

# Evaluation of an Audio-haptic Sensory Substitution Device for Enhancing Spatial Awareness for the Visually Impaired

Rebekka Hoffmann, MSc,<sup>1,2\*</sup> Simone Spagnol, PhD,<sup>2,3</sup> Árni Kristjánsson, PhD,<sup>1</sup> and Runar Unnthorsson, PhD<sup>2</sup>

**SIGNIFICANCE:** Visually impaired participants were surprisingly fast in learning a new sensory substitution device, which allows them to detect obstacles within a 3.5-m radius and to find the optimal path in between. Within a few hours of training, participants successfully performed complex navigation as well as with the white cane.

**PURPOSE:** Globally, millions of people live with vision impairment, yet effective assistive devices to increase their independence remain scarce. A promising method is the use of sensory substitution devices, which are human-machine interfaces transforming visual into auditory or tactile information. The Sound of Vision (SoV) system continuously encodes visual elements of the environment into audio-haptic signals. Here, we evaluated the SoV system in complex navigation tasks, to compare performance with the SoV system with the white cane, quantify training effects, and collect user feedback.

**METHODS:** Six visually impaired participants received eight hours of training with the SoV system, completed a usability questionnaire, and repeatedly performed assessments, for which they navigated through standardized scenes. In each assessment, participants had to avoid collisions with obstacles, using the SoV system, the white cane, or both assistive devices.

**RESULTS:** The results show rapid and substantial learning with the SoV system, with less collisions and higher obstacle awareness. After four hours of training, visually impaired people were able to successfully avoid collisions in a difficult navigation task as well as when using the cane, although they still needed more time. Overall, participants rated the SoV system's usability favorably.

**CONCLUSIONS:** Contrary to the cane, the SoV system enables users to detect the best free space between objects within a 3.5-m (up to 10-m) radius and, importantly, elevated and dynamic obstacles. All in all, we consider that visually impaired people can learn to adapt to the haptic-auditory representation and achieve expertise in usage through well-defined training within acceptable time.

*Optom Vis Sci* 2018;95:757–765. doi:10.1097/OPX.0000000000001284

Copyright © 2018 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Optometry. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

**Supplemental Digital Content:** Direct URL links are provided within the text.



#### Author Affiliations:

- <sup>1</sup>Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland  
<sup>2</sup>Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik, Iceland  
<sup>3</sup>Department of Information Engineering, University of Padova, Padova, Italy  
 \*rebekkah@hi.is

Currently, 253 million people worldwide live with vision impairment, of which 38.5 million are totally blind.<sup>1</sup> This entails a major clinical and scientific challenge to develop effective visual rehabilitation methods to improve their independence and overall life quality. The white cane has been the primary assistive device for mobility, but it provides very limited information about the immediate environment and does not enable the visually impaired to plan the most efficient path through a cluster of obstacles beyond its reach.<sup>2</sup> During the last two decades, new technologies have been developed aimed at filling this gap and advancing traditional assistive navigation.<sup>3,4</sup> Invasive technologies such as artificial retinal prostheses are costly and have, so far, only resulted in low-resolution vision.<sup>5</sup> Sensory substitution devices are a promising alternative. They are noninvasive human-machine interfaces that draw on the central nervous system by bypassing the nonfunctioning visual system, transforming visual information into auditory or tactile information.<sup>6,7</sup> Sensory substitution devices challenge developers and designers because many issues must be considered if the visually impaired are to be expected to use a particular device, for example, usability, portability, comfort, real-time and long-term operation,

accessibility of interface, and appearance.<sup>7,8</sup> Sensory substitution devices based on computer vision, as opposed to ultrasonic or radar methods, have been developed,<sup>9–11</sup> but most are still prototypes or not widely used because of a lack of functionality, low ergonomics, lack of end-user involvement during development, and high costs.<sup>8</sup> Despite the increasing number of assistive technologies, most solutions fail to provide satisfying independence beyond known environments that would substantially increase the mobility of visually impaired people<sup>12</sup> and solve only parts of the problems faced by the visually impaired during wayfinding tasks.<sup>3</sup>

In the Sound of Vision project,<sup>13,14</sup> a sensory substitution device for visually impaired people was developed, which provides a continuous real-time multisensory representation of the environment. The acquisition system of the Sound of Vision device scans the environment and creates a three-dimensional model that encodes the relevant elements into both haptic and audio cues. The challenges of multisensory integration of haptics and audio, such as sensory overload and interference, are discussed in previous work.<sup>15</sup> Besides developing a sensory substitution device, a key aim was to develop a training along with learning software to

effectively train the visually impaired in using the Sound of Vision system for mobility, because there is good evidence that training is important for distal attribution or externalization.<sup>7,16</sup>

Our aims were to evaluate the efficiency of the Sound of Vision system for mobility, compare it with the white cane, and quantify training effects. Visually impaired participants received training on the Sound of Vision system with repeated performance assessments, where they were asked to navigate through standardized scenes. During navigation, participants relied on the assistive device(s) to avoid collisions with obstacles. We counted the number of collisions, correctly identified obstacles, and completion time.

## METHODS

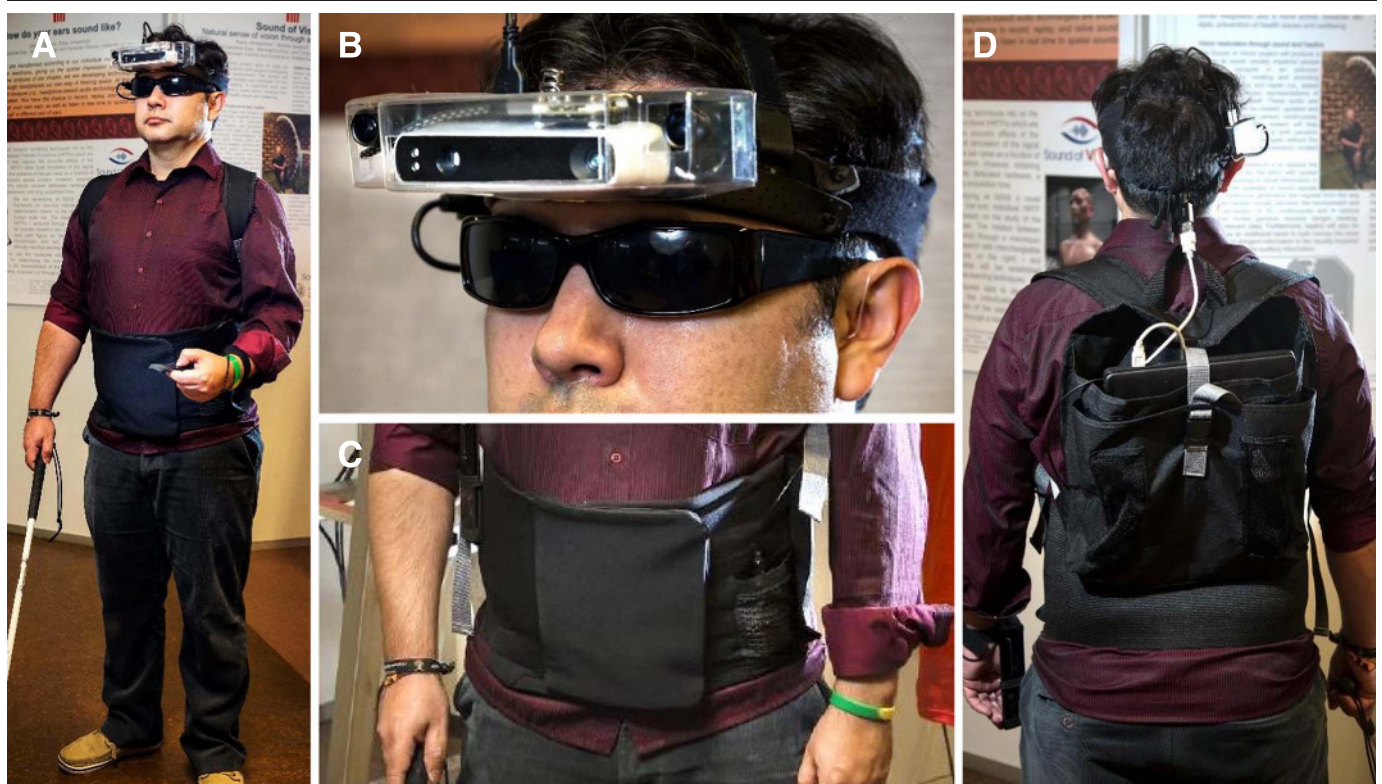
### Participants

Six visually impaired participants were tested (two females; age, 29 to 46 years; average,  $34.67 \pm 7.39$  years). According to the definition for visual impairment by the World Health Organization,<sup>17</sup> two participants belonged to category 3, meaning that visual acuity is between 3/60 and 1/60 (finger counting) or that the radial visual field is smaller than  $10^\circ$ . Two participants belonged to category 4, meaning that the visual acuity is less than 1/60 or that only light perception is possible. Finally, two participants were totally blind with no light perception (category 5). Four were born with visual impairment (one of them blind), two became visually impaired between the age of 10 and 15 years, and all were trained and

experienced white cane users. Participants from categories 3 and 4 were blindfolded (totally blind) to simulate environments with inadequate lighting or poor contrast, which would force them to rely on assistive devices. The participants were recruited from the Icelandic National Institute for Blind and gave written informed consent. The experiments conformed to the Declaration of Helsinki and were approved by the National Bioethical Committee of Iceland (VSN-15-107). The Icelandic community of visually impaired is small, so we do not list specifics by individual and refer to all participants as male independently of their sex, to preserve anonymity.

### Apparatus

The Sound of Vision system includes customized hardware and software allowing real-time operation. As shown in Fig. 1, the system comprises the following: a three-dimensional camera unit worn on the head (panel A), a pair of in-ear headphones allowing environmental sounds to enter the ear canal (panel B), a haptic belt with a matrix of  $6 \times 10$  vibrating motors placed on the abdomen (panel C), and a processing unit carried on the back (Lenovo Ideapad Y700 laptop, panel D). The system scans the environment through cameras while tracking both head and body position by means of two inertial measurement units, placed on head and shoulder, respectively. Although the full Sound of Vision system includes many different functionalities (e.g., object recognition, detecting dangerous objects, and reading text), only a small number of functionalities were used here: depth information collected from a high-performance three-dimensional Structure Sensor



**FIGURE 1.** The Sound of Vision (SoV) system. (A) A participant equipped with the complete SoV system. (B) The headgear with three-dimensional cameras and sound-permeable in-ear earphones (as required for outdoor navigation). (C) The tactile belt, with 60 vibrating motors at the abdomen. (D) The processing unit (laptop) carried inside a backpack.

camera and rendered in real time through auditory and haptic signals with two encoding mechanisms, the *fluid flow* model and the *closest point*, respectively.<sup>18,19</sup> The  $640 \times 480$ -pixel depth maps, which span the camera's entire field of view ( $58^\circ$  horizontal,  $45^\circ$  vertical, and a 0.4- to 3.5-m depth range) with high depth precision ( $<1\%$ ), are acquired at 10 frames per second. Custom software converts the depth map into sound and vibrations. The total system lag (from depth map collection to the audio-haptic representation) was less than 100 milliseconds.

### Audio Encoding

The audio encoding is based on the head position measured by the head-mounted inertial tracker. The fluid flow sound model is designed for constantly changing scenarios, relying on continuous depth information from the cameras. Complex liquid sounds are created from a population of bubble sounds defined from an empirical phenomenological model of bubble statistics.<sup>20</sup> The acquired visual scene is divided into 15 equally sized rectangular sectors ( $3 \times 5$ ). Depth information within each sector is mapped into bubble sound features, whereas the direction of the sector is mapped into spatial sound features. The higher the density of obstacles close to the user, the more bubbles sound, and the closer the obstacles, the higher the sound intensity.<sup>21</sup>

### Haptic Encoding

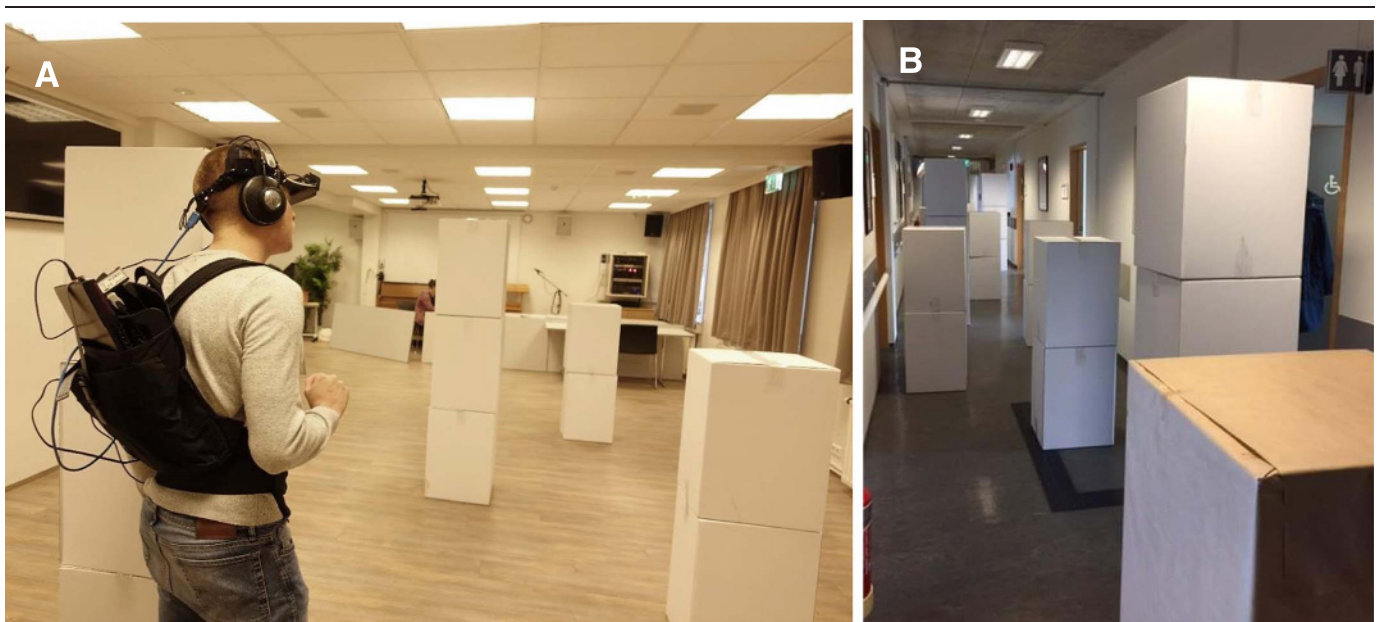
The haptic encoding is based on the users' body position in relation to their head position, so the haptic representation of the location of stationary objects is kept constant when users stay steady and only rotate their heads. For the closest point haptic model, the direction of the surface nearest to the user within a 3.5-m radius is haptically represented, through activation of the spatially corresponding motor in the  $6 \times 10$  motor array of the haptic belt. Frequency and amplitude (intensity) of the activated motor are

inversely proportional to object distance, with higher vibration intensity for closer objects. To amplify the vibrotactile feedback, the motor representing the closest cell in the depth map is augmented with neighboring motors, resulting in a  $2 \times 2$  array of activated motors.

### Training

Every task in the training program was designed to enable participants to correctly interpret the audio and haptic signals and to use the multisensory feedback for navigation. The training time was based on previous usability studies with the Sound of Vision system. During training (and testing), participants wore over-ear headphones with perforated ear cups for minimal sound attenuation, which allowed them to hear the target while focusing on the audio feedback (Fig. 2).

The first four hours involved training in a virtual environment, which allowed participants to get acquainted with the Sound of Vision system through carefree exploration without distractors or the need to coordinate their body. Participants sat at a table with a keyboard and joystick in front of them while wearing the Sound of Vision system. The virtual environment software ran on a laptop, to which the Sound of Vision system was connected, and the Sound of Vision system conveyed the virtual scenes rendered by the software, instead of using data from the camera. Participants first learned to distinguish properties of single objects (e.g., size, direction, distance), followed by properties of multiple objects. Subsequently, using a joystick, participants actively navigated through virtual scenes of increasing difficulty (e.g., finding one, passing between two, or numerous obstacles). Participants had the option to learn (actively manipulating the scene), to train (receiving feedback on their responses in a computer-generated task), and, eventually, to test themselves (without feedback). A detailed description of the virtual



**FIGURE 2.** Training and testing complex navigation with a sensory substitution device. Panel (A) shows a visually impaired participant during training equipped with the Sound of Vision system. He is wearing over-ear headphones with minimal sound-attenuation allowing to hear the target. Panel (B) shows an example navigation scene setup for assessing performance.

training apparatus and tasks can be found in the study of Moldoveanu et al.<sup>22</sup>

The next four hours of training took place in real-world settings with cardboard boxes as obstacles. The training started with the exploration of an empty room, followed by a single obstacle of varying height. The obstacle was made of cardboard boxes (each 40 × 40 cm wide/deep), where either one (60-cm high), two (120-cm high), or three (180-cm high) boxes were placed on top of one another. Participants were encouraged to systematically explore the system output when changing their own position in relation to the obstacle (approach/diverge, circle, etc.). Afterward, participants trained on scenes with a single obstacle by pointing at its direction, judging distance and size. A similar approach was used for two and three obstacles but with the additional task of passing between them. Finally, participants were trained on navigation scenes similar to the performance assessments. They were asked to find their way to a loudspeaker playing music on the opposite side of the room while avoiding collisions with three to nine randomly placed obstacles and the walls (Fig. 2A). Beyond learning to interpret the audio-haptic signals, the training was aimed at enhancing general navigation skills. Participants learned to scan the environment horizontally to assess the layout of objects and to scan objects vertically to infer height. Participants also became acquainted with the range of the Sound of Vision system (0.4 to 3.5 m) and learned how to exploit this for navigation.<sup>23</sup>

## Experimental Setup

The navigation scenes have to be of comparable difficulty but also need to be changed between trials because any performance increase might otherwise reflect memory of scene layout. The scenes should also be difficult to avoid ceiling effects that limit variance in the performance. Detailed rules for setting up a standardized navigation scene with random properties within a defined range were therefore determined beforehand. Each of the five assessments (including baseline) were conducted in the same setting: 15-m distance between starting point and target (in a straight corridor, 190 cm of width) and a constant number of 10 obstacles, where three were high (180 × 40 × 40 cm) and seven were low (120 × 40 × 40 cm). Obstacle location varied randomly between scenes, but either their placement always left a free passageway of at least 100 cm (or 20 cm wider than the participants' width) around each obstacle or the path on one side was blocked by a wall (Fig. 2B).

## Procedure

The participants underwent extensive training for four weeks, with training sessions of two hours, one to two times per week, and repeated performance assessments. Experienced mobility instructors assisted in all sessions.

Data from initial performance assessments were used as baseline where participants completed the navigation scene three times, once for each of the following conditions: (a) the Sound of Vision system only, (b) the Sound of Vision system and the white cane, and (c) the white cane only. During each assessment, participants were instructed to navigate as directly as possible toward the target sound, placing highest priority on avoiding collisions with obstacles. To avoid accidental contact, participants were instructed to keep their arms close to their bodies. Participants were naive about scene layout and setup rules. After participants completed the first two hours of virtual training, their performance was reevaluated, where they repeated the navigation scene two times,

(a) with the Sound of Vision system and (b) with the Sound of Vision system and the white cane. After another two hours of virtual training, performance was assessed in the same way for a third time. Subsequently, participants trained for two hours in real-world settings, followed by the fourth performance assessment. After another two hours of real-world training, performance was assessed for the fifth time. Overall, this procedure resulted in eight hours of training, five assessments of the Sound of Vision system, five assessments of the Sound of Vision system and the white cane, and one assessment of the white cane per participant.

Several performance measures were used. For the Sound of Vision system (a), the hands were free, so participants were asked to verbally indicate the presence of an obstacle, followed by a pointing gesture toward it, to countercheck if obstacles were successfully passed collision-free because of obstacle awareness or by coincidence. Each of the 10 obstacles per scene was rated as either a *Hit* (correctly reported) or a *Miss* (not reported). Completion time was measured, and collisions were counted as follows: major collisions (with nonreported obstacles) and minor collisions (with previously reported obstacles or minor lateral brushes). In conditions with the white cane, participants were instructed to use the cane normally but to avoid body contact with obstacles. The number of body collisions with obstacles and missed obstacles (neither touched by cane nor participant) were assessed, but cane hits were not counted as collisions. Because participants did not have their hands free in conditions involving the white cane, obstacle awareness was not assessed.

## User Feedback

All participants answered a user feedback questionnaire at the end of the last testing session. The questionnaire comprised 11 items with a five-level Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Items 1 to 10 were taken from the System Usability Scale to assess effectiveness, efficiency, and satisfaction,<sup>24</sup> and item 11 assessed the system's potential for enhancing leisure activities (see Appendix Table 1, available at <http://links.lww.com/OPX/A356>).<sup>25</sup>

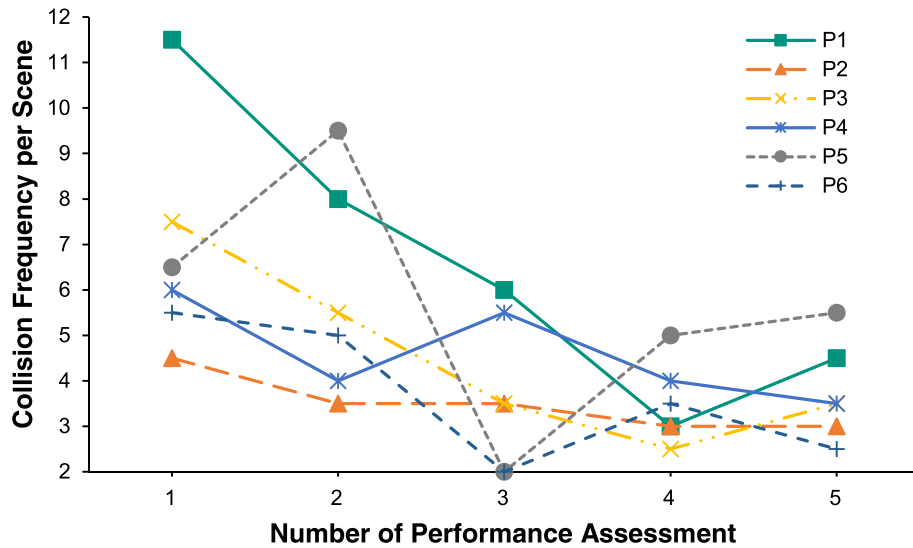
## Data Analysis

For all analyses (conducted in R),<sup>26</sup> major and minor collisions were merged into one outcome variable, referred to as *collisions*, where the former is weighted with 1.0 and the latter with 0.5. The group medians are reported, complemented by the interquartile range ( $Q_3$  to  $Q_1$ ). For each group comparison, the estimate of the shift in locations, meaning the median of the differences between the groups based on the Wilcoxon signed rank test (before vs. after training) and the Whitney-Mann  $U$  test (comparing assistive devices),<sup>27</sup> is reported, along with the associated confidence intervals (exact level).

# RESULTS

## Training Improvement

(a) Sound of Vision system only. Overall, performance improved with training. Collisions by participant are shown in Fig. 3. Median collision frequency decreased from before training (median, 6.25; interquartile range, 2.0) to after training (median = 3.50; interquartile range, 1.5), with a median of differences of 2.75 (95% confidence interval, 1.0 to 7.0). The number of correctly identified



**FIGURE 3.** Progress of individual obstacle avoidance with the Sound of Vision (SoV) system. The course of collision frequency with obstacles per scene when using the SoV system, plotted for each individual as a function of training progress. Participants trained two hours with the SoV system between each performance assessment.

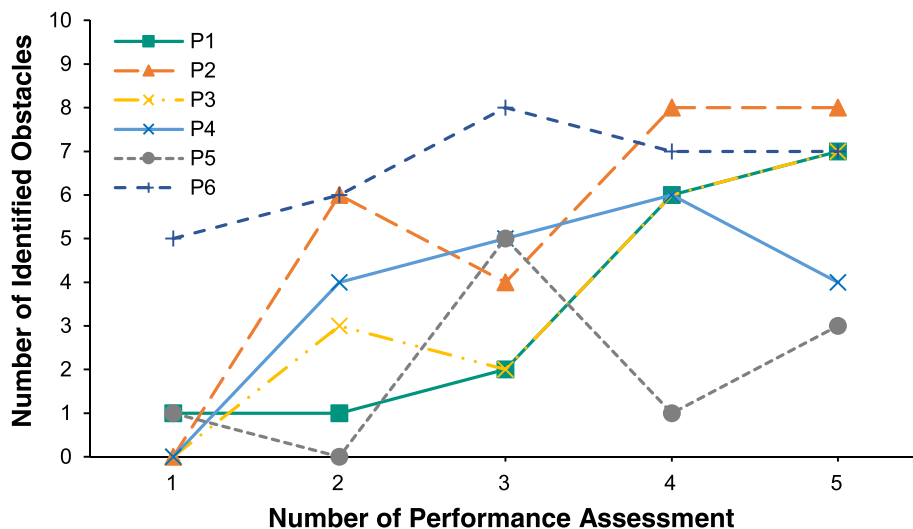
obstacles per participant is shown in Fig. 4 and increased from before (median, 0.50; interquartile range, 1.0) to after training (median, 7.0; interquartile range, 3.0; median of the differences, 5.0; 90% confidence interval, -7.0 to -2.0). After four hours of training (third assessment), the collision frequency reached a plateau of 3.25 to 3.50 per scene but not the number of correct obstacle identifications. There was no effect of completion time when comparing before (median, 170.0; interquartile range, 39.0) with after training (median, 177.0; interquartile range, 91.0; see Appendix Fig. A1, available at <http://links.lww.com/OPX/A357>).

(b) Sound of Vision system and white cane. Participants did not show a similar performance improvement as with Sound of Vision system only (see collision medians in Fig. 5). The number of collisions did not decrease from before (median, 4.0; interquartile

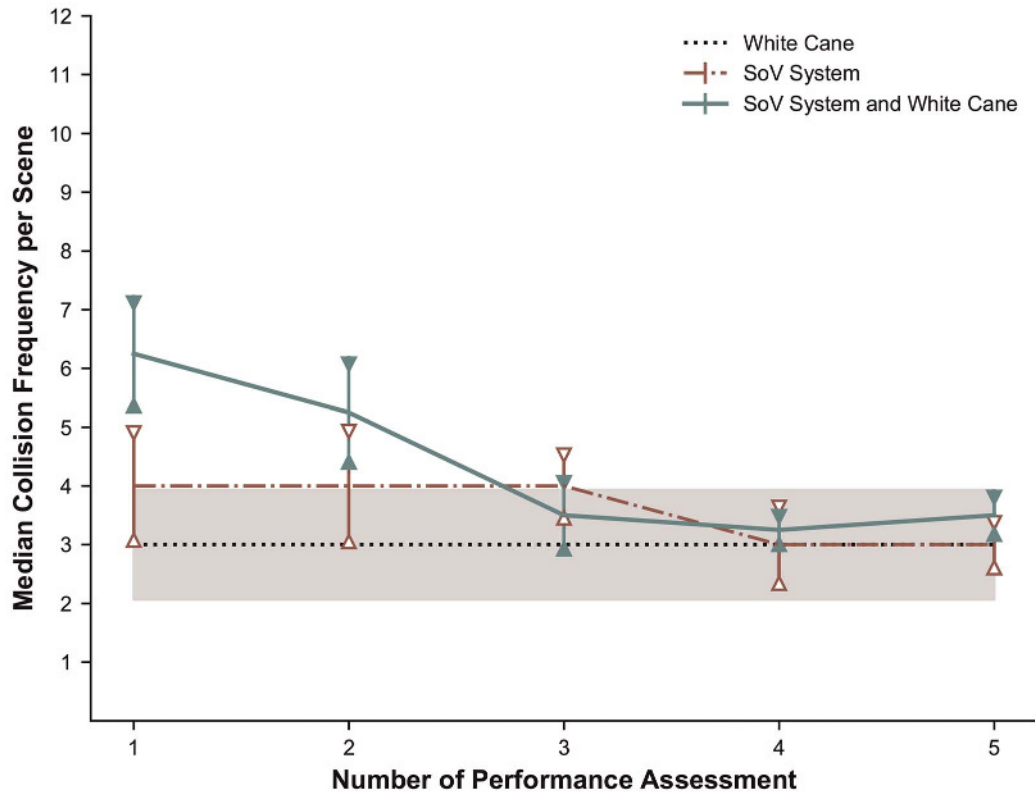
range, 3.0) to after training (median, 3.0; interquartile range, 2.0), with a median of differences of 0.80 (80% confidence interval, -2.50 to +4.0). The course of collision frequency per participant with both assistive devices is presented in Appendix Fig. A2, available at <http://links.lww.com/OPX/A357>. Completion times also did not change from before (median, 71.0; interquartile range, 33.0) to after training (median, 74.5; interquartile range, 49.0).

### Assistive Device Comparison

Fig. 5 plots the median collision frequency per scene for the three assistive devices. As expected, participants performed best with (c) the white cane, with a median of 3.0 (interquartile range, 1.0) collisions and completion time of 55.0 seconds (interquartile range, 24.0 seconds).



**FIGURE 4.** Progress of individual obstacle awareness with the Sound of Vision (SoV) system. The number of obstacles per scene that were correctly identified (verbally reported before colliding or passing) by each participant when using the SoV system as a function of training progress. Participants trained two hours with the SoV system between each performance assessment.



**FIGURE 5.** Comparison of obstacle avoidance progress between assistive devices. Median collision frequency per scene for the three assistive device conditions: Sound of Vision (SoV) system only, white cane only, and both assistive devices. Between each performance assessment, participants received two hours of training with the SoV system. Error bars show the standard error of the median. Because performance with the white cane was assessed once, its standard error of the median is depicted as assessment-spanning area rather than point errors.

More interestingly, however, before training, performance was lowest with the Sound of Vision system only, indicated by more collisions (median, 6.25; interquartile range, 2.0) relative to the white cane (median of the differences, -3.50; 95% confidence interval, -7.50 to -1.50). The number of collisions with both assistive devices (median, 4.0; interquartile range, 3.0) compared with the white cane only, however, did not differ (median of the differences, -1.0; 95% confidence interval, -5.0 to +2.0). The same pattern was found when comparing completion time with the white cane to the other conditions. Before training, the time was comparable with the condition with both assistive devices (median, 71.0; interquartile range, 33.0), with a median of differences of 17.0 (95% confidence interval, -16 to +77) but was higher in the Sound of Vision system only condition (median, 170.0; interquartile range, 39.0; median of the differences, -115.37; 95% confidence interval, -183.0 to -42.0).

Fig. 5 shows a trend for improved performance with the Sound of Vision system until the median collision frequency per scene aligns for all three conditions. At the third assessment, the number of collisions with the Sound of Vision system (median, 3.5; interquartile range, 2.5) did not differ from the number of collisions with the white cane (median of the differences, -0.5; 95% confidence interval, -3.50 to +1.0). The number of collisions with both assistive devices (median, 4.0; interquartile range, 0.0) was still comparable with the white cane (median of the differences, -1.0; 95% confidence interval, -4.0 to +1.0). Finally, no participant completed

any navigation scene (also when using the white cane) collision free or with the maximum rate of identified obstacles.

### User Feedback

No participant judged the Sound of Vision system to be unnecessarily complicated (item 2) or cumbersome to use (item 8). Although three participants found the ease of use to be rather low (item 3), there was strong variance in participants' opinions, ranging from 1 to 5. Most participants felt like they would need the support of a technical person to use the system (item 4). One participant, however, was very positive about the ease of use and did not expect to need assistance. Most participants agreed that the functions of the system were well integrated and consistent (items 5 and 6). Whereas three participants felt that they had to learn a lot to be able to use the system, two had the opposite opinion (item 10). Except for one participant, all estimated that visually impaired participants would learn to use the Sound of Vision system very quickly (item 7). All participants (except for one participant being neutral) felt very confident in using the Sound of Vision system (item 9). Four participants would like to use the Sound of Vision system frequently (item 1), and three participants agreed that the system would enhance their capacity for leisure activities (item 11). The complete results can be found in Appendix Table 1, available at <http://links.lww.com/OPX/A356>.

## DISCUSSION

Our aims were to compare performance with the Sound of Vision system to the white cane, quantify training progress, assess mobility with the system, and collect user feedback. The results indicate that with increased training participants achieved better performance, with fewer obstacle collisions and higher obstacle awareness. After four hours of training, participants using the Sound of Vision system were able to complete the navigation scenes with an equally low number of collisions as with the white cane. Training using the Sound of Vision system results in surprisingly rapid improvements in mobility, which quickly reaches performance levels seen with the white cane.

There was no consistent trend for completion time, and more training with the Sound of Vision system would be required to decrease the completion time to a comparable level with the white cane. As indicated by Spagnol et al.,<sup>23</sup> time does not appear to be a meaningful metric for measuring navigation performance, because it needs to be interpreted in context. Also, we emphasized accuracy rather than speed. Paradoxically, participants tend, if anything, to slow down once they gain an understanding of the audio-haptic representation, trying to apply lessons learned from training.

### Individual Differences

When evaluating the Sound of Vision system, it is important to consider individual performance differences. Some variation may reflect participants' backgrounds, their responses on the usability questionnaire, and their feedback throughout the study.

Participant P05 might represent a subgroup of visually impaired that cannot be easily convinced of using sensory substitution devices. The participant is totally blind, yet socially well integrated and is particularly skilled in using various assistive devices. The training with the new Sound of Vision system progressed relatively slowly. This may partly reflect that, over many decades, he has gotten highly accustomed to his way of encountering daily challenges and therefore might have doubted the value of investing time in mastering a new assistive device. Although he was repeatedly instructed otherwise, he tried to complete the assessments as fast as possible, as individual completion times reveal (Appendix Fig. A1, available at <http://links.lww.com/OPX/A357>). His frequent collisions understandably increased frustration and lowered motivation for further training. Accordingly, he did not confirm to feel confident with the Sound of Vision system and criticized it for being too complicated and bulky, stating that he would prefer to ask friends or coworkers for assistance.

Remarkably, P02 and P04 chose to rely on the Sound of Vision system instead of the white cane when provided with both. Neither is totally blind, which tends to result in less experience with the white cane. They had a favorable attitude toward the Sound of Vision system and were keen to explore its possibilities. All other participants, however, chose to use the white cane rather than the Sound of Vision system when both were available, which is reflected in the assessment outcomes. The white cane only condition resulted in the lowest number of collisions and fastest completion times. Although the number of collisions and completion time between (b) both assistive devices and (c) white cane only were similarly constant across all assessments, the condition (a) with the Sound of Vision system revealed a different pattern, with initially more collisions that significantly decreased with training. It

is important to highlight that eight hours of training with the Sound of Vision system is not comparable with years of training with the white cane.

Participant P06 might represent a group of visually impaired that could benefit most from using the Sound of Vision system. He has been totally blind from birth, and despite good experience with the cane, he prefers navigating without it. He is young and highly engaged in social and leisure activities but limits his independent navigation to familiar environments. Using the Sound of Vision system seemed intuitive for him, which is reflected in very few collisions and high obstacle awareness already in his baseline assessment. He was eager to learn, and the training proceeded fast, resulting in the highest performance, with him stating, "[When using the Sound of Vision system,] I feel as if I can actually see the things around me." Accordingly, he rated the system as being easy to use and learn, confirmed he would like to use it frequently and that it would enhance his numerous leisure activities. After the study and two more training hours, he mastered complex navigation scenes with 13 realistic obstacles (steady pedestrians, suitcases, plants, chairs, etc.) without collisions.

### Limitations

An obvious shortcoming of repeating a similarly constructed navigation scene, even if randomized, is that the improved performance in our experiment might stem from learning the principles of scene construction. Note, however, that the number of collisions did not decrease when the navigation scene was performed with the white cane and the Sound of Vision system, indicating that users' improvement reflected learning of the system rather than learning of scene setup. If the latter was the case, performance should have improved in all conditions.

Although no scene was mastered without collisions, it is important to note that the scenes were designed to be very difficult. Previous experiments with the Sound of Vision system revealed that participants mastered navigation tasks of lower difficulty without collisions after only two hours of training with the Sound of Vision audio encoding.<sup>23</sup> With much more training planned for the current study, precautions were taken to avoid a ceiling effect, limiting variance in performance. Obstacles were placed so close that even sighted persons had to slow down while navigating the scenes, and in the cane-only condition, participants still collided with obstacles.

Although the reported findings are highly promising, caution should be applied owing to the limited number of participants. Note also that the average age of our group was 35 years, and the oldest participant was aged 46 years. Generalization to older groups therefore requires caution; however, 81% of people who suffer from vision impairment are older than 50 years.<sup>28</sup> Jóhannesson et al.<sup>29</sup> demonstrated vibrotactile spatial acuity of 13 mm for the motor type used in the Sound of Vision system for 20- to 26-year-old participants. Assuming that tactile acuity decreases by 1% annually,<sup>30</sup> the 30-mm intermotor distance in the Sound of Vision system should enable sufficient tactile acuity for older visually impaired. The highly prevalent problem of age-related hearing loss compromising accurate perception of audio signals in sensory substitution devices<sup>31</sup> might therefore be compensated for by tactile encoding. In addition, general cognitive functions for learning and integrating multisensory feedback decrease in old age, even though neural plasticity still occurs.<sup>32</sup> Overall, Sound of Vision should be introduced to young and middle-aged visually impaired people so that long-term use could compensate for decreasing cognitive functions in aging.

Although not all participants showed a large increase in performance after a few hours of training, they may master the Sound of Vision system with longer training. As stated earlier, learning to confidently navigate with the white cane requires several months of training, and this probably also applies to mastering a sensory substitution device. In fact, such devices typically involve a training program, like the vOICE system,<sup>33</sup> which requires up to 1 year of training.<sup>34</sup> Note that Spagnol et al.<sup>23</sup> directly compared performance between the vOICE system and the audio encoding of the Sound of Vision system in the same navigation task. Their results reveal faster learning and higher performance with the audio encoding of the Sound of Vision system than the vOICE system.

When navigating, visually impaired people face a particular challenge with the detection of dynamic obstacles (pedestrians, bicycles, and vehicles). Even though our study did not include moving objects, we believe that additional training with the Sound of Vision system would enable users to recognize dynamic objects, providing a crucial advantage over the white cane. Unlike other sensory substitution devices, like the vOICE system that is based on left-right scans,<sup>33</sup> the Sound of Vision system facilitates the tracking of dynamic objects through real-time representations. Furthermore, the Sound of Vision system tracks both head and body position so that, when participants scan their environment by head rotation while standing, dynamic objects are represented as a cascade of motor activations, with participants reporting that they felt a sweep of vibrations along the torso when pedestrians walked past.

Many systems for supporting the mobility of the blind have been developed, offering a wide range of solutions for specific problems, for instance, TapTapSee (a mobile camera application that photographs and recognizes objects),<sup>35</sup> Trekker Breeze+ (a handheld talking Global Positioning System [GPS]),<sup>36</sup> the Drishti navigation system (using a series of maps with associated information),<sup>37</sup> and TANIA (a portable GPS device with movement sensors, providing acoustic information about position or distance).<sup>38</sup> Even though such systems are very useful, they are limited to known environments because they require either a database to store the environment information or hardware and software components placed in predefined locations (bus stations, train stations) that interact with the wearable assistive device. In contrast, the Sound of Vision system requires no maps or databases because it renders the current environment to the user in real time. Furthermore, in contrast with many existing solutions, it integrates solutions for a wide range of problems, having modes for specific object recognition

(stairs, doors, holes in the ground), text reading, and collision danger warning, and can be used indoors and outdoors, in daylight or at night.

Furthermore, one issue facing many substitution devices is that the modality used for bypassing the nonfunctioning sense may not have the required spatiotemporal resolution necessary to convey sufficient information, which is especially relevant for the sense of touch.<sup>29,30</sup> To the best of our knowledge, the Sound of Vision system is the only sensory substitution device involving multisensory feedback. Although challenges of multisensory integration need to be addressed,<sup>15</sup> audio-haptic representations may be a promising solution to compensate for the lower bandwidths of these modalities compared with the visual sense.

Another challenge for many sensory substitution devices is that, despite successful proofs of concept, the solutions are not widely used. To be accepted by the blind community, the Sound of Vision system should be scalable to run on systems with low processing power, such as smartphones. This might be achieved if the system directly receives reliable information conveyed through an off-the-shelf depth sensor. We are currently investigating the experiential quality with limited computing power. Future versions could include enhanced functions, such as GPS-based navigation, to further support mobility, and the haptic belt could comprise thinner piezoelectric actuators that provide a wider range of frequencies with high acceleration. Smaller versions of the cameras will become available soon, allowing a subtler design resembling a baseball cap, with the option for individual customization of color, fabric, and fit.

## CONCLUSIONS

Overall, our findings provide encouraging evidence for the usability of the Sound of Vision system. We believe that visually impaired users can adapt to a haptic-auditory representation of their environment and achieve expertise in usage through well-defined training within adequate time frames. In contrast to the white cane, the Sound of Vision system enables visually impaired people to determine obstacle-free spaces between objects and identify dynamic and elevated obstacles, substantially increasing spatial awareness.<sup>2</sup> The Sound of Vision system can therefore lead to a more active lifestyle and improved wellbeing for many visually impaired individuals.

## ARTICLE INFORMATION

**Supplemental Digital Content:** Appendix Table A1 (available at <http://links.lww.com/OPX/A356>): Complete results of the User Feedback Questionnaire per Participant. After study completion, all participants answered a user feedback questionnaire. The questionnaire items and the responses of the six participants (P1–P6) are shown in the table. The Likert scale response options were as follows: *I strongly disagree (1), I disagree (2), I am undecided (3), I agree (4), and I strongly agree (5)*.

Appendix Figure A1 (available at <http://links.lww.com/OPX/A357>): Individual completion time when navigating with the Sound of Vision (SoV) system. The completion time (in seconds) with the SoV system that was needed by each individual per scene as a function of training progress (x). Participants trained two hours with the SoV system between each performance assessment. Appendix

Figure A2 (available at <http://links.lww.com/OPX/A357>): Individual obstacle avoidance when navigating with the Sound of Vision (SoV) system and the white cane. The course of collision frequency with obstacles per scene (y) when using the SoV system and the white cane, plotted per individual as a function of training progress (x). Participants trained two hours with the SoV system between each performance assessment.

**Submitted:** February 26, 2018

**Accepted:** July 3, 2018

**Funding/Support:** European Union Horizon 2020 Research and Innovation Program (643636).

**Conflict of Interest Disclosure:** None of the authors have reported a financial conflict of interest.

**Author Contributions:** Conceptualization: RH, SS, ÁK, RU; Data Curation: RH; Formal Analysis: RH; Funding Acquisition: ÁK, RU; Investigation: RH, SS; Methodology:

RH; Project Administration: ÁK, RU; Resources: ÁK, RU; Software: SS; Supervision: ÁK, RU; Validation: RH, SS; Visualization: RH; Writing – Original Draft: RH; Writing – Review & Editing: SS, ÁK, RU.

## REFERENCES

1. World Health Organization (WHO). Media Centre — Visual Impairment and Blindness Fact Sheet No. 282. Available at: <http://www.who.int/mediacentre/factsheets/fs282/en/>. Accessed May 31, 2018.
2. Wiener WR, Welsh RL, Blasch BB. Foundations of Orientation and Mobility, Vol. 1, 3rd ed. New York, NY: American Foundation for the Blind; 2010.
3. Caraiman S, Moldoveanu BA, Morar A, et al. Computer Vision for the Visually Impaired: The Sound of



- Vision System. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2017;1480–9.
4. Bhowmick A, Hazarika SM. An Insight into Assistive Technology for the Visually Impaired and Blind People: State-of-the-art and Future Trends. J Multimodal User Interfaces 2017;11:149–72.
  5. Striem-Amit E, Bubic A, Amedi A. Neurophysiological Mechanisms Underlying Plastic Changes and Rehabilitation Following Sensory Loss in Blindness and Deafness. In: Murray MM, Wallace MT, eds. The Neural Bases of Multisensory Processes. Boca Raton, FL: CRC Press/Taylor & Francis; 2012: Chapter 21.
  6. Reich L, Maidenbaum S, Amedi A. The Brain as a Flexible Task Machine: Implications for Visual Rehabilitation Using Noninvasive vs. Invasive Approaches. Curr Opin Neurol 2012;25:86–95.
  7. Kristjánsson Á, Moldoveanu A, Jóhannesson ÓI, et al. Designing Sensory-substitution Devices: Principles, Pitfalls and Potential. Restor Neurol Neurosci 2016; 34:769–87.
  8. Dakopoulos D, Bourbakis NG. Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. IEEE Trans Syst Man Cybern (C) Appl Rev 2010;40: 25–35.
  9. Costa P, Fernandes H, Martins P, et al. Obstacle Detection Using Stereo Imaging to Assist the Navigation of Visually Impaired People. Procedia Comput Sci 2012;14:83–93.
  10. Filipe V, Fernandes F, Fernandes H, et al. Blind Navigation Support System Based on Microsoft Kinect. Procedia Comput Sci 2012;14:94–101.
  11. Ribeiro F, Florencio D, Chou PA, et al. Auditory Augmented Reality: Object Sonification for the Visually Impaired. IEEE Multimed Signal Proc 2012; 319–24.
  12. Giudice NA, Legge GE. Blind Navigation and the Role of Technology. In: Helal A, Mokhtari M, Abdulrazak B, eds. The Engineering Handbook of Smart Technology for Aging, Disability, and Independence. New York, NY: John Wiley & Sons, Inc; 2008:479–500.
  13. Sound of Vision Project (SoV). Available at: <https://soundofvision.net/>. Accessed June 22, 2018.
  14. Strumiňto P, Bujacz M, Baranski P, et al. Different Approaches to Aiding Blind Persons in Mobility and Navigation in the “Naviton” and “Sound of Vision” Projects. In: Pissaloux E, Velazquez R, eds. Mobility of Visually Impaired People. New York, NY: Springer; 2018: 435–68.
  15. Jóhannesson ÓI, Balan O, Unnthorsson R, et al. The Sound of Vision Project: On the Feasibility of an Audio-haptic Representation of the Environment, for the Visually Impaired. Brain Sci 2016;6:20.
  16. Hartcher-O'Brien J, Auvray M. The Process of Distal Attribution Illuminated through Studies of Sensory Substitution. Multisens Res 2014;27:421–41.
  17. World Health Organization. The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva, Switzerland: World Health Organization; 1992.
  18. Janeczek M, Skulimowski P, Owczarek M, et al. Adaptive Edge-based Stereo Block Matching Algorithm for a Mobile Graphics Processing Unit. Sig P Algo Arch 2017;201–6.
  19. Csapo A, Spagnol S, Herrera Martínez M, et al. Usability and Effectiveness of Auditory Sensory Substitution Models for the Visually Impaired. Audio Eng Soc Convention 2017;142.
  20. Van den Doel K. Physically-based Models for Liquid Sounds. ACM Trans Appl Percept 2005;2:534–46.
  21. Spagnol S, Baldan S, Unnthorsson R. Auditory Depth Map Representations with a Sensory Substitution Scheme Based on Synthetic Fluid Sounds. IEEE Workshop Multimed Signal Proc 2017;1–6.
  22. Moldoveanu AB, Ivascu S, Stanica I, et al. Mastering an Advanced Sensory Substitution Device for Visually Impaired through Innovative Virtual Training. IEEE Consumer Electronics-Berlin 2017;120–5.
  23. Spagnol S, Hoffmann R, Herrera Martínez M, et al. Blind Wayfinding with Physically-based Liquid Sounds. Int J Hum Comput Stud 2018;115:9–19.
  24. Brooke J. SUS: A Quick and Dirty Usability Scale. In: Jordan PW, Thomas B, Weerdmeester BA, et al, eds. Usability Evaluation in Industry. London, United Kingdom: Taylor & Francis; 1986:4–7.
  25. Ryu YS. Development of Usability Questionnaires for Electronic Mobile Products and Decision Making Methods [doctoral dissertation]. Virginia Polytechnic Institute and State University; 2005.
  26. Yang S, Berdine G. Non-parametric Tests. SW Respiratory Crit Care Chron 2014;2:63–7.
  27. Mundry R, Fischer J. Use of Statistical Programs for Nonparametric Tests of Small Samples Often Leads to Incorrect P Values: Examples from Animal Behaviour. Anim Behav 1988;56:256–9.
  28. World Health Organization (WHO). Fact Sheet: Blindness and Visual Impairment. Available at: <http://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment>. Accessed May 31, 2018.
  29. Jóhannesson ÓI, Hoffmann R, Valgeirsdóttir VV, et al. Relative Vibrotactile Spatial Acuity of the Torso. Exp Brain Res 2017;235:3505–15.
  30. Stevens JC, Patterson MQ. Dimensions of Spatial Acuity in the Touch Sense: Changes over the Life Span. Somatosens Mot Res 1995;12:29–47.
  31. Cruickshanks KJ, Wiley TL, Tweed TS, et al. Prevalence of Hearing Loss in Older Adults in Beaver Dam, Wisconsin: The Epidemiology of Hearing Loss Study. Am J Epidemiol 1998;148:879–86.
  32. Baltes PB, Lindenberger U. On the Range of Cognitive Plasticity in Old Age as a Function of Experience: 15 Years of Intervention Research. Behav Ther 1988;19: 283–300.
  33. Meijer PB. An Experimental System for Auditory Image Representations. IEEE Trans Biomed Eng 1992;39: 112–21.
  34. Pasqualotto A, Esenkaya T. Sensory Substitution: The Spatial Updating of Auditory Scenes “Mimics” the Spatial Updating of Visual Scenes. Front Behav Neurosci 2016;10:79.
  35. TapTapSee. Available at: <https://taptapseeapp.com/>. Accessed May 31, 2018.
  36. Trekker Breeze+ by Humanware. Available at: <http://store.humanware.com/hau/trekker-breeze-plus-handheld-talking-gps.html>. Accessed May 31, 2018.
  37. Helal A, Moore SE, Ramachandran B. Driшти: An Integrated Navigation System for Visually Impaired and Disabled. Proc IEEE Wearable Comput 2001;149–56.
  38. Hub A. Precise Indoor and Outdoor Navigation for the Blind and Visually Impaired Using Augmented Maps and the TANIA System. Proc Int Conf Low Vis 2008;2–5.