Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:40 09 November 2016 (PT)

Emerald Insight

Journal of Assistive Technologies

Multimodal and alternative perception for the visually impaired: a survey

Wai Lun Khoo Zhigang Zhu

Article information:

To cite this document:
Wai Lun Khoo Zhigang Zhu , (2016),"Multimodal and alternative perception for the visually impaired: a survey", Journal of Assistive Technologies, Vol. 10 Iss 1 pp. 11 - 26
Permanent link to this document:
http://dx.doi.org/10.1108/JAT-04-2015-0014

Downloaded on: 09 November 2016, At: 20:40 (PT)

References: this document contains references to 34 other documents.

To copy this document: permissions@emeraldinsight.com

The fulltext of this document has been downloaded 57 times since 2016*

Users who downloaded this article also downloaded:

(2016),"Smartphone use and internet literacy of senior citizens", Journal of Assistive Technologies, Vol. 10 Iss 1 pp. 27-38 http://dx.doi.org/10.1108/JAT-03-2015-0006

(2016),"HeadBanger: controlling switchable software with head gesture", Journal of Assistive Technologies, Vol. 10 Iss 1 pp. 2-10 http://dx.doi.org/10.1108/JAT-04-2015-0015

Access to this document was granted through an Emerald subscription provided by emerald-srm:563821 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Peer-reviewed paper

Multimodal and alternative perception for the visually impaired: a survey

Wai Lun Khoo and Zhigang Zhu

Wai Lun Khoo is Graduate Research Assistant and Zhigang Zhu is Professor, both at Department of Computer Science, City College of New York, New York, New York, USA.

Abstract

Purpose – *The purpose of this paper is to provide an overview of navigational assistive technologies with various sensor modalities and alternative perception approaches for visually impaired people. It also examines the input and output of each technology, and provides a comparison between systems.*

Design/methodology/approach – *The contributing authors along with their students thoroughly read and reviewed the referenced papers while under the guidance of domain experts and users evaluating each paper/technology based on a set of metrics adapted from universal and system design.*

Findings – *After analyzing 13 multimodal assistive technologies, the authors found that the most popular sensors are optical, infrared, and ultrasonic. Similarly, the most popular actuators are audio and haptic. Furthermore, most systems use a combination of these sensors and actuators. Some systems are niche, while others strive to be universal.*

Research limitations/implications – *This paper serves as a starting point for further research in benchmarking multimodal assistive technologies for the visually impaired and to eventually cultivate better assistive technologies for all.*

Social implications – *Based on 2012 World Health Organization, there are 39 million blind people. This paper will have an insight of what kind of assistive technologies are available to the visually impaired people, whether in market or research lab.*

Originality/value – *This paper provides a comparison across diverse visual assistive technologies. This is valuable to those who are developing assistive technologies and want to be aware of what is available as well their pros and cons, and the study of human-computer interfaces.*

Keywords *Games, Assistive technologies, Auditory, Haptic, Prostheses, Virtual environment*

Paper type *Literature review*

1. Introduction

In 2012, there were 285 million visually impaired people worldwide; 39 million of which were blind and the rest (246 million) had low vision. This is an approximately 77 percent increase from 2002 (161 million visually impaired people)[1][2]. As the world population ages, the number will only increase unless a significant breakthrough is found. In the USA alone, with aging baby boomers, the number will double[3]. Furthermore, as one gets older, he/she will eventually become impaired one way or another. The need for assistive technology (AT), therefore, is and will be there. More importantly, the need of an evaluation system and usability study for AT is imperative to assess the usefulness of AT, to provide scientific measurement of AT, and to establish benchmarks for heterogeneous systems, so that visually impaired people who are looking for AT in the market can make an informed decision.

ATs include a number of products and services that aid a person due to loss of autonomy (e.g. visual, physical, hearing, etc.) (Plos *et al.*, 2012). Commercial products such as UltraCane[4]

Received 18 April 2015
Accepted 31 August 2015

This work is supported by US National Science Foundation Emerging Frontiers in Research and Innovation Program under Award No. EFRI-1137172. The authors would like to thank CCNY CS/CpE/EE/Psy Joint Senior Design students for reading the papers and providing evaluation. The authors thank the following experts for their feedback: Dr William Seiple from Lighthouse Guild, Dr Edgardo Molina from Vista Wearable Inc., Ms Barbara Campbell, Mr Daniel Diaz, and Mr Masud Kalim from NYS Commission for the Blind.

DOI 10.1108/JAT-04-2015-0014

and Miniguide[5], are available to users to purchase. This paper, however, will focus on ATs that address navigation assistance for visually impaired that have not been commercialized, except for the newer ones such as Argus II and Brainport. Furthermore, there are overlapping similarities between the commercial products and research projects cited, but research projects have published results, unlike commercial products' white paper. It would be ideal to evaluate commercial products, but that's not our goal. In other words, this paper considers sensory substitution or alternative perception devices that transform visual information into non-visual signals (e.g. auditory and vibrotactile). There is a wide range of ATs available for visually impaired people ranging from simple to specialized, from commercial-off-the-shelf to medicalization. These ATs can be categorized into three categories: physical, digital, and medical. Since the advent of remote sensing devices and computing technology in the 1960s, scientists and engineers are able to build specialized devices (e.g. electronic travel aid, haptic gloves, etc.) for the visually impaired. Some of these are "alternative perception" devices. Another type of ATs is retinal prosthesis. Unlike alternative perception devices, retinal implants convert incident light ray into electrical signals, and deliver the signals to retinal neurons other than photoreceptors, since it is generally damaged as a result of an eye disorder (Weiland and Humayun, 2014; Weiland *et al*., 2011). The main advantage of retinal prostheses is partial vision restoration, albeit very low resolution (i.e. depends on the electrode array's dimension).

This paper first describes how AT systems are selected for this study, and what features are used to evaluate these systems in Section 2. Then, various ATs are discussed with respect to the selected features and grouped into three major categories: physical, digital, and medical, in Sections 3-5. In Section 6, a comparison study framework and its results are presented. Lastly, concluding remarks, limitations of this study, and expert's feedback are presented.

2. The methodology of the survey

This paper primary focus is the study of AT's inputs (sensors) and outputs (actuators) in order to investigate what is the best combination of sensors and actuators for an AT. The secondary focus is to compare ATs by providing an ordinal ranking analysis on each system. While it is difficult to compare AT of one approach with another of different approach, the spirit here is to provide useful insights about what kind of systems people embrace. It is not our goal to use this analysis to critique these systems, but rather to study how such systems are used.

AT solutions can be categorized into one of the following three approaches: physical; digital; or medical. These classification is significant for practical purposes (i.e., low vision may not require a surgery to gain partial sight, but a physical or digital solution to enlarge a document or scene in front), however, the study of its interfaces cut across approaches. Nonetheless, for simplicity, various systems are labeled in alphabetical order based on these approaches. While there are commercially available systems such as UltraCane (Mândru *et al*., 2007) and Miniguide (Hill and Black, 2003), we selected the 13 systems with research articles for this comparative study using keywords such as "multimodal," "alternative perception," and "virtual environment."

While some devices have one form of sensors or actuators, others have a combination of heterogeneous sensors or actuators (i.e. multimodal). Digital approaches to design AT can be completely or partially virtual. Unlike physical devices, digital systems can receive input ranging from standard keyboard and mouse, body tracking, to interacting and manipulating other objects. Although haptic is a popular feedback mechanism, digital systems generally use a combination of haptic and audio to provide an immersive experience (Torres-Gil *et al*., 2010). Software can be used to construct an environment (based on reality or otherwise) for free exploration (or task completion) or a game to improve certain skills such as navigation and cognitive mapping (Ghali *et al*., 2012). Unlike physical and digital approaches, retinal prostheses are not substitution devices. Instead, they tackle the underlying problem directly by either stimulating the optic nerve or in the worst case, the visual cortex in the occipital lobe (i.e. cortical implant). Due to strict medical regulation, retinal prostheses that have made it to the consumer market are rare and few. As such, in addition to advances in medical technology that make the surgical techniques more efficient and effective, other researchers are making improvements to retinal prostheses itself by developing improved processing algorithms and simulations.

This paper analyzes various systems and devices using the following ten features (adapted from Plos *et al*., 2012; Dakopoulos and Bourbakis, 2010), which are much needed in most of the AT solutions, for use and further development. These features are chosen in an attempt to address the universality of an AT, while retaining good system performance and functionalities:

1. modular design: design with interchangeable parts that can meet a variety of needs;
2. functional acceptability: design that is useful, reliable, and robust;
3. accessibility: design that can extend to several groups of people, if not the general population;
4. social integration: design that is unobtrusive, but with aesthetic features and social values;
5. real time: the system operates fast enough such that information exchange with the user is useful;
6. low cost: affordable to most users;
7. friendly: low learning curve for the system; easy to use;
8. wireless: the device is connected wirelessly to a remote computer;
9. availability: the device is ready to use and further experiments can be conducted; and
10. performance: overall performance.

In order to have a more objective evaluation, five experts from New York State Commission for the Blind, Lighthouse Guild, and Vista Wearable, were consulted to provide a weighted rating for each of the ten features. Furthermore, 19 senior design capstone students, who were working on AT projects in their senior year, were also recruited to provide feature ratings as well. We might view this as a new type of critical review approach; instead of just making observations and drawing conclusions by a few authors, more people in the fields are involved in the critical review. Table I shows the normalized feature ratings by experts and students. Each feature was rated with a range 1-10 (i.e. 10 is the most important). While the expert ratings are higher than the students' in absolute values, its relative values are similar. This can be shown in values in parentheses, after quantile normalization.

Since the normalized feature ratings between experts and students are similar, the 19 students are also recruited to provide evaluation feedback by reading the papers outside the classes, and presenting/discussing them in classes, and then individually rate the systems based on the features described above. For every feature and for every system, students were asked to assign a Likert-scale value from a range 1-5, where 1 stands for "strongly disagree" and 5 stands for "strongly agree." This was done thoroughly over one semester, with two class meets every week. The students are guided by both the instructor of the course (Zhu) and some of the experts/visually impaired users; by providing comments and feedback. Papers were read by all and each paper presented by two students (one for a survey, and one for QAs). For a manageable amount of reading assignments, only 13 papers were selected, and since the class was concluded in 2014, which is why only systems before 2014 were included in the comparison. However, eight more relevant papers since conclusion of the class are briefly discussed in the three categories.

3. Physical systems

This section discusses five physical systems with haptic and auditory feedback. Table II is a summary of the systems in four aspects: sensors, actuators, processing, and functions.

Table I Normalized feature ratings by experts and students

	F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*	*F9*	*F10*
Experts	0.50 (0.35)	1.00 (0.91)	0.72 (0.54)	0.84 (0.62)	0.96 (0.77)	0.82 (0.60)	0.90 (0.70)	0.58 (0.43)	0.86 (0.67)	0.92 (0.73)
Students	0.29 (0.43)	0.49 (0.70)	0.39 (0.62)	0.37 (0.57)	0.58 (0.77)	0.37 (0.57)	0.53 (0.73)	0.19 (0.35)	0.47 (0.67)	0.81 (0.91)

Table II Physical systems

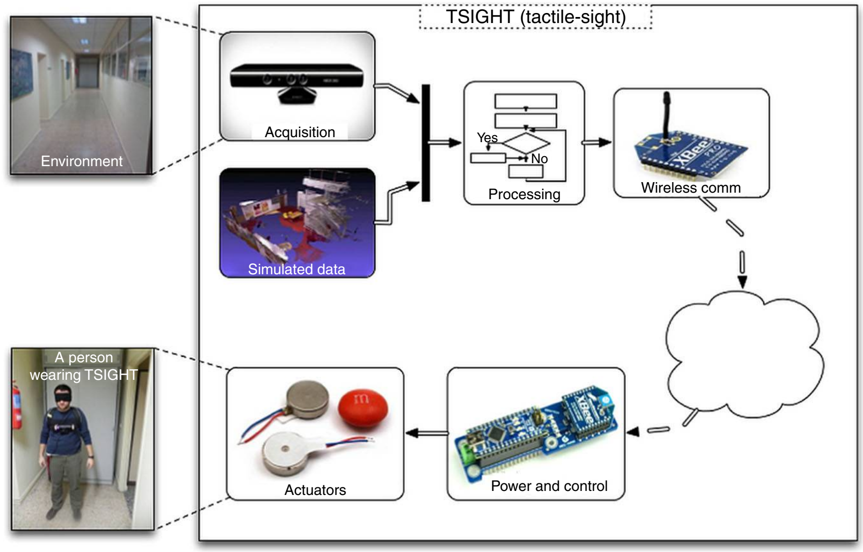
Devices	*Sensors*	*Actuators*	*Processing units*	*Functions*
TSIGHT (Cancar *et al*., 2013)	Depth	Vibrator; 72 units placed on abdominal area	Integrated circuit	Capturing distance of objects and convey it via an array of vibrators on waist
Rehab shoes and glasses (Abu-Faraj *et al*., 2012)	Ultrasonic	Vibrator/buzzer; 3 pairs on toe cap and 1 pair on glasses	Micro-controller	Detect gaps on grounds and obstacles on or above head level
SpaceSense (Yatani *et al*., 2012)	Smartphone	Vibrator; 9 units on back of phone	iPhone app	Convey step-by-step navigational directions
Talking Point 3 (Yang *et al*., 2011)	Smartphone	Text-to-speech	Determine location data to include and provide interactive control	Automatically provide nearby point of interests (POIs) with the ability to query other POIs
ConWiz (Polacek *et al*., 2012)	Smartphone	Text-to-speech	WOz prototype	Synthesizing voice commands for navigation

There are more recent systems in this category such as smartphone-based indoor localization system (Basso *et al*., 2015), electronic travel support system using 3D glasses (Mattoccia and Macri, 2015), and vibrotactile feedback on body and foot (Meier *et al*., 2015).

TSIGHT (System A)

The TSIGHT (Cancar *et al*., 2013) receives processed data (i.e. distance) from a Microsoft Kinect that is attached to the torso with a belt. Through an XBee wireless device, commands are sent to an I^2C control module that receives activation intensity for each actuator, and resends the information to the corresponding actuators (Figure 1). In their experiment, 72 (6 × 12) vibrotactile actuators are housed in a waistband placed on the abdominal area of the user for navigation.

Figure 1 General architecture of TSIGHT



Source: Cancar *et al*. (2013)

Even though TSIGHT is not very modular, it is performing in real time. Our experts believe that both systems can convey information to users in at least 30 Hz.

Rehab shoes and spectacles (System B)

Abu-Faraj *et al*. (2012) designed a prototype of a rehabilitative pair of shoes and eyeglasses for the blind. The device is used to detect obstacles at ground and head levels, including stair ascent and descent. They first retrofitted each shoe with three pairs of ultrasonic transducers placed on the toe cap with spacing in between to allow maximum coverage. Vibrating motors are attached inside the shoes underneath the corresponding locations of the transducers. Second, they retrofitted a pair of ultrasonic transducers above the bridge of a pair of Ray-Ban eyeglasses, and a buzzer is mounted at one of the temples (Plate 1). This design is useful in navigation, especially in situations like detecting potholes and knowing when you're at the top (or bottom) of stairs. It, however, has a low social integration value.

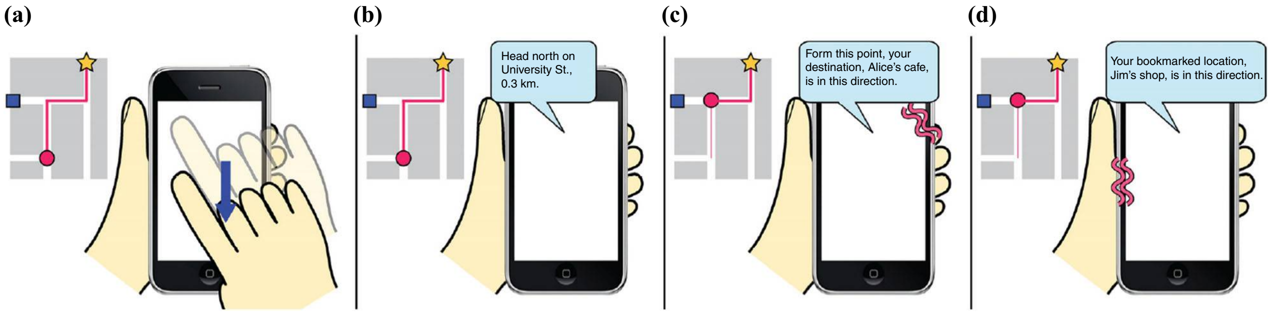
SpaceSense (System C)

SpaceSense is an iPhone map application with a custom spatial tactile feedback system (Yatani *et al*., 2012). The feedback system consists of nine (3×3) vibrating motors on the backside of the mobile device. The application allows the user to select places of interest and retrieves up to 20 nearby locations of the user. The user can use flick gestures to navigate the list (Figure 2). When a location is selected, the application reads out the name and address of the place, and

Plate 1 Rehabilitative shoes and spectacles

Source: Abu-Faraj *et al*. (2012)

Figure 2 Interactions supported by SpaceSense



Source: Yatani *et al*. (2012)

use spatial tactile feedback to indicate the location's distance and cardinal direction from user. The distance is conveyed by vibration intensity. If the location is to the left of the user (cardinal direction), the left side of the device vibrates. With a selected location as a destination, the application presents each step of the walking instructions in the cardinal directions along with its walking distance, one at a time. The system is useful in navigation for the visually impaired as well as sighted users when navigating in an urban area. Mobile smartphones are ubiquitous nowadays, as is navigating around a city with a smartphone; therefore, the social integration value for this system is high.

Talking Point 3 (System D)

Yang *et al*. (2011) designed a location- and orientation-aware smartphone-based system called Talking Point 3 (TP3). It allows the user to interactively access local environment information and convey the information through text-to-speech. The information is stored in a central database that can be accessed freely and updated by anyone. This information can include location, name, hours of operation, detailed descriptions, and user comments. The system pushes selected information about the immediate surroundings to the user, while allowing the user to pull more available detailed information about the immediate and distant surroundings (Figure 3). User interaction is achieved using touch screen gestures, a single button press, and a shake gesture. TP3 is useful for the visually impaired as well as sighted users. It is also accessible and has high social integration value.

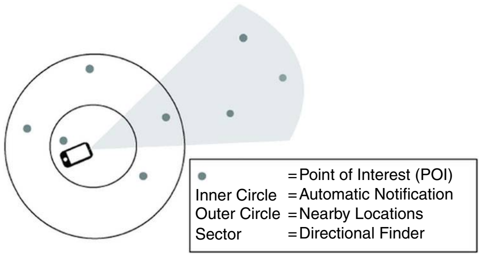
ConWiz (System E)

Similar to TP3, Polacek *et al*. (2012) uses a Wizard of Oz (WOz) testing approach to develop their WOz prototype, which is part of the Contextual Wizard (ConWiz) system. WOz experiment is an experiment in which subjects interact with a system that subjects believe to be autonomous, but which is actually being controlled or partially controlled by a person. The WOz prototype is an Android smartphone application that uses synthesized voice commands for navigation and controlling a vibro-wristband for tactile commands (Figure 4). The human wizard is controlling a separate device, Mobile Wizard, that can send navigational and vibration commands to the WOz prototype. Both the Mobile Wizard and WOz prototype are connected to a server that has data and an evaluator can monitor the experiment progress. The vibration is used to alert the user of any dangerous situation. The choice of employing WOz testing approach is to simplify the design and focus on investigation of navigational commands and usability issues. Although all components are *wir*eless, the system needs a human wizard to give navigational commands, it is therefore, currently not available.

4. Digital approaches

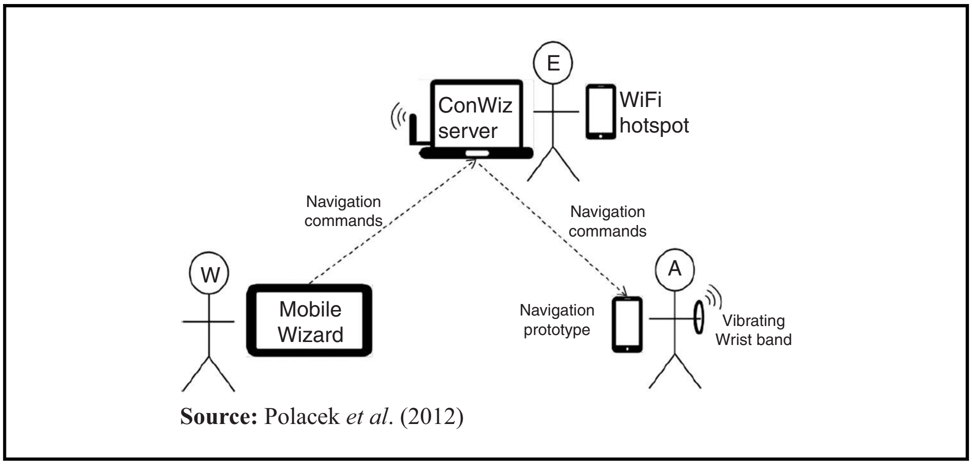
Unlike physical systems, digital approaches, mostly with virtual environments (VEs), allow for quick prototyping and testing without dealing with the physical limitations of devices. VEs are also

Figure 3 Talking Point 3 mechanisms



Source: Yang *et al*. (2011)

Figure 4 ConWiz setup



Source: Polacek *et al.* (2012)

less restrictive, as the experimenter can configure various virtual sensors with real actuators to investigate different setup. Table III lists VEs for navigational tasks. More recent VEs that complement this section are VEs using soundscape (Chandrasekera *et al.*, 2015) and EyeCane (Chebat *et al.*, 2015).

BlindAid (System F)

Lahav *et al.* (2008, 2012) developed the BlindAid system (Plate 2) to study the exploration process of an unknown space in a VE, the cognitive processing of it, and applying the knowledge in real space. It uses a Microsoft SideWinder Force Feedback Joystick as input and haptic feedback. The VE is based on a real space with randomly placed obstacles. Subjects were asked to freely explore the VEs without time limits, and then use building block components to map the layout. Finally, with the mental map, subjects were asked to perform orientation tasks (e.g. reach and identify an object) in the real environments. This system is quite useful to the visually impaired in that it helps them cognitively map an environment before visiting the actual space.

Hara et al. Project (System G)

Hara *et al.* (2010) designed an augmented reality to evaluate multimodal feedback strategies. The authors used a Wiimote[6] to provide audio and vibrotactile feedback to the subject. The Wiimote is attached to an aluminum stick (like a long cane) as a handle. On the other end of

Table III Digital approaches

Devices	*Sensors*	*Actuators*	*Processing units*	*Functions*
Lahav *et al.* (2008, 2012)	Force feedback joystick	Force feedback joystick	VE based on real space with randomly placed obstacles	Cognitive mapping and preplanning orientation aid
Hara *et al.* (2010)	Active marker/ tracking system	Wiimote for audio and haptic	Track user and map onto VE, and alert user if user gets close to a virtual wall	To evaluate multimodal feedback strategies
MOVA3D (Sanchez *et al.*, 2010)	Digital Clock Carpet	Spatial sounds	Microsoft XNA	A game to evaluate Digital Clock Carpet
Audio-based Environment Simulator (Connors *et al.*, 2012)	Keyboard	Audio	Render environment from building floor plan or virtual rendering of modern building and gamified it	Simulated navigation and exploration of existing physical building

Plate 2 The BlindAid system

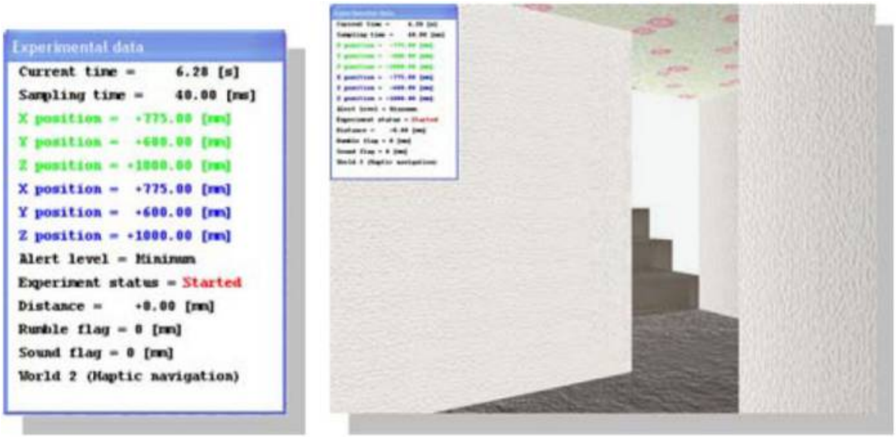
Source: Lahav *et al*. (2012)

the stick is an active marker. There is an active marker on a cap that the subject wears, which is tracked by an optical tracking system and mapped onto the VE (Figure 5). As the subject gets close to a virtual wall, the Wiimote will alert the user via audio or haptic feedback. Even though the entire system can be expensive (mainly the optical tracking system), each of the components are modular and wireless.

MOVA3D (System H)

As an application of the Digital Clock Carpet (DCC), MOVA3D is a game to help improve navigation skills for visually impaired children (Sanchez *et al*., 2010). The game receives inputs from DCC or keyboard. With keyboard, users navigate using the arrow keys. With DCC, users use the clock system to step in the direction they want to take, with 12 o'clock always in front of you. The spatial sound is relative to the user's orientation. The results of their experiment show that the system is useful and easy to use. Visually impaired children, furthermore, can play the game along with their sighted friends since MOVA3D GUI also renders normal textures for walls and floors (Figure 6). The system is relatively low cost since the customized haptic device is built

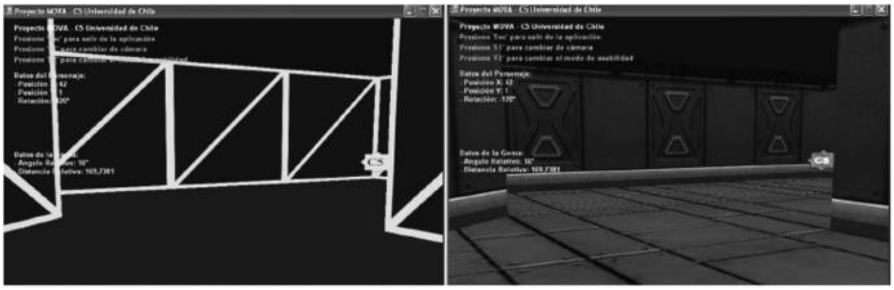
Figure 5 A status window and a visualized virtual world



```
Current time =      6.28 [s]
Sampling time =     40.00 [ms]
X position =   +775.00 [mm]
Y position =   +600.00 [mm]
Z position = +1000.00 [mm]
X position =   +775.00 [mm]
Y position =   +600.00 [mm]
Z position = +1000.00 [mm]
Alert level = Minimum
Experiment status = Started
Distance =      +0.00 [mm]
Rumble flag = 0 [mm]
Sound flag = 0 [mm]
World 2 (Haptic navigation)
```

Source: Hara *et al*. (2010)

Figure 6 MOVA3D



Source: Sanchez *et al*. (2010)

from commercial-off-the-shelf parts and the game is developed in-house. Overall, the system has good performance and it had captured subjects' interest from the beginning. An extension of this project is MovaWii (Sanchez *et al*., 2014).

Connors *et al*. (2012) developed the Audio-based Environment Simulator (AbES) that allows simulated navigation and exploration of existing physical building. The AbES can render an environment from a building floor plan or virtual rendering of a modern building. Users interact with AbES by using keyboard as input and spatialized sound cues are generated via two speakers placed left and right of the users. The user has to collect as many jewels as possible hidden in various rooms while avoiding roving monsters that can steal the jewels from them. This can be beneficial for visually impaired people in improving their navigation skills and building a cognitive map of a particular place. This system is accessible to people who may be interested in exploring a new place, either for fun or in anticipation of a visit.

5. Medical solutions

In contrast with physical systems, retinal prostheses are very expensive and involved invasive surgery. In contrast with digital, prostheses simulation and algorithms seek to improve the functionalities of an implant while satisfying very restrictive constraints (e.g. low resolution, power usage, etc.) and safety requirements. Table IV lists various prosthesis devices, and simulation and processing algorithms to improve retinal prosthesis. Some recent retinal implant or related simulation algorithms are simulated prosthetic vision using photorealistic VEs (Zapf *et al*., 2014), peripheral visual prosthesis (Zapf *et al*., 2015), and bi-modal visual representation (Feng *et al*., 2014).

Table IV Medical solutions

Devices	*Sensors*	*Actuators*	*Processing units*	*Functions*
Argus II	Camera	60-electrode array; epiretinal	Convert img to electrical signals	Partial vision restoration
Brainport	Camera	400-electrode array; on tongue	Convert img to electrical signals	Alternative perception using tongue
Weiland *et al*. (2012)	Stereo camera	Vibrotactile	Simultaneous localization and mapping algorithm	Using computer vision algorithm to develop smart image processing system for retinal prosthesis
Time-to-contact map (McCarthy and Barnes, 2012)	Depth	Monitor display (testing)	Compute ratio of surface distance and object's relative velocity toward observer	Faster way of computing "depth" map than phosphene modeling and detect fast approaching object or person

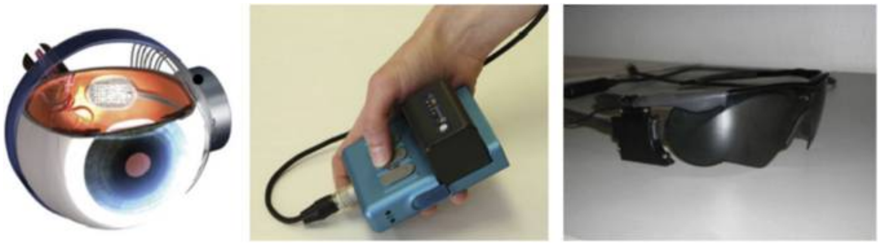
Argus II (System J)

A notable retinal prosthesis that is currently in clinical trials in the USA is the Argus II system (Plate 3) by Second Sight[7]. It is the world's first FDA approved retinal prosthesis that attempts to restore partial vision for VIP (Weiland and Humayun, 2014; Weiland *et al.*, 2011; Ahuja and Behrend, 2013). The system consists of an implantable device surgically implanted on the eye via an epiretinal approach, and can communicate with an external unit worn by the user. The system also consists of a 60-electrode array. The external unit consists of glasses, a video processing unit, and a battery. A small camera (510 × 492 resolution) is mounted on the glasses, which captures video data, sends it a processor, and then converts the image to electronic signals that are sent to the implant. Lastly, the implant wirelessly receives signals and sends the appropriate electrical stimulus pulses to the electrode array (Weiland and Humayun, 2014; Ahuja and Behrend, 2013). The system, while is useful, it has a low social integration value because users have to wear the external unit and a cable that connects to the glasses. The system is also expensive.

Brainport (System K)

Unlike invasive retina implants, Brainport (Plate 4) from Wicab[8] is a tongue-based electrical stimulation device that conveys brightness contrast of a scene in front of the user through a

Plate 3 Argus II



Source: Ahuja and Behrend (2013)

Plate 4 Brainport

Source: Khoo *et al*. (2012)

20 × 20 electrode array pressed against the tongue. A camera is mounted on a pair of glasses that captures videos, which is then sent to a base unit that processes the data. This involves converting it to a grayscale image and then downsamples it to 20 × 20 pixels. The electrotactile stimulus data is then transmitted to the tongue display unit (Bach-y Rita *et al*., 2003; Danilov and Tyler, 2005; Khoo *et al*., 2013). Users have to interpret information via the tongue and thus impedes their ability to speak. It, therefore, has low social integration value.

Weiland et al. Project (System L)

Weiland *et al*. (2012) proposed a smart image processing system using computer vision algorithms for retinal prosthesis, in two parts (Figure 7). First, the navigation component uses a stereo camera[9] to generate a dense 3D map, which is used to localize the user on a global frame of reference and continuously maps new obstacles detected. Navigational directions are computed based on the map and feedback given via vibrotactile signals (vibration motors on left/right shoulders). Second, the object detection component has a webcam mounted on a head-mounted display, where the video stream is modified to simulate prosthetic vision and feed into the display. The simulation of prosthetic vision can be controlled by specifying the number of active pixels, pixel location, and pixel size. In addition to these pixels, there are also eight cuing pixels that indicate the direction of the most important object. This system has a low social value because special hardware has to be outfitted onto the person and the components are connected by wires.

Time-to-contact map (System M)

McCarthy and Barnes (2012) presented an alternative visual representation for visual prosthesis that encode the distance to a surface with respect to their time-to-contact. Time-to-contact is defined as the ratio of the surface distance and the object's relative velocity toward the observer. Such encoding not only captures spatial information, but temporal information as well. This ratio is computed from the temporal changes of depth. This representation provides a faster way of computing the "depth" map and clearest visualization than phosphene modeling (Figure 8). It is intuitive and easy to understand that a brighter region in the image means a fast approaching object or person.

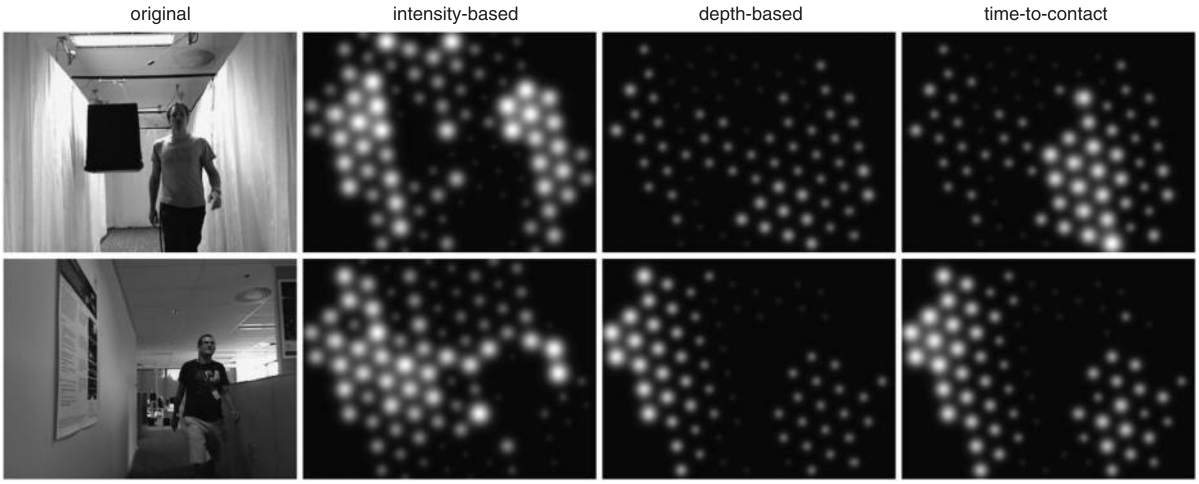
6. Comparative study

This section attempt to quantitatively compare the 13 aforementioned assistive technologies with respect to the features described in Section 2. This study is not designed as an experiment, but

Figure 7 Prototype of the smart image processing system

Source: Weiland *et al*. (2012)

Figure 8 Visual comparisons of scenes using simulated prosthetic vision



Source: McCarthy and Barnes (2012)

as a new kind of critical review for providing researchers more useful information than just the discussions we provided in Sections 3-5.

Formulation

In Section 2, five experts were asked to provide a rating from a scale of 1-10 for each feature, where 1 stands for "not very important" and 10 stands for "very important." The feature weights, then, are computed as follows:

$$w_i = \frac{\sum_{j=1}^{5} ER_{ij}}{5N} \tag{1}$$

where N is the number of features (10, in our case) and ER_{ij} is the jth expert's rating for $feature_i$. The average feature weights being used for this study are shown in the first row of Table V. Finally, the ranking can be computed using the following formula:

$$S_k = \sum_{i=1}^{N} \frac{w_i x_i}{N} + b \tag{2}$$

$$b = \sum_{k=1}^{M} \frac{\left| \sum_{i=1}^{N} \frac{w_i x_{k,i}}{N} - \sum_{i=1}^{N} \frac{x_{k,i}}{N} \right|}{M} \tag{3}$$

where M is the number of systems (13, in our case); i and j refer to specific features; k refers to a specific system; $x_{k,i}$ is the rating of system k and $feature_i$; S_k is the score for system k; and b is bias (0.62, in our case). Since each feature is weighted, the ranking for each system is biased. To correct this, the bias is computed by calculating the difference between the average feature values weighted and un-weighted.

For every feature F_i and for every system (A-M), a value (x_i) is assigned using the Likert-scale from a range 1-5, where 1 stands for "strongly disagree" and 5 stands for "strongly agree." When there is not enough information for a feature of a system, a value of 3, which stands for "neutral," is chosen.

Results

Table V shows the normalized average feature weights by experts, the average rating results by students, and the overall system score. The average standard deviation for all ratings and

Table V Rating for all systems

	F1 *0.50*	*F2* *1.00*	*F3* *0.72*	*F4* *0.84*	*F5* *0.96*	*F6* *0.82*	*F7* *0.9*	*F8* *0.58*	*F9* *0.86*	*F10* *0.92*	*System Score*
A	2.63	3.84	3.32	2.32	4.37	3.53	3.42	3.47	3.53	4.05	3.46
B	3.05	3.63	3.11	3.58	4.16	3.95	3.68	2.84	2.74	3.68	3.45
C	2.89	3.68	3.68	4.21	4.37	3.68	4.05	4.26	3.32	3.95	3.73
D	3.26	3.63	3.74	4.32	4.11	3.74	3.63	4.47	3.32	3.63	3.68
E	3.11	3.32	3.37	4	3.37	3.74	3	3.16	2.26	3.11	3.25
F	2.68	3.63	3.42	3.16	3.84	2.32	3.74	2.74	2.84	3.89	3.28
G	3.58	4.16	3.37	3.32	4.21	3.21	3.89	4.42	3.05	4.26	3.66
H	3.16	3.84	3.53	3.42	4.11	4.05	4.11	2.16	2.95	4.42	3.58
I	3	3.95	3.68	3.58	4.05	3.95	4.16	3.16	3.37	4.37	3.69
J	2.68	3.68	2.74	3.68	4.21	2.11	2.74	2.95	2.68	3.53	3.18
K	2.68	3.32	3	1.84	3.63	2.37	2.37	1.95	1.84	2.95	2.75
L	2.95	3.74	2.89	2.58	4	2.84	3.63	3.21	2.58	3.58	3.24
M	3.37	3.68	3.05	4.26	3.89	3.05	3.95	4.47	2.37	3	3.45

Notes: A-E: physical; F-I: digital; J-M: medical

features are 0.87 and 0.38, respectively. The feature that was most valued by our experts is F2, functional acceptability. The least valued is F1, modular design. Ranked No. 1 for physical systems is SpaceSense (System C). Ranked No. 1 for digital approaches is AbES (System I). Lastly, ranked No. 1 for medical solutions is time-to-contact map (System M). Overall, SpaceSense and AbES are tied for No. 1, while Brainport (System K) is overall last.

7. Discussion

This paper surveyed the recent development of ATs for visually impaired persons with three types of solutions: physical, digital, and medical, and analyzed 13 multimodal assistive technologies. For each system, its interfaces have been described and its processing unit explained. The most popular sensors are optical, infrared, and ultrasonic. Similarly, the most popular actuators are audio and haptic. It is interesting to note, however, that all systems scored around three. This tells us that there does not exist a system that solves all and is very popular, instead ATs become niche products solving very specific needs. Even then, each system by itself is not as strongly desirable as one hopes. Note that this is a limited comparative study based on what is reported in the literature and evaluated by developers, instead of actually trying out the systems. However, experts were consulted on what features are important, some of whom are visually impaired, and systems were compared accordingly. Furthermore, this paper provide a new kind of approach to evaluate and compare systems before they come to the market, which could be cost-effective and valuable for both researchers and developers.

Surprisingly, our experts did not think modular design is a very important feature. “I don’t care about modular design, as long as it works,” said one expert who is visually impaired (Campbell). Overall, they believe that Hara *et al*. (System G) is the most modular of them all. The main purpose of that system is to evaluate multimodal feedback, so once the tracking system is setup, the experimenters can interchangeably use audio or haptic feedback in whatever augmented environment they want. Our collective experts also believe that System G is more functionally acceptable than others. The least functional acceptable systems are ConWiz and Brainport. ConWiz is still a conceptual prototype, therefore, usefulness is difficult to judge. Brainport, however, requires the users to interpret information via the tongue and it impedes their ability to speak. One expert who used Brainport said “it’s very awkward and you can’t talk while using it.”

While both MOVA3D and AbES are virtual based systems, MOVA3D is rated the most affordable and AbES the easiest to use. MOVA3D’s customized haptic device is built from commercial-off-the-shelf parts and the environment is developed in-house. AbES relied upon our ability to spatially localized objects. Furthermore, the finding jewel game is fun and intuitive to play. Both TP3 and time-to-contact map rated the best for wireless feature. Understandably, TP3 is a smartphone app, and time-to-contact map is a processing algorithm that can be incorporated into any retinal implant processing unit. However, since wireless is one of the least important features, they are weighted less among the ranking.

Note that all of the systems are still in research phase or clinical trial. Our experts, however, believe that TSIGHT shows the most promise of maturity and availability. Lastly, our experts believe that System G is the most functionally acceptable, which is the most important feature.

Notes

1. World health organization, "Visual impairment and blindness," July 1, 2015, www.who.int/mediacentre/factsheets/fs282/en
2. World health organization, "Magnitude of blindness and visual impairment," July 1, 2015, www.who.int/blindness/causes/magnitude/en/
3. National Federation of the Blind, "Blindness and Low Vision: Fact sheet," July 1, 2015, https://nfb.org/fact-sheet-blindness-and-low-vision
4. Sound Foresight Technology Ltd, "Ultracane," July 29, 2015, www.ultracane.com/
5. GDP research, "The Miniguide mobility aid," July 29, 2015, www.gdp-research.com.au/minig_1.htm
6. Nintendo, "Wii U," July 1, 2015, www.nintendo.com/wii/
7. Second sight, "Argus II Retinal Prosthesis System," July 1, 2015, www.secondsight.com/g-the-argus-ii-prosthesis-system-pf-en.html
8. Wicab Inc., "Brainport V100," July 1, 2015, www.wicab.com/en_us
9. TYZX, "Sight in an Unstructured World," July 1, 2015, www.tyzx.com/

References

Abu-Faraj, Z.O., Jabbour, E., Ibrahim, P. and Ghaoui, A. (2012), "Design and development of a prototype rehabilitative shoes and spectacles for the blind", *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference, IEEE*, pp. 795-9.

Ahuja, A.K. and Behrend, M.R. (2013), "The Argus II retinal prosthesis: factors affecting patient selection for implantation", *Progress in Retinal and Eye Research*, Vol. 36 No. 5, pp. 1-23.

Bach-y Rita, P., Tyler, M.E. and Kaczmarek, K.A. (2003), "Seeing with the brain", *International Journal of Human-Computer Interaction*, Vol. 15 No. 2, pp. 285-95.

Basso, S., Frigo, G. and Giorgi, G. (2015), "A smartphone-based indoor localization system for visually impaired people", *Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium*, pp. 543-8.

Cancar, L., Diaz, A., Barrientos, A., Travieso, D. and Jacobs, D. (2013), "Tactile-sight: a sensory substitution device based on distance-related vibrotactile flow", *International Journal of Advanced Robotic System*, Vol. 10 No. 272.

Chandrasekera, T., Yoon, S.Y. and D'Souza, N. (2015), "Virtual environments with soundscapes: a study on immersion and effects of spatial abilities", *Environment and Planning B: Planning and Design*, Vol. 42 No. 6, pp. 1003-19.

Chebat, D.R., Maidenbaum, S. and Amedi, A. (2015), "Navigation using sensory substitution in real and virtual mazes", *PLoS ONE*, Vol. 10 No. 6, p. e0126307. doi: 10.1371/journal.pone.0126307.

Connors, E., Yazzolino, L., Sanchez, J. and Merabet, L. (2012), "Development of an audio-based virtual gaming environment to assist with navigation skills in the blind", *Journal of Visualized Experiments: JoVE*, No. 73, p. e50272. doi: 10.3791/50272.

Dakopoulos, D. and Bourbakis, N. (2010), "Wearable obstacle avoidance electronic travel aids for blind: a survey", *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions*, Vol. 40 No. 1, pp. 25-35.

Danilov, Y. and Tyler, M. (2005), “Brainport: an alternative input to the brain”, *Journal of Integrative Neuroscience*, Vol. 4 No. 4, pp. 537-50.

Feng, D., Walker, J., Barnes, N. and McCarthy, C. (2014), “A bi-modal visual representation can enhance orientation and mobility performance with less than 20 phosphenes”, *Investigative Ophthalmology & Visual Science*, Vol. 55 No. 13, p. 1799.

Ghali, N.I., Soluiman, O., El-Bendary, N., Nassef, T.M., Ahmed, S.A., Elbarawy, Y.M. and Hassanien, A.E. (2012), “Virtual reality technology for blind and visual impaired people: reviews and recent advances”, in Gulrez, T. and Hassanien, A.E. (Eds), *Advances in Robotics and Virtual Reality*, Springer, Berlin, Heidelberg, pp. 363-85.

Hara, M., Shokur, S., Yamamoto, A., Higuchi, T., Gassert, R. and Bleuler, H. (2010), “Virtual environment to evaluate multimodal feedback strategies for augmented navigation of the visually impaired”, *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 975-78.

Hill, J. and Black, J. (2003), “The miniguide: a new electronic travel device”, *Journal of Visual Impairment & Blindness*, Vol. 97 No. 10, pp. 1-6.

Khoo, W.L., Seidel, E.L. and Zhu, Z. (2012), “Designing a virtual environment to evaluate multimodal sensors for assisting the visually impaired”, in Miesenberger, K., Karshmer, A., Penaz, P. and Zagler, W. (Eds), *Proceedings of the 13th International Conference on Computers Helping People with Special Needs – Volume Part II, ser. ICCHP’12*, Springer-Verlag, Berlin and Heidelberg, pp. 573-80.

Khoo, W.L., Knapp, J., Palmer, F., Ro, T. and Zhu, Z. (2013), “Designing and testing wearable range-vibrotactile devices”, *Journal of Assistive Technologies*, Vol. 7 No. 2, pp. 102-17.

Lahav, O. and Mioduser, D. (2008), “Haptic-feedback support for cognitive mapping of unknown spaces by people who are blind”, *International Journal of Human-Computer Studies*, Vol. 66 No. 1, pp. 23-35.

Lahav, O., Schloerb, D., Kumar, S. and Srinivasan, M. (2012), “A virtual environment for people who are blind – a usability study”, *Journal of Assistive Technologies*, Vol. 6 No. 1, pp. 38-52.

McCarthy, C. and Barnes, N. (2012), “Time-to-contact maps for navigation with a low resolution visual prosthesis,”, *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 2780-83.

Mândru, D., Lungu, I., Mociran, A. and Tatar, O. (2007), “Development of a mechatronic blind stick”, *Annals of the University of Oradea, Fascicle of Management and Technological Engineering*, Vol. 6 No. 26, pp. 797-802.

Mattoccia, S. and Macri, P. (2015), “3D glasses as mobility aid for visually impaired people”, in Agapito, L., Bronstein, M. and Rother, C. (Ed.), *Computer Vision -– ECCV 2014 Workshops Lecture Notes in Computer Science*, Springer International Publishing, Zurich, pp. 539-54.

Meier, A., Matthies, D.J.C., Urban, B. and Wettach, R. (2015), “Exploring vibrotactile feedback on the body and foot for the purpose of pedestrian navigation”, *Proceedings of the 2nd International Workshop on Sensor-based Activity Recognition and Interaction, ACM, Rostock*, pp. 1-11.

Plos, O., Buisine, S., Aoussat, A., Mantelet, F. and Dumas, C. (2012), “A universalist strategy for the design of assistive technology”, *International Journal of Industrial Ergonomics*, Vol. 42 No. 6, pp. 533-541.

Polacek, O., Grill, T. and Tscheligi, M. (2012), “Towards a navigation system for blind people: a wizard of oz study”, *SIGACCESS Accessible Computing*, No. 104, pp. 12-29, available at: http://dl.acm.org/citation.cfm?id=2388820

Sanchez, J., de Borba Campos, M. and Espinoza, M. (2014), “Multimodal gaming for navigation skills in players who are blind”, *Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems, Sociedade Brasileira de Computação, Porto Alegre*, pp. 169-79.

Sanchez, J., Saenz, M. and Garrido, J.M. (2010), “Usability of a multimodal video game to improve navigation skills for blind children”, *ACM Transactions on Accessible Computing (TACCESS)*, Vol. 3 No. 2, pp. 7:1-7:29.

Torres-Gil, M., Casanova-Gonzalez, O. and Gonzalez-Mora, J. (2010), “Applications of virtual reality for visually impaired people”, *WSEAS Transactions on Computers*, Vol. 9 No. 2, pp. 184-93.

Weiland, J.D. and Humayun, M.S. (2014), “Retinal prosthesis”, *Biomedical Engineering, IEEE Transactions*, Vol. 61 No. 5, pp. 1412-24.

Weiland, J., Cho, A.K. and Humayun, M. (2011), “Retinal prostheses: current clinical results and future needs”, *Ophthalmology*, Vol. 118 No. 11, pp. 2227-37.

Weiland, J.D., Parikh, N., Pradeep, V. and Medioni, G. (2012), “Smart image processing system for retinal prosthesis”, *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 300-3.

Yang, R., Park, S., Mishra, S.R., Hong, Z., Newsom, C., Joo, H., Hofer, E. and Newman, M.W. (2011), “Supporting spatial awareness and independent wayfinding for pedestrians with visual impairments”, *The Proceedings of the 13th international ACM SIGACCESS Conference on Computers and accessibility, ACM*, pp. 27-34.

Yatani, K., Banovic, N. and Truong, K. (2012), “Spacesense: representing geographical information to visually impaired people using spatial tactile feedback”, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM*, pp. 415-24.

Zapf, M.P., Boon, M.Y., Matteucci, P.B., Lovell, N.H. and Suaning, G.J. (2015), “Towards an assistive peripheral visual prosthesis for long-term treatment of retinitis pigmentosa: evaluating mobility performance in immersive simulations”, *Journal of Neural Engineering*, Vol. 12 No. 3, p. 036001, available at: http://iopscience.iop.org/article/10.1088/1741-2560/12/3/036001

Zapf, M.P., Matteucci, P.B., Lovell, N.H., Zheng, S. and Suaning, G.J. (2014), “Towards photorealistic and immersive virtual-reality environments for simulated prosthetic vision: integrating recent breakthroughs in consumer hardware and software”, *Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE*, pp. 2597-600.

About the authors

Wai Lun Khoo received his BS and MS in Computer Science at the City College of New York – CUNY in 2009 and 2012, respectively. He is currently a Doctoral Student at the Graduate Center – CUNY. His research interests include computer vision, multimodal sensing, and human-computer interaction. He is also interested in the application of computer vision to assistive technology for the blind and visually impaired people. Wai Lun Khoo is the corresponding author and can be contacted at: khoo@cs.ccny.cuny.edu

Zhigang Zhu is the Herbert G. Kayser Chair Professor of Computer Science, at the City College of New York and The Graduate Center – CUNY, and directs the City College Visual Computing Laboratory (CCVCL). His research interests include 3D computer vision, multimodal sensing, virtual/augmented reality, video representation, and various applications in assistive technology, environment, robotics, surveillance, and transportation. He has published over 140 technical papers in the related fields. He is an Associate Editor of the *Machine Vision Applications Journal*, and has been a Technical Editor of the *ASME/IEEE Transactions on Mechatronics*.