenes, indicating the cross-scenario robustness of our system.

In comparison to the traditional white cane, the system showcased smoother turning and more efficient pathfinding, as evidenced by the traces in Fig. 6a(ii). Further traces are available in Supplementary Fig. 25. Statistical analysis revealed that walking speed with the wearable system was on par with cane usage<sup>31,32</sup>. The navigation time and walking distance were significantly reduced by approximately 25%. The incorporation of A-skins notably decreased collision incidents (Fig. 6a(iii)).

Our structured training and validation on scenes 1 to 3 demonstrate the effectiveness of the wearable system. To further assess its impact across diverse real-world environments with a broader participant pool, we engaged an additional eight VIPs in real-world testing across four varied scenarios: indoor and outdoor environments, static and dynamic obstacles, work-related and daily living scenarios, and a comprehensive task mirroring real-world challenges (Fig. 6b(i),(ii) and Supplementary Fig. 26). The proficient completion of these tasks underscores the system's effectiveness in meeting real-life challenges. Detailed overview of these representative tests can be found in Supplementary Videos 5–8.

After all experiments, the VIPs completed the System Usability Scale, which is widely used for system usability evaluation<sup>32,33</sup>. They gave the wearable system average usability scores of 79.6 (Supplementary Table 2), corresponding to the 85th in a distribution of 5,000 commercial and research devices<sup>34</sup>. In open-ended questions, participants



**Fig. 5 | Metaverse for immersive training.** **a**, Workflow of VR training. **b**, Verification of the VR platform by comparing the real and reconstructed traces. **c**, Training process in scene 1: direct navigation. (i) Scene schematic. (ii) Representative training process. (iii) Statistical results. After training, the VIPs ( $n = 12$ ) completed the tasks fast and smoothly with fewer collisions. The instantaneous speed (depicted by colour darkness) and curve smoothness reflect

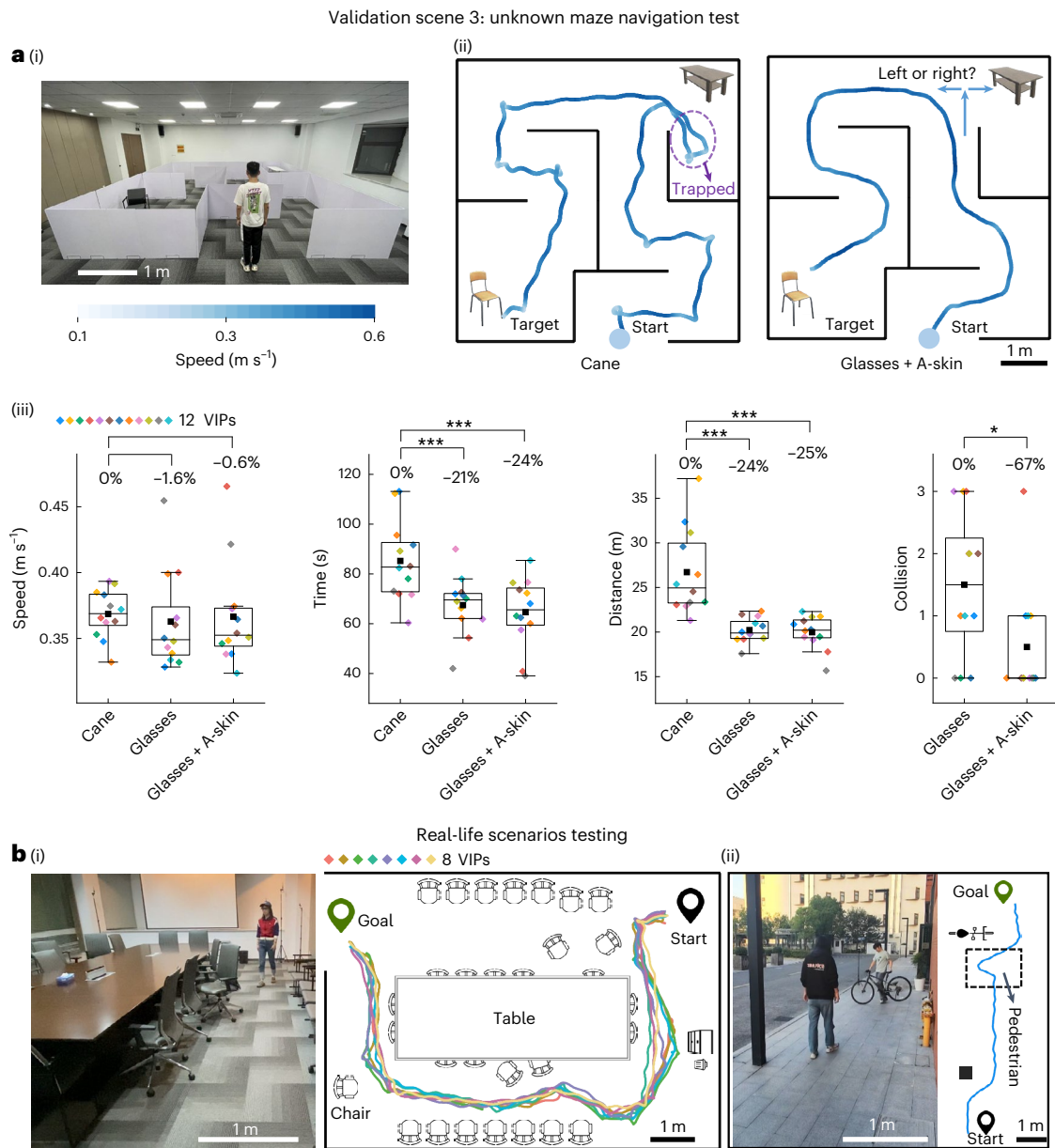
their increasing confidence on the system. **d**, Training process in scene 2: indirect navigation. This is to show the navigation to an initially unseen target without the need of global map. (i) Scene schematic. (ii) Representative trace before and after training. (iii) Statistics of walking speed of all the 12 VIPs ( $*P < 0.05$ ,  $**P < 5 \times 10^{-3}$ ,  $***P < 5 \times 10^{-4}$ , two-way ANOVA: random effect, participant; fixed effect, training condition).

reported that the wearable system provided intuitive feedback and engaging training games.

## Discussion

We engineered a human-centric, wearable visual assistance system by synergizing advancements in AI and hardware. It caters to the unique needs of VIPs while ensuring minimal physical and cognitive strain. The integration of audio and haptic components offers effective visual

support. The audio segment focuses on processing intricate data within a frontal 70° field using AI algorithms but operates at a slower update rate of 250 ms per cycle. In contrast, the A-skin element swiftly handles sparse information, particularly detecting side obstacles at an update rate of 18 ms per cycle. The system mirrors the human eye configuration, allocating a central region for detailed processing and a peripheral area for surveillance. By blending dense and sparse data processing, fast and slow response rates, and front-side observation capabilities,



**Fig. 6 | Real-world testing.** **a**, Validation in new scene 3. After training in scenes 1 and 2, the VIPs ( $n = 12$ ) navigated through a maze, evaluating the training efficacy and the system’s performance in unknown real-world scenarios. (i) The photo of scene 3. (ii) Typical traces of a VIP guided by a cane and this system and (iii) the statistical results of all 12 VIPs. Compared with the cane, the system showed smoother turning. It maintains the high walking speed of the cane, but greatly reduces the navigation time and walking distance (by -25%). The A-skins markedly reduced the collisions (by 67%) without compromising other

metrics ( $*P < 0.05$ ,  $**P < 5 \times 10^{-3}$ ,  $***P < 5 \times 10^{-4}$ , two-way ANOVA, random effect, participant; fixed effect, assistive tool). The boxes extend from the lower to upper quartile values of the data, with a line at the median and a dot at the mean. The whiskers extend between the lowest and highest data points within 1.5 times the interquartile range. **b**, Testing in real-life scenarios. (i) Additional VIPs ( $n = 8$ ) completed a navigation task in a cluttered conference room, highlighting the system’s effectiveness in cluttered indoor environments. (ii) A VIP navigated a dynamic outdoor environment, demonstrating its applicability outside.

the system efficiently monitors a broad area with minimal power usage and low latency. In terms of hardware, the system introduces a flexible sensory-motor A-skin and self-powered triboelectric smart insoles, offering better wearability compared to conventional rigid devices. Experiments with VIPs showed significant increases in various metrics. The system’s personalized algorithm and lightweight hardware provide insights for user-friendly wearable visual assistance systems.

The system showcases an overall cycle time that aligns with human reaction time (200–300 ms), as illustrated in the time sequence outlined in Supplementary Fig. 27. Operating at 4 Hz, the system consumes 6.4 W of power (Supplementary Fig. 28). Further analysis indicates that within one cycle, the RGB-D camera consumes the most energy (0.85 J

per cycle, 56% of the total energy). Additionally, a substantial portion of the power is allocated to support the Raspberry Pi system (0.45 J per cycle, 29% of the total energy). The A-skin consumes roughly 0.1–0.2 W of power depending on whether the actuator is activated or not. The system consumes less energy than many existing systems<sup>14,24,32</sup>, however, the overall energy consumption exceeds that of biological visual systems<sup>35</sup>. Transitioning to advanced cameras such as biomimetic eyes<sup>36–38</sup> and energy-efficient application-specific integrated circuit chips in lieu of Raspberry Pi could further mitigate the overall power consumption.

Overall, the system stands as a promising research prototype, setting the stage for the future advancement of wearable visual assistance.

Envisioned as an open-source platform, it welcomes interdisciplinary collaborations to drive its progression, including enhancements in vision models, integrated wearable electronics, insights from neuroscience and personalized generative training environments. Crucially, the participation of a broader, more diverse group of VIPs is vital for advancing navigation aids tailored to their specific needs.

## Methods

### Overall system hardware

A Depth Camera D435i (Intel RealSense) is mounted on a 3D-printed glasses frame. It is connected to a Raspberry Pi 4 Model B microcontroller via USB. All computations are performed locally, without the need for remote server or other services. Audio cues (50 milliseconds spatialized cues or verbal instructions) are delivered through a wireless bone-conducting earphone (OpenRun Pro, Shokz). The system is powered by a commercial lithium-ion battery. Excluding the battery, the entire system weighs 195 g. The smart insoles include a three-axis magnetometer (HMC5883L, Adafruit) and triboelectric sensors. Angle and pressure data are used to estimate walking and rotation. A humanoid robot (TonyPi Pro, Huaner Electronic Technology) was used in the demonstration. Its head camera was replaced with D435i and demonstration-related functions were programmed into its operating system.

### Preparation of artificial skins

The electronic components, including the ToF (VL53L0X, STMicroelectronics) sensor, MCU (ATMEGA328P-AU, Microchip Technology Inc.) and other components are integrated on flexible PET substrate by a flexible printed circuit board fabrication process. For the actuation system, a thin permanent magnetic (thickness 0.5 mm, diameter 5 mm) is fixed on bottom of PET substrate. A single layer copper coil (diameter 5 mm) is fixed on a narrow PET belt to form the vibrator. The copper coil is electrically connected to the output of MCU. The stretchable connection between the ToF sensor and control circuit is fabricated by printing silver particles onto a PDMS (polydimethylsiloxane) substrate (sylgard 184) by the material printer (Scientific 3, Prtronic). It is further packaged by another PDMS layer to improve the flexibility and stretchability.

### Fabrication of triboelectric smart insoles

The triboelectric smart insole was fabricated by 3D printing of concave pyramid array (100- $\mu\text{m}$  bottom diameter, 200  $\mu\text{m}$  in pitch). The microcones on PTFE (polytetrafluoroethylene) were prepared by hot embossing PTFE film (30  $\mu\text{m}$ ) against the mask for 2 min at 170  $^{\circ}\text{C}$  under a pressure of roughly 10  $\text{kg cm}^{-2}$ . The Al mesh was manually weaved by aluminium fibres (100  $\mu\text{m}$  in diameter), which were then covered by an aluminium film of 100  $\mu\text{m}$ . Small springs (1 mm in diameter) were used to separate the Al mesh and PTFE substrate. The electronic signal was conducted from the aluminium electrode.

### Participants

The minimum number of participants with visual impairment was estimated using a power analysis of data from a pilot study. The mean walking speeds of novices and those trained were 0.283 and 0.336  $\text{m s}^{-1}$ , respectively, and the standard deviation is 0.038  $\text{m s}^{-1}$ . The power analysis used a power of 0.9 at a significance level of 0.05. The power analysis recommended 9 participants and 12 were selected to provide a factor of safety in case data from any participants were found to be unusable during later analysis. We recruited an additional eight VIPs in the real-world test. A total of 20 VIPs participated in the experiment. The participants were selected randomly from diverse backgrounds, such as education, age, gender and personal habits. All participants had no experience with similar assistive devices, minimizing the effect of previous knowledge that participants may have had. All participants were further blindfolded to avoid the residual vision of some VIPs.

All participants were volunteers and provided written informed consent before completing the protocol (IRB HRP ADOIV01), which was approved by the Shanghai Jiao Tong University Institutional Review Board. Consent was obtained for the publication of identifiable images and videos of the research participants. The detailed information about the participants with visual impairment is shown in Supplementary Table 1.

### Metaverse rehabilitation training and real-world navigation experiments

Unity (v.2020.3.36f1) was used to build the virtual environment. Camera and ray casts are used to simulate the RGB-D camera. Scene 1 features two parallel walls and two obstacles. The walls are 2.5 m apart and 1 m high, while the obstacles are 1.2 m wide and 1 m high. Scene 2 includes a table in the upper left corner, a chair in the upper right and surrounding walls 1 m high. The chair is invisible from the starting point due to wall obstructions. Collisions are simulated with a collision sound effect. During the pretest and posttest, we built the virtual scenarios in the real world, using the wearable system for navigation.

To avoid the influence of scene memorization, scenes 1 and 2 were randomly alternated during the training sessions, and the layouts of these scenes were randomly mirrored. After each session, a 2-min break was given, during which participants could ask and answer questions. These measures ensured that participants did not repeatedly train on the exact same scene configuration.

To evaluate the training effect, scene 3 was devised for validation. It consisted of a maze, and the total path length was approximately 25 m. A table was placed as a landmark. After training in scenes 1 and 2, the VIPs proceeded directly to scene 3, which was entirely new and had not been previously encountered. Participants wore hats with tags on them, and an Ultra-Wideband system recorded the tags' two-dimensional positions at a rate of 50 Hz. Average walking speed was calculated by dividing the total path length by the completion time. Collisions were counted whenever a participant made contact with walls or obstacles. The training regimen was meticulously organized to guarantee that users attained consistent performance levels without excessive exertion.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All the data required to assess the conclusions of the article are available via Figshare at <https://doi.org/10.6084/m9.figshare.26103583> (ref. 39) and are available for reuse for ethical and scientific purposes.

### Code availability

The exemplary scripts for data processing and analysis for this study are available in the GitHub repository at <https://github.com/Jian-Tang2000/wearableSystem> (ref. 40).

### References

1. Roska, B. & Sahel, J. A. Restoring vision. *Nature* **557**, 359–367 (2018).
2. Maidenbaum, S., Abboud, S. & Amedi, A. Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci. Biobehav. Rev.* **41**, 3–15 (2014).
3. Tapu, R., Mocanu, B. & Zaharia, T. Wearable assistive devices for visually impaired: a state of the art survey. *Pattern Recogn. Lett.* **137**, 37–52 (2020).
4. Herskovitz, J. et al. Hacking, switching, combining: understanding and supporting DIY assistive technology design by blind people. In *Proc. 2023 CHI Conference on Human Factors in Computing Systems* (eds Albrecht, S. et al.) 1–17 (ACM, 2023).

5. Shah, D., Osiński, B. & Levine, S. Lm-nav: robotic navigation with large pre-trained models of language, vision, and action. In *Proc. 2024 Conference on Robot Learning* (eds Marc, T. et al.) 492–504 (PMLR, 2024).
6. Wu, W. et al. Vision-language navigation: a survey and taxonomy. *Neural Comput. Appl.* **36**, 3291–3316 (2024).
7. Spoendlin, H. & Schrott, A. Analysis of the human auditory nerve. *Hear. Res.* **43**, 25–38 (1989).
8. Bach-y-Rita, P. Tactile sensory substitution studies. *Ann. N. Y. Acad. Sci.* **1013**, 83–91 (2004).
9. Nguyen, T. H. et al. A wearable assistive device for the blind using tongue-placed electro-tactile display: design and verification. In *Proc. 2013 International Conference on Control, Automation and Information Sciences* (eds Jin, B. et al.) 42–47 (IEEE, 2013).
10. Huang, Y. et al. A skin-integrated multimodal haptic interface for immersive tactile feedback. *Nat. Electron.* **6**, 1020–1031 (2023).
11. Guo, H. et al. A highly sensitive, self-powered triboelectric auditory sensor for social robotics and hearing aids. *Sci. Robot.* **3**, eaat2516 (2018).
12. Liu, Y. et al. Soft, miniaturized, wireless olfactory interface for virtual reality. *Nat. Commun.* **14**, 2297 (2023).
13. Hoffmann, R. et al. Evaluation of an audio-haptic sensory substitution device for enhancing spatial awareness for the visually impaired. *Optom. Vis. Sci.* **95**, 757–765 (2018).
14. Li, G., Xu, J., Li, Z., Chen, C. & Kan, Z. Sensing and navigation of wearable assistance cognitive systems for the visually impaired. *IEEE Trans. Cogn. Dev. Syst.* **15**, 122–133 (2022).
15. Xu, C. et al. A physicochemical-sensing electronic skin for stress response monitoring. *Nat. Electron.* **7**, 168–179 (2024).
16. Yang, Q. et al. Mixed-modality speech recognition and interaction using a wearable artificial throat. *Nat. Mach. Intell.* **5**, 169–180 (2023).
17. Gu, L. et al. A biomimetic eye with a hemispherical perovskite nanowire array retina. *Nature* **581**, 278–282 (2020).
18. Yao, K. et al. Encoding of tactile information in hand via skin-integrated wireless haptic interface. *Nat. Mach. Intell.* **4**, 893–903 (2022).
19. Jocher, C. et al. Ultralytics yolov8. *GitHub* <https://github.com/ultralytics/ultralytics> (2023).
20. Krausz, N. E., Lenzi, T. & Hargrove, L. J. Depth sensing for improved control of lower limb prosthesis. *IEEE Trans. Biomed. Eng.* **62**, 2576–2587 (2015).
21. González, J. & Yu, W. Multichannel audio aided dynamical perception for prosthetic hand biofeedback. In *Proc. 2009 IEEE International Conference on Rehabilitation Robotics* (eds Kiyoshi, N. et al.) 240–245 (IEEE, 2009).
22. George, J. A. et al. Biomimetic sensory feedback through peripheral nerve stimulation improves dexterous use of a bionic hand. *Sci. Robot.* **4**, eaax2352 (2019).
23. Meijer, P. B. An experimental system for auditory image representations. *IEEE Trans. Biomed. Eng.* **39**, 112–121 (1992).
24. Levy-Tzedek, S., Hanassy, S., Abboud, S., Maidenbaum, S. & Amedi, A. Fast, accurate reaching movements with a visual-to-auditory sensory substitution device. *Restor. Neurol. Neurosci.* **30**, 313–323 (2012).
25. Yang, Y. et al. Breathable electronic skins for daily physiological signal monitoring. *Nano-Micro Lett.* **14**, 161 (2022).
26. Chortos, A., Liu, J. & Bao, Z. Pursuing prosthetic electronic skin. *Nat. Mater.* **15**, 937–950 (2016).
27. Yu, Y. et al. All-printed soft human-machine interface for robotic physicochemical sensing. *Sci. Robot.* **7**, eabn0495 (2022).
28. Yu, X. et al. Skin-integrated wireless haptic interfaces for virtual and augmented reality. *Nature* **575**, 473–479 (2019).
29. Sanes, J. N. & Donoghue, J. P. Plasticity and primary motor cortex. *Annu. Rev. Neurosci.* **23**, 393–415 (2000).
30. Susal, J., Krauss, K., Tsingos, N. & Altman, M. Immersive audio for VR. In *Proc. 2016 AES International Conference on Audio for Virtual and Augmented Reality* (eds Andres, M. et al.) 99–124 (Audio Engineering Society, 2016).
31. Zhao, Y. et al. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems* (eds Regan, M. et al.) 1–14 (ACM, 2018).
32. Slade, P., Tambe, A. & Kochenderfer, M. J. Multimodal sensing and intuitive steering assistance improve navigation and mobility for people with impaired vision. *Sci. Robot.* **6**, eabg6594 (2021).
33. Jordan, P. W., Thomas, B., McClelland, I. L. & Weerdmeester, B. *Usability Evaluation in Industry* (CRC, 1996).
34. Sauro, J. *A Guide to the System Usability Scale: Background, Benchmarks & Best Practices* (Measuring Usability LLC, 2011).
35. Country, M. W. Retinal metabolism: a comparative look at energetics in the retina. *Brain Res.* **1672**, 50–57 (2017).
36. Liao, F. et al. Bioinspired in-sensor visual adaptation for accurate perception. *Nat. Electron.* **5**, 84–91 (2022).
37. Park, J. et al. Avian eye-inspired perovskite artificial vision system for foveated and multispectral imaging. *Sci. Robot.* **9**, eadk6903 (2024).
38. Zhang, Z. et al. All-in-one two-dimensional retinomorphic hardware device for motion detection and recognition. *Nat. Nanotechnol.* **17**, 27–32 (2022).
39. Jian, T. et al. Human-centered design and fabrication of a wearable multimodal visual assistance system. *FigShare* <https://doi.org/10.6084/m9.figshare.26103583> (2025).
40. Jian, T. et al. Human-centered design and fabrication of a wearable multimodal visual assistance system. *Zenodo* <https://doi.org/10.5281/zenodo.14752720> (2025).

## Acknowledgements

This work was funded by STI 2030—Major Projects (grant no. 2022ZD0210000), National Science Foundation China grant (no. 62274110) and Shanghai Rising-Star Program (grant no. 21QA1404000). The individuals involved in the 2022ZD0210000 project include L.G. and B.Y., the 62274110 project includes L.G. and the 21QA1404000 project includes L.G. We sincerely thank all the participants who generously volunteered their time, provided valuable feedback and contributed to the study.

## Author contributions

L.G. and J.T. were responsible for the system design. J.T., Y.Z., G.J., L.X. and W.R. conducted the experiments. J.T., L.G., H.B. and Q.G. developed the AI algorithm, B.Y., J.Z., X.W. and Z.F. assisted the data collection and analysis. All authors participated in manuscript preparation.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42256-025-01018-6>.

**Correspondence and requests for materials** should be addressed to Leilei Gu.

**Peer review information** *Nature Machine Intelligence* thanks Wei Gao and Xinge Yu for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with

the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection The data collection process was conducted using Python (version 3.9) and Unity (version 2020.3.36f1) on a laptop that communicated with the Raspberry Pi on the device via Wi-Fi.

Data analysis The data analysis was performed using Python (version 3.9), and the code is available at <https://doi.org/10.5281/zenodo.14752720>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All the data required to assess the conclusions of the article have been archived in an online repository at <https://doi.org/10.6084/m9.figshare.26103583>

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	In our study, we used self-reporting methods to determine the gender of participants. We did not differentiate treatment or analysis based on gender. Therefore, no sex- or gender-specific analyses were conducted.
Population characteristics	The participants in our study were from diverse backgrounds and were all diagnosed with visual impairments. No further covariate-specific population characteristics, such as age or genotypic information, were specified for the purposes of this study.
Recruitment	The researchers recruited all participants by visiting the visually impaired community. All participants needed to meet the criteria of being over 18 years old and free from mobility or communication impairments to avoid any interference with learning to use assistive devices and analyzing their mobility indicators in the study. Additionally, factors such as gender, education level, and occupation did not affect recruitment. None of the participants had prior experience with the assistive devices used in the study to prevent any influence from prior knowledge.
Ethics oversight	The study protocol was approved by the Shanghai Jiao Tong University Institutional Review Board with the identifier IRB HRP AD01V01 2023. All participants have signed informed consent forms and are aware of and have agreed to the publication of their likenesses.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study aims to design an assistive device to help visually impaired individuals tackle challenges in their daily lives. We recruited visually impaired participants to engage in training sessions and indoor and outdoor scenario tests to assess the impact of the assistive device. We utilized a two-way ANOVA analysis to examine performance differences under different assistive conditions.
Research sample	The minimum number of visually impaired participants was determined based on a power analysis of data from a pilot test. In that test the mean walking speeds for novice and trained individuals were 0.283 m/s and 0.336 m/s, respectively, with a standard deviation of 0.038 m/s. The power analysis was conducted with a power of 0.9 and a significance level of 0.05, recommending 9 participants. To account for potential data loss during later analysis, 12 participants were selected. In addition, 8 more visually impaired persons were recruited for the real-world test. In total, 20 visually impaired participants took part in the experiment.
Sampling strategy	The minimum number of visually impaired participants was estimated using a power analysis of data from pilot study. The power analysis used a power of 0.9, a significance level of 0.05. The power analysis recommended 9 participants, and 12 were selected to provide a factor of safety in case data from any participants were found to be unusable during later analysis.
Data collection	The data collection process involved participants completing related experiments, after which they filled out questionnaires using pen and paper. Cameras were used to capture first-person and third-person videos of the participants during their movements. The Ultra Wideband system recorded real time positional data during experiments, which was then analyzed to assess mobility related metrics. Only the participants and the researchers were present during the data collection process; no additional observers or assistants were involved. The study design was single-blind, meaning the researchers were aware of the experimental conditions, while the participants were not informed of the study's hypothesis or specific conditions.
Timing	Data collection continued throughout the entire duration of the experiments (from 08/2023 to 11/2024), until all sessions were completed. Participants completed multiple experiments in a specified order, with designated breaks between experiments to avoid fatigue. No significant gaps occurred between data collection periods for different cohorts.
Data exclusions	No data were excluded. All data from all participants were included in each corresponding analysis.
Non-participation	No participants dropped out or declined participation; all recruited participants completed the corresponding experiments.
Randomization	All participants were exposed to every assistance condition, as the study aimed to test differences between these assistance conditions. The allocation was systematic rather than random, ensuring that each participant experienced all conditions. This design

allowed each participant to serve as their own control, thereby controlling for individual differences and ensuring consistency in the comparison of assistance conditions.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging