Sensitive interfaces for blind people in virtual visits inside unknown spaces

Nancy E. Guerrón , Antonio Cobo , José J. Serrano Olmedo ,
Carlos Martín

PII: S1071-5819(19)30101-6
DOI: https://doi.org/10.1016/j.ijhcs.2019.08.004
Reference: YIJHC 2356

To appear in: *International Journal of Human-Computer Studies*

Received date: 24 November 2017
Revised date: 1 August 2019
Accepted date: 14 August 2019

Please cite this article as: Nancy E. Guerrón , Antonio Cobo , José J. Serrano Olmedo , Carlos Martín , Sensitive interfaces for blind people in virtual visits inside unknown spaces, *International Journal of Human-Computer Studies* (2019), doi: https://doi.org/10.1016/j.ijhcs.2019.08.004

HIGHLIGHTS:

- The virtual visits in advance, aid to blind people in the tasks of recognition of objects and structures on unknown indoor spaces.
- Participants evaluated three sensitive interfaces: voice, beeps and gestures. The effectiveness, efficiency and usefulness of each one was evaluated using up to thirty virtual reality applications.
- The combination of three sensitive interfaces and a cognitive interface in virtual reality applications facilitated the construction of cognitive maps in who are blind.

Sensitive interfaces for blind people in virtual visits inside unknown spaces

Nancy E. Guerrón [a, b], Antonio Cobo[a, c], José J. Serrano Olmedo[a, c], Carlos Martín[a]

[a] Centre for Biomedical Technology. Universidad Politécnica de Madrid, 28223, Madrid-España
[b] Universidad de las Fuerzas Armadas ESPE. 170501, Sangolquí-Ecuador
[c]Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, 28029, Madrid- España

nancy.guerron@ctb.upm.es, antonio.cobo@ctb.upm.es, josejavier.serrano@ctb.upm.es, carlos.martin.lucendo@gmail.com

ABSTRACT: Virtual reality applications with sensitive interfaces for blind users were developed and tested on smart mobiles running Android as operating system with the participation of twenty blind adults; each application was built with a cognitive interface and a combination of three sensitive interfaces (voice, beeps and gestures) to allow the user to explore the inside of a virtual scenario beforehand. The vibration was an extra, fixed sensitive interface that informed the user that the avatar was walking through the virtual world. The combination of different interfaces within the applications allowed to (1) provide multisensory information to the user while exploring an unknown virtual scenario, (2) build a model of the space in which the user is located and (3) successfully perform tasks in the real world. This research compared different patterns for each interface: 3 for voice, 3 for the beep and 4 for gestures, during the exploration of an unknown space, with the goal of determining which ones were the most efficient and effective at detecting, identifying and locating structures and objects in an unfamiliar environment, aiding blind users to develop a cognitive map of the explored place and successfully perform tasks within the real scenario. Evidence shows that the voice interface provides a statistically significant difference on effectiveness, efficiency, and cognitive mapping compared to other interfaces. Most decisively, pairwise comparisons have shown that, when compared to a normal speech rate, a speech rate 2 times faster manifested an improvement of 5.9% on object and structures detection, a reduction on exploration time of 39.72%, an improvement of 6% on quality of cognitive map development and greater success in carrying out the absolute tasks. The beep interface with 500Hz frequency and 100 milliseconds spacing provided a statistically significant difference for cognitive mapping, showing an improvement of 8.9%. Finally, in the context of gesture interfaces, a single touch plus vertical swiping, produced statistically significant differences in the reduction of discovery time by 34.28%.

Keywords: Human-computer interface; sensitive interfaces for blind people; virtual reality applications; indoor navigation

1. INTRODUCTION

1.1 Background

Visiting an unfamiliar place has a strong cognitive load on a blind person, causing him stress and discomfort (Martinez et al., 2014). When a visually impaired person walks the interior of a place, high, medium or low objects and protruding structures or concavities of the environment, could represent great obstacles along his route. To move more freely and feel more secure, he needs to obtain information from the environment using his other sensory modalities (Proulx et al., 2014). The blind person can also count on the support of people or rehabilitation technicians, who will verbally describe the place and distances between objects (Spinosa et al., 1998); or he could explore the place on his own using a white cane when there is little activity on the site (Picinali et al., 2014) and thus form a mental representation of the place for subsequent visits.

Before the journey, the person needs to get prepared for his path's difficulties, find alternative routes for inaccessible sections and trust in his own abilities (Koutny and Miesenberger, 2016). A virtual tour on a smart device can help sighted people recognize a place they do not know; however, blind people need non-visual information to learn what the place is like, where the objects are, and other relevant details to be more autonomous. e-Glance project uses virtual reality applications to help blind people to explore an unknown place through virtual visits in advance (Cobo et al., 2017).

Virtual spaces consist of a set of software tools that allow developing knowledge without people needing to move to acquire information (Malecki, 2017), these spaces can be constructed on smart electronic devices through virtual reality applications (VRA) that allow people with blindness to explore an unknown space in virtual form. The training of blind people using virtual environments has been carried out successfully, so that users acquire skills for orientation and mobility; also some studies showed that users created robust and comprehensive cognitive maps (Lahav et al., 2012; O. Lahav, Schloerb, & Srinivasan, 2015; Espinoza et al., 2014; Sánchez, 2010). Furthermore, VRA are used as support tools for education (Nincarean et al., 2013; Sánchez et al., 2014), work (Akhavian and Behzadan, 2016; (Kwon and Lee, 2016), health (O. Lahav et al., 2015), mobility support and accessible tourism (Miesenberger et al., 2014), trade and leisure (Fernandes et al., 2012). VRA use a mix of interfaces to facilitate the learning process; these interfaces can be sensitive and cognitive. Sensitive interfaces such as beep, voice or vibration make usage of the senses whereas cognitive interfaces refer to the way a person learns.

A research on how audio-tactile maps allow blind people to form mental representations of space (Gomez et al., 2012; Picinali et al., 2014) presented scenarios where blind people developed spatial knowledge by listening for audio events and interacting with these events within a virtual reality (VR) experience. Likewise, another study (Papadopoulos and Koustriava, 2017) described how individuals with blindness constructed cognitive maps of a familiar and unknown space through audio-tactile maps, where participants presented a well-formed cognitive map of a familiar area compared to the cognitive map of an unknown area.

Because many VRA available are based on the sense of sight, sensory substitution devices had been developed (Johnson and Higgins, 2006; Kajimoto et al., 2014; Proulx et al., 2014), which enabled blind people to recognize shapes and characteristics of some objects. A study with 39 participants found significant differences in the increase of the speed of navigation and a better comprehension and satisfaction of users when using Enhanced Text Interfaces (Leuthold et al., 2008).

Some mobile applications require the use of external sound sources (Štorek et al., 2013), haptic devices and other alternative modes of interaction with the environment (Bourbakis et al., 2013; Guerreiro et al., 2017) to provide real-world information. The use of these technologies is not yet widespread, either because of the high cost of extending their physical implementation, the discomfort or complexity of their use, their inaccessible design, or the margin of error in the accuracy to preserve the person's integrity, which makes these systems undesirable for most blind people.

The limitations that users with blindness find in previously built systems, besides the costs, are related to the user interface, which include physical difficulties and cognitive barriers to understand procedures and content; the key to interface design to ease accessibility services is in mobile applications (Rodriguez-Sanchez et al., 2014).

The adoption of words and sounds is the main source of information in applications aimed to blind people (Shimomura et al., 2010). There is a mobile application used by people with blindness to detect doors and objects in the corridors using a sound interface (Moreno et al., 2012). The ARGUS European project employed binaural audio technology to provide the user with spatial information of a predefined route. Vibration is another alternative to represent information using a mobile and several attached vibration motors. Yatani (2012) built a system that offers information about the distance and the direction towards a destination for people with visual disabilities. Rodriguez (2014) described the possibilities of guiding a blind user within a route giving auditory and tactile feedback during navigation, the user makes gestures with finger movements on the touch screen and in fixed regions around the mobile's corner to read the software on the screen and navigate the menu; the system employs text-to-speech, vibration, and audio feedback as physical interfaces.

The Audio-Touch Interfaces (Gomez et al., 2012) contributed to improved spatial perception, allowing blind people to selectively explore environments. Some Haptic-Acoustic Interaction Metaphors were developed using virtual environments (VE) (Fabio et al., 2011) that provided information in three dimensions for the acquisition of the environment's description.

Our team developed mobile applications with virtual spaces taken from reality, so that blind people could virtually explore an unknown environment beforehand. Sensitive interfaces as voice, beeps and gestures (7) plus the proximity cognitive interface (9) were used to provide multisensory information to users, with the aim of evaluating the contribution of these interfaces in the construction of a cognitive map of the explored place.

1.2 The use of sensitive interfaces and virtual space for cognitive map construction

Cognitive mapping implies the creation of mental structures based on previously identified objects/spaces and memories rooted in a person's brain. Studies on the development of spatial knowledge indicate that blind people are able to form mental representations of space, equally to people with normal vision, but only if they have enough information (Kitchin, 2001). Recent evidence suggests that the type of information provided during a spatial task can improve their orientation and performance skills (Papadopoulos and Koustriava, 2017), so it will be necessary to combine absolute (objects in space are located relative to the body of the person) and relative location activities (one object in relation to other objects) for the codification of space by users. F. De Felice (2011) indicates that a VE allows to represent only the part of the world that is considered relevant for the user, which can greatly simplify the perception and interpretation efforts required by users; they also developed metaphors of interaction to facilitate access to information using VR tools.

According to a study (Sanabria, 2007) for blind people, a tactile map is the only approach for acquiring a structured knowledge to recognize spaces and places. Furthermore, it describes the creation of cognitive maps is a powerful strategy that provides spatial information for the mobility of sighted and blind people. Traditional VR focused mainly on visual feedback, which is not accessible to people with visual impairment (Zhao et al., 2018). For this reason it is important to investigate and assess if there is a combination of interfaces in VRA that provides enough support to blind people to develop a better knowledge of their environment. The present study used VRA so that visually impaired users acquire enough information about an unknown place in order to recreate that space in their mind.

Spatial mapping in a person with visual disabilities should be integral and distinctive, because the learning process of an individual is a conjunction between cognitive and perceptual skills (Majerova, 2017; Orly Lahav and Mioduser, 2008). Verbal assistance is very important to overcome many challenges, and the strategies used to understand the environment's information must be customized because blind people with no previous visual experience seem to be more dependent on egocentric strategies for coding and representing space during exploration than with allocentric strategies that are more useful to people with prior visual experience (Pasqualotto et al., 2013). Some studies mention the need to add tactile devices to guide the user in a virtual space (Papadopoulos and Koustriava, 2017; De Felice et al., 2007; Orly Lahav and Mioduser, 2008) because these provide force feedback for a better identification of structures during virtual visits. Multimodal interactions supply users with more natural ways to manipulate virtual space; the combination of auditory and tactile information seems to be able to better represent the information and, therefore, would allow a more efficient mental construction of the space.

A common approach to assess the quality of the cognitive map is to directly test its usefulness to navigate real spaces (Cobo et al., 2017; Merabet and Jaime, 2016), here quality is usually modeled as success or failure when completing a set of particular tasks. Another approach is to ask the participants to provide a verbal description of the scene and other could be one where they construct realistic or schematic models of the explored scene. In addition, there are examples of questionnaires with predefined questions that assist the analysis of dimensions of different outsourcing in isolation (Giraud et al., 2017), but there is no index capable of providing a general evaluation.

The current study will use VRA with three sensitive interfaces and a cognitive interface, with the purpose that blind people can explore an unknown place and acquire enough information about the structure of the environment and the obstacles existing inside it, and then perform tasks in the real environment equivalent to the virtual place explored.

1.3 Hypotheses

As described earlier, the act of "learning" is a conjunction between cognitive and perceptual learning, which means the construction of spatial knowledge through VRA will require the use of appropriate interfaces that provide accurate and continuous information at the conceptual and perceptual level (Cobo et al., 2017; Majerova, 2017; Lahav et al., 2015). In our study, blind people used an application to virtually explore an unknown space. As a cognitive interface, the proximity exploration was used, where the avatar walked through the virtual environment to detect obstacles. As sensitive interfaces, the voice, beeps and gestures were used to inform about the obstacles that fall within the security zones two in the front and two on the sides.

The research proposes that the combination of sensitive interfaces used during the virtual exploration of unknown spaces, allows blind users an efficient and effective exploration process to produce a reasonable cognitive map of the place visited. We tried to prove that there is an appropriate pattern for beeps, voices and gestures used during the exploration, that maximize the percentage of detection and facilitates the identification and location of obstacles in an unknown space.

It was hypothesized that a VRA that uses: double speed in the reproduction of messages; beeps with intermediate and low frequencies to alert on different types of objects; and simple gesture patterns on the mobile screen are better than other proposed combinations, because it would improve efficiency and efficacy of the exploration process. A greater efficiency would be associated with a shorter duration during the overall exploration stage. A greater effectiveness would be associated to a greater percentage of objects and structures detected in this exploration stage. It was also hypothesized that the sensitive interfaces used during the virtual exploration, would provide enough information to the user to create a mental representation of the environment, which increases the percentage of success in completing tasks in the real space. An optimal cognitive mapping would be associated with: the correct identification of obstacles and internal structures with reasonable accuracy. Finally, the usefulness of the cognitive map would be associated with success in absolute, relative and orientation tasks in real space.

1.4 Our contributions

A study was carried out to determine the utility of information delivered to the user by the VRA so he could explore a virtual space in advance. In this study, three types of sensitive interfaces were investigated separately; their combination, together with a cognitive interface, could better detect and locate obstacles in an unknown environment, enabling in this way the construction of a cognitive map of the place visited, whose utility was tested with tasks in a real environment.

We compared the differences between using the voice with three different reproduction speeds and a language adapted to the needs of the user, in the same way we compared three beep patterns with three different frequencies depending on the type of object and the distance to the object; finally we compared four types of gestures used to start or stop the exploration action and also to obtain more information about the obstacles.

Previously, for two years, we had made four workshops with blind and low vision users, to analyze general aspects of the application, correct certain errors, make the tool more effective and successful and also to choose the patterns that would be used in this study (Guerrón et al., 2018). This last study describes a full methodology for building a VRA whereas the current paper presents the results of a single final workshop specific to the analysis of sensitive interfaces.

A part of the study was presented by Cobo (2017), where it was concluded that there is no difference between the cognitive maps reached by users with remote and proximity exploration. In this paper, we present the contribution of the voice, beep and gestures sensitive interfaces, during the proximity exploration, in the detection and location of objects in the environment, for the acquisition of spatial knowledge.

Other similar studies used a haptic device and a smartphone (Orly Lahav and Mioduser, 2008; Picinali et al., 2014; Papadopoulos & Koustriava, 2017; Zhao et al., 2018) to provide individuals with blindness valuable information associated with the structure and content of space, to help building mental spatial maps. In this study, we wanted to determine which of all the proposed interfaces improved the process of detection and location of structures and objects inside an unknown environment, using only a mobile phone. This work could be a reference in the types of sensitive interfaces used in VRA for blind people, which provides greater efficiency and effectiveness during the learning of new environments.

2 MATERIALS AND METHODS

There are five sections: Section 2.1 holds a description of the participants in the study; Section 2.2 contains a description of the interfaces used. Section 2.3 describes the implementation of the virtual rooms. Section 2.4 defines study variables. Finally, the section 2.5 details the design of the study.

2.1 Participants

Twenty blind participants (11 men, 9 women, average age 52.2, SD = 14.1) tested thirty VRA (See 2.5.1). All the participants had become blind more than seven years ago. They knew how to use a mobile phone, and one of the participants had a slight deafness, which did not interfere with the development of the tests, because during the training the volume of the application was adjusted to each participant. The participants were recruited among people affiliated to ONCE (National Organization for Spanish Blind people), workers of the ONCE Foundation, and from the support service for students with disabilities of the Technical University of Madrid.

2.2 Users Interfaces

Two types of interfaces were considered: sensitive and cognitive. The sensitive interfaces studied were: voice, beeps and gestures; in addition, a vibration pattern was used to inform the participant that the avatar walked through the virtual environment. The cognitive interface chosen was proximity exploration. These interfaces were chosen based on 2 criteria.

Firstly, the authors used the information available from studies related to the research topic. Although no similar studies were found, a predilection for the usage of voice and haptic feedback was observed. Also, the selection of the interfaces took into consideration that blind people use traditional interfaces and have a preference for them (Oliveira et al., 2011). Likewise, because of the several concepts related to the presence of obstacles inside the space, it was necessary to define the appropriate terminology to report about their presence (Tönnis et al., 2013).

Another priority was to have the smartphone as the only work tool for the user; this in turn reduces the effort in the perception and interpretation of information made by the participant (Fabio et al., 2011), reduces the cost of the research and allows greater user's satisfaction since blind people appreciate discrete tools that are used daily.

In a previous research, the most representative patterns of the current interfaces were selected (Guerrón et al., 2018). The workshop tested 20 vibration patterns, 36 beeps patterns, 10 gesture patterns and 20 voice patterns with blind users.

The generated patterns were a combination of: random values, participant suggestions gathered during previous workshops and also product of the researchers reasoning. For example, high frequency beeps were associated with obstacles located more than a meter and a half high above the floor or when the user was very close to the obstacle. The lowest frequencies were associated with objects on the ground up to sixty centimeters.

Because the parameters associated with the voice interfaces such as speed, distance units and the employed terminology generated different patterns; the researches decided to ask participants from a prior workshop to choose three representative combinations. The three combinations that scored the highest in regard of importance were selected. Similarly, the chosen gestures were the four that user found the simpler to perform.

2.2.1 Sensitive Interfaces

B1-B2-B3 (Three beep patterns): Square waves of different frequency (f), duty-cycle ($0.1 \leq \delta \leq 0.2$) and spacing ($0.1s \leq d \leq 0.3s$), were used to warn of the objects located at 1m and 1.5m from the avatar, inside of the frontal security zone (See Fig. 1); these waves differ in intensity and frequency depending on the object's height and the distance from the avatar to the object as shown in Table 1.

Table 1
Beep patterns used as sensitive interface

Pattern	Distance (m)	High object	Middle object	Low object
B1	1,5	f5, δ2, d2	f3, δ2, d2	f1, δ2, d2
B1	1	f6, δ1, d1	f4, δ1, d1	f2, δ1, d1
B2	1,5	f5, δ1, d1	f4, δ2, d1	f2, δ1, d2
B2	1	f4, δ2, d3	f3, δ1, d1	f2, δ2, d1
B3	1,5	f3, δ1, d3	f4, δ1, d2	f1, δ1, d1
B3	1	f3, δ1, d2	f4, δ2, d2	f2, δ1, d3

f1=100Hz, f2=200Hz, f3=500Hz, f4=1KHz, f5=2KHz, f6=5KHz
δ1=0.1, δ2= 0.2
d1= 100ms, d2=200ms, d3=300ms

V4-V5-V6 (Three voice patterns): RTVoice application was used to generate three female voice patterns with different playback speeds, which informed about the name of nearby objects and structures that entered or left the frontal and lateral security zones of the avatar, while he was walking through the virtual space (See Fig. 1). These interfaces also reported on the start/end of the exploration, as shown in **Table 2**. The user had to allow the avatar to go through the entire virtual scenario, to structure in his memory the arrangement of internal objects and structures; in order to recognize these in real space. We tried to avoid excessive auditory information (Leuthold et al., 2008) using software functionalities.

Table 2
Voice patterns used as a sensitive interface

Pattern	Speed	Action modes	Objects within the Security Zones
V4	Normal (x)	Active/Deactivate	Object* enters/out by left/right
V5	Fast (2x)	Start/End	Object* left/right, free way left/right
V6	Very Fast (3x)	Enabled/Disabled	Object* left/right, nothing thing left/right

* Refers to the name of the object

G7-G8-G9-G10 (Four gesture patterns): The gestures on the mobile screen allowed to start the actions of walking and stopping. While the avatar was walking, the application reproduced the name of a nearby object or structure once. Also, it was possible to reproduce this information with a horizontal or vertical stroke by the user on the mobile screen, as indicated in **Table 3**, column 3.

If the user wished to remember or to know the name of the object or structure that was within the frontal security zone of the avatar, he could repeat the same gesture on the screen; this has been referred to as "object changing". The gesture was recognized in any available space of the touch screen, with a significant variation in stroke speed, regardless of the inclination of the mobile (Kane et al., 2011).

Table 3
Gesture patterns used as a sensitive interface

Pattern	Walk/Stop	Object changing
G7	Single touch	Vertical swipe
G8	Double touch	Horizontal swipe
G9	Single touch with sound	Vertical swipe with sounds

G10	1.5 seconds sustained touch	Horizontal swipe with sound

In addition, a vibration pattern was used to inform the participant that the avatar walked through the virtual environment.

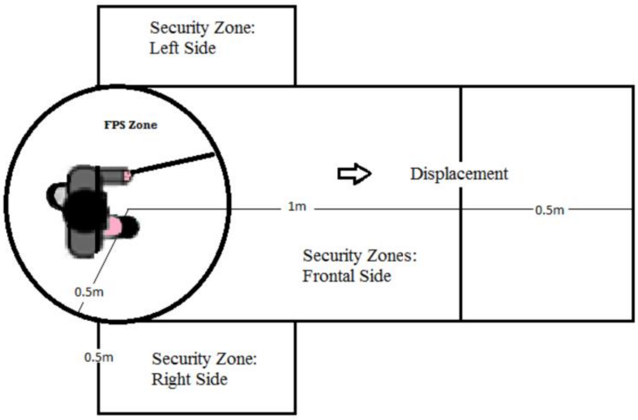

Fig. 1 Security zones activated during the exploration

2.2.2 Cognitive Interface

Proximity exploration: While the user's avatar walked through the virtual room, the user received auditory and tactile feedback. The user remained seated in a chair and could easily turn in all directions to change the orientation of the avatar. The vibration informed the user that the avatar was walking. This interface uses security zones called No-FoA and FPS Zones that are activated when the avatar begins to walk through the virtual room.

No-FoA (focus of attention): Rectangular prisms of 0.5m and 1m in length were placed on the sides and in front of the avatar (Fig. 1), which made up the security zones. These were activated during the proximity exploration. They were used to provide information and warnings to the user about nearby obstacles. These prisms allowed users to reach objects that were further from the reach of a white cane.

FPS Zone: It is a 0.5m ratio cylinder that acts as the First Personal Security Zone assigned to the user. This works as a protection to avoid the obstacles that are presented around the user similar to a white cane when walking through a room. This cylinder contains the user and moves with him when he moves around the room.

2.3 Implementation of Virtual Rooms

Since large spaces are difficult to control and the participants require sufficient time in the environment to have adequate knowledge of the indoor space (Papadopoulos and Koustriava, 2017), three appropriate work environments were defined to develop this research: an office, a pub and a bedroom. Each of these places had 10mx6mx3m and contained six objects, four walls, two windows, a column and two doors; only the pub had three doors (See Fig. 2). This number of obstacles permitted users to remember the names of objects, structures and their location in space after an exploration session (Robison and Unsworth, 2017). The position of the objects and their orientation was changed, to generate the necessary number of spaces for the test with users (**Table 6**, **Table 7** and **Table 8**); a total of nine scenarios were virtualized. To test the interface in each application and scenario a color code was assigned accordingly; the beep was black, the voice has yellow and the gesture was white.

These applications were installed on smartphones with Android versions 4.4.4 and 5.0.2. **Unity 3D, (2017)** was chosen to virtualize the rooms and build the VRA, due to its ability for modifying the program and change the structures and interior objects and also because it is cross-platform.

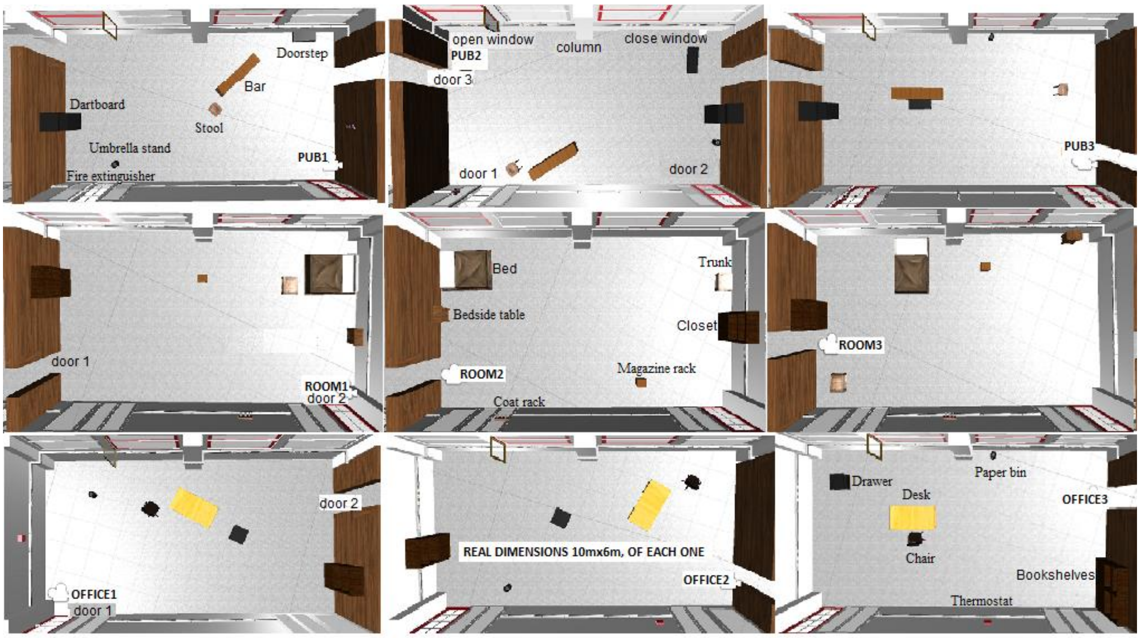

Fig. 2 Zenital plane of the virtual rooms (pub, bedroom and office)

2.4 Study variables and exploration tasks

The participants carried out three major activities: 1) Exploration of the virtual space using the smartphone. 2) Cognitive mapping using a model of cardboard and small scale objects, and finally, 3) Execution of three types of tasks (relative, absolute and orientation tasks) in a real space equivalent to the virtual room previously explored, whit the goal to know the usefulness of the cognitive map. The variables in **Table 4** were evaluated for this study.

Table 4
Variables assessed in the study

Activity	Variables	
Exploration	Efficiency	Effectiveness
Cognitive Mapping	Quality cognitive map	
Usefulness	Success in real tasks	

2.4.1 Exploration Tasks

The exploration of the virtual space allowed to measure the efficiency and effectiveness of the tool in the detection of obstacles.

Efficiency was defined as the time spent in discovering all the obstacles that are detected (structures and objects) during a given exploration session. In each session there was an inactivity time, in which useful information was not provided to the participant and was eliminated from the total time used for the exploration. See equation (1). It is possible that not all the obstacles in the scene were detected during this time, only a maximum number of them, which can be lower than the total number.

$$\boldsymbol{time}_{\boldsymbol{used\ on\ detecting}} = (\boldsymbol{time}_{\boldsymbol{exploration}} - \boldsymbol{downtime}) \qquad (1)$$

Effectiveness was defined as the number of objects and structures detected during the exploration of a virtual room. The greater the number of obstacles detected, the more effective the tool was.

2.4.2 Cognitive Mapping Process

The cognitive mapping determined the quality of the spatial information that the user acquired with the application. The examined variables were: the correct number of obstacles identified (both objects and structures), the number of objects correctly located, and the number of obstacles inserted or omitted by mistake (See Table 5). Four quadrants were assigned to place the objects within the model; the objects and pairs of objects (pairs of small scale objects) correctly located, were modeled as a dichotomous variable. The quality of the cognitive map was determined as a result of the correct identification of the scenario and internal obstacles, as well as the proper distribution of objects within the space; where 1 would be the maximum value.

Table 5
Variables for the cognitive mapping process

Nomenclature	Description	Value
I1*	correct-obstacles identified	Continuous, between 0 and 1
I2*	correct- objects located	Continuous, between 0 and 1
IS*	empty room identified	Dichotomous 1 (if model choice = success) = 0 (if model choice = failure)
N	total number obstacles	Discrete, number of the objects and structures without the empty room, max 16
n1	total number objects	Discrete, n1=6
n2	total number interior structures	Discrete, depending on the type of room 9 or 10
n3	total number correct- obstacles identified	Discrete, max 16
n4	total number correct- objects identified	Discrete, max 6
n5	total number correct-identification pairs	Discrete, max 15
E1	error in the objects identified	Continuous, between 0 and 1
EO1	error-omission	Continuous= total omissions/n1
EO2	error-insertion	Continuous=total insertions/ n1
EA	absolute-location error for objects	Dichotomous 1 (if it falls in a assigned quadrant) = 0 (otherwise)
ER	relative-location error for objects (pairs small scale objects)	Dichotomous 1 (if the relative location of two miniatures was equivalent to the relative position of the corresponding objects in the original
Q*	quality of spatial information obtained (cognitive map)	Continuous, between 0 and 1

* evaluated variables

Where I1 can be obtained, by:

$$\boldsymbol{I1} = \frac{\mathbf{n3}}{\mathbf{N}} \tag{2}$$

In the case of incorrect identification of objects, the identification error (E1) can be obtained by:

$$\boldsymbol{E1} = \frac{(\mathbf{EO1} + \mathbf{EO2})}{\mathbf{2}} \tag{3}$$

In the case of locating the objects inside the space, can be obtained by:

$$\boldsymbol{I2} = \frac{\sum_{1}^{\mathbf{n4}} \mathbf{EA}}{\mathbf{2*n4}} + \frac{\sum_{1}^{\mathbf{n5}} \mathbf{ER}}{\mathbf{2*n5}} \tag{4}$$

The Quality of mapping of space can be obtained, as follows:

$$\boldsymbol{Q} = \frac{(\mathbf{I1} + \mathbf{I2})}{\mathbf{2}} - \frac{\mathbf{E1}}{\mathbf{2}} \tag{5}$$

2.4.3 Usefulness

The usefulness of the cognitive map was evaluated with the success in the accomplishment of three tasks (absolute, relative and orientation tasks) in the real physical space equivalent of the virtual room previously explored. Success in the accomplishment of each task was modeled as a dichotomous variable, 1 was assigned when the participant reached the objective, 0 otherwise.

a. Absolute task: The participants must walk from the initial exploration point to a target point; usefulness of absolute locations in the map is assessed.
b. Relative task: The participants must walk from an object to another one; usefulness of relative locations in the map is assessed.
c. Orientation task: The participants must walk back to the starting point; participants' ability to update their maps while walking through a real room is assessed.

2.5 Study design

A within-subjects cross-sectional study was conducted. Eligible subjects are those that have become blind at least two years before the beginning of the study, suffer total blindness and are over eighteen years old.

2.5.1 Configuration of user applications

In each configuration, three physical interfaces and one cognitive interface were assigned, but only one of the physical interfaces was evaluated at a time, having the other 2 interfaces fixed. In that way, when the three voice interfaces were evaluated, the fixed interfaces were B1 for the beep and G7 for the gesture; when the three beep interfaces were evaluated, the fixed interfaces were V4 for the voice and G7 for the gesture and finally when the gesture interfaces were evaluated the interface V4 for the voice and B1 for the beep were used. The selection of V4, B1 and G7 seemed to be the most appropriate, because they were the first to be generated and tested as satisfactory by the members of the work team

Participants were involved in three experiences for each type of sensitive interface, with a total of nine experiences. During each experience, they explored a different room (office, pub, or bedroom). To minimize the impact of possible carryover effects, both the sensitive interface and room were randomized, as well as the execution sequence, according to the **Table 6**, **Table 7** and **Table 8**.

In order to test the beep interfaces (B1-B2-B3), we built nine applications, one for every virtual room and for each sensitive interface. The code of color black was assigned to locate the objects within each room; this means that several labels of this color were placed on the floor of each room. The V4 voice interface and the G7 gesture interface remained constant during the evaluation of the beep interfaces.

Table 6
Configuration of user's applications for the beep interfaces

Participant id	Experience 1	Experience 2	Experience 3
1	B1-office-black	B3-bedroom-black	B2-pub-black
2	B3-pub-black	B2-office-black	B1-bedroom-black
3	B2-bedroom-black	B1-pub-black	B3-office-black
4	B2-pub-black	B3-bedroom-black	B1-office-black
5	B1-bedroom-black	B2-office-black	B3-pub-black
6	B3-office-black	B1-pub-black	B2-bedroom-black

To evaluate the voice interfaces (V4-V5-V6), nine other applications were also required. They were assigned the yellow code to locate the objects inside the room.

The B1 beep interface and the G7 gesture interface remained constant during the evaluation of the voice interfaces.

Table 7
Configuration of user's applications for voice interface

Participant id	Experience 4	Experience 5	Experience 6
1	V4-bedroom-yellow	V6-pub-yellow	V5-office-yellow
2	V6-office-yellow	V5-bedroom-yellow	V4-pub-yellow
3	V5-pub-yellow	V4-office-yellow	V6-bedroom-yellow
4	V5-office-yellow	V6-pub-yellow	V4-bedroom-yellow
5	V4-pub-yellow	V5-bedroom-yellow	V6-office-yellow
6	V6-bedroom-yellow	V4-office-yellow	V5-pub-yellow

To evaluate the gesture interfaces (G7-G8-G9-G10), twelve applications were built. They were assigned the white color code to locate the objects inside the room. The B1 beep interface and the V4 voice interface remained constant during the evaluation of the gesture interfaces. Each participant only evaluated three of the four interfaces, due the restrictions imposed by the availability of physical spaces and the size of the research teams.

Table 8
Configuration of user's applications for gesture interface

Participant id	Experience 7	Experience 8	Experience 9
1	G7-pub-white	G8-bedroom-white	G10-office-white
2	G9-office-white	G7-pub-white	G8-bedroom-white
3	G10-bedroom-white	G9-office-white	G7-pub-white
4	G8-office-white	G10-bedroom-white	G9-pub-white
5	G7-pub-white	G8-office-white	G10-bedroom-white
6	G9-bedroom-white	G7-pub-white	G8-office-white
7	G10-pub-white	G9-bedroom-white	G7-office-white
8	G8-office-white	G10-pub-white	G9-bedroom-white
9	G7-bedroom-white	G8-office-white	G10-pub-white
10	G9-office-white	G7-bedroom-white	G8-pub-white
11	G10-pub-white	G9-office-white	G7-bedroom-white
12	G8-bedroom-white	G10-pub-white	G9-office-white

2.5.2 Procedure

All participants signed an informed consent prior to the beginning of the study. The activities were recorded by a camera, through observation sheets and through records sent over the network. A total time of eight hours per user was established to carry out all the planned experiences, with two work teams. Although there were no time constraints dedicated to training, nor to virtual exploration, users took an average of forty minutes to complete each experience. Since there were nine experiences, six were made in the morning and the other three in the afternoon. After each experience the participants had a few minutes of rest and then continued with the next work session.

At the beginning of the session, each participant received a warm welcome and a member of the research team summarized the objectives of the study. Each participant was involved in nine experiences as shown in **Table 6**, **Table 7** and **Table 8**. Each experience consisted of five stages:

1. Training: The participants used a training application, which consisted of a virtual space with few obstacles, to become familiar with the interfaces that would be used during the exploration. The virtual space Fig. 3b was constructed based on Fig. 3a. The participants received the support of a team member at all time; they were given no time constraints.
2. Exploration stage: The participant explored one of the three virtual rooms (office, pub, or bedroom) using one of the configurations showed in the **Table 6**, **Table 7** and **Table 8**. They were asked to explore the room until they felt they could describe its structures and objects. The participant began the exploration, sliding his/her finger on the touch screen (See Fig. 3c); and ended when he notified the researcher, who was responsible for closing the exploration screen. The participants had no time restrictions in this activity. In this stage the duration of the exploration, the time of inactivity, the name of the obstacle and the route taken by the user were recorded.
3. Post-test questionnaire: Each participant was asked to indicate the amount of walls, windows, doors and columns that he remembered from the previous exploration stage. The answers were manually recorded by a member of the research team.

Fig. 3 Procedure: a) real scenario, b) virtual scenario-training, c) exploration, d) model, e) tasks

4. Model: The user selected a cardboard model from three available options, only one corresponded to the empty virtual room previously explored. Twelve objects were delivered one by one to the participant, where six of them were in the room and another six did not correspond. Finally, the participant places the objects to scale within the model. A photograph of the resulting model was taken at the end (See Fig. 3d). In this stage it was recorded the: identification of the empty room, identification of objects, time dedicated to identification, location of objects and pairs of objects within the model.
5. Performing tasks: The participant was driven to the real physical space equivalent of the virtual room. At the beginning he/she was placed at the same location and position as the avatar, and then the participant was asked to complete three tasks: a relative one, an absolute one, and an orientation one. A member of the research team told to the participant toward which objects he/she must walk (Fig. 3e). The duration of the task and its success or failure was recorded.

3 EVALUATION AND RESULTS

In this study we analyzed the differences in the cognitive mapping of blind people, using sensitive interfaces during the exploration. The variables were described in section 2.4.

3.1 Exploration

3.1.1 Efficiency of the exploration process

Efficiency was modeled as time used for discovery, which was defined in section 2.4.1. The values of discovery duration with the sensitive interfaces of beep, voice and gesture under study are shown in **Table 9**.

Table 9
Average duration of the discovery stage

Beep interface			Voice interface			Gesture interface			
B1	B2	B3	V4	V5	V6	G7	G8	G9	G10
597s	526s	681s	705s	591s	982s	692s	725s	1053s	775s

The results of the statistical analysis, performed in each group of sensitive interfaces, are shown below. The dependence of discovery duration with the sensitive beep interface was assessed with a Friedman test that resulted no statistically significant. The sensitive interface B2 was achieved by a shorter detection duration.

Table 10
Statistic analysis on efficiency: Beep Interface

Test	B1	B2	B3
Shapiro-Wilk	p=0	p=0.29	p=0.002
Friedman	Chi-square=0.77, df=2, p-value=0.84		

The dependence of discovery duration with the sensitive voice interface was assessed with a Friedman test that resulted statistically significant. The pairwise comparison between the types of voice-sensitive interfaces was performed by applying a Tukey test.

Table 11
Statistic analysis on efficiency: Voice interface

test	V4	V5	V6
Shapiro-Wilk	p=0.94	p=0.88	p=0.26
Friedman	Chi-square=6.33, df=2, p-value=0.042		
Tukey	V4-V5: p-value=0.114	V5 obtained a 39.72% of reduction time compared to V6	
	V4-V6: p-value=0.92	V5-V6: p-value=0.043	

The dependence of discovery duration with the sensitive gesture interface was assessed with a Friedman test that resulted statistically significant. The pairwise comparison between the types of voice-sensitive interfaces was performed by applying a Tukey test.

Table 12
Statistic analysis on efficiency: Gesture interface

test	G7	G8	G9	G10
Shapiro-Wilk	p=0.99	p=0.71	p=0.82	p=0.78
Friedman	Chi-square=6.33, df=3, p-value=0.038			
Tukey	G7-G9: p-value=0.0019	G8-G9: p=0.0028		G7-G9: p=0.007
	G7 entailed a 34.28% of reduction in discovery time compared to G9	G8 entailed a 31.1% of reduction compared to G9		G10 entailed a 26.59% of reduction compared to G9

3.1.2 Effectiveness of the exploration process

It was defined as the percentage of objects and structures detected during the exploration of a virtual room (See **Table 13**).

Table 13
Average effectiveness of the exploration stage

Beep interface			Voice interface			Gesture interface			
B1	B2	B3	V4	V5	V6	G7	G8	G9	G10
82.16%	80.12%	81.11%	89.61%	94.04%	88.24%	90.14%	93.33%	90.27%	97.22%

The dependence of effectiveness with the sensitive beep interface was assessed with a Friedman test that resulted no statistically significant. The sensitive interface B1 reached a higher percent in obstacles detection.

Table 14
Statistic analysis on effectiveness: Beep Interface

test	B1	B2	B3
Shapiro-Wilk	p=0.894	p=0.969	p=0.638
Friedman	Chi-square=0.87, df=2, p-value=0.63		

The dependence of effectiveness with the sensitive voice interface was assessed with a Friedman test that resulted statistically significant. The pairwise comparison between the types of voice-sensitive interfaces was performed by applying a Tukey test.

Table 15
Statistic analysis on effectiveness: Voice interface

test	V4	V5	V6
Shapiro-Wilk	p=0.568	p=0.485	p=0.282
Friedman	Chi-square=6.816, df=2, p-value=0.033		
Tukey	V4-V5: p-value=0.033	V5 improved a 4.6% on detection obstacles compared to V4	
	V4-V6: p-value=0.17	There was not found to statistically significant	
	V5-V6: p-value=0.032	V5 improved a 5.9% on detection obstacles compared to V6	

The dependence of effectiveness with the sensitive gesture interface was assessed with a Friedman test that resulted no statistically significant. G10 interface achieved highest percentage concerning obstacles detection.

Table 16
Statistic analysis on effectiveness: Gesture interface

test	G7	G8	G9	G10
Shapiro-Wilk	p=0.620	p=0.156	p=0.683	p=0.196
Friedman	Chi-square=1.5, df=3, p-value=0.628			

3.2 Cognitive Mapping Process

The selection of the cardboard model and the selection and location of the miniature objects within the model are summarized in **Table 17**.

Table 17
Summary of the average percentages obtained in the selection, identification and location of obstacles.

Interface	Is			I1			I2
	To (%)	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)
B1	73,68	97,37	75,88	88,95	68,42	59,26	6,67
B2	52,63	90,79	84,21	77,63	68,42	75,44	1,67
B3	68,42	96,05	76,32	73,68	50,00	71,93	2,22
V4	60,32	95,99	76,42	71,11	67,20	80,73	14,02
V5	66,67	98,81	83,25	77,79	85,71	86,51	21,42

V6	65,56	97,22	88,00	88,33	97,22	75,93	12,50
G7	48,33	94,17	70,28	86,94	76,67	75,56	23,33
G8	70,00	93,75	83,75	98,33	96,67	83,33	29,44
G9	86,67	98,33	95,00	93,33	93,33	87,78	23,33
G10	70,00	91,67	72,53	77,55	81,67	77,55	29,63

T0: Selection of the cardboard model
T1-T5: Identification of walls, doors, windows, columns and objects inside.
T6: Location of objects (absolute and relative).

The error in the identification of objects was obtained by means of equation (3), this is showed in the **Table 18**

Table 18
Average percentage of objects that have been incorrectly identified (E1)

B1	B2	B3	V4	V5	V6	G7	G8	G9	G10
9.17	8.33	7.46	4.82	3.57	6.94	3.33	3.33	4.17	5.0

3.2.1 Identification of obstacles (I1), location of objects (I2), identification of empty room (Is)

The identification of obstacles was assessed by means of equation (2). The location of objects was evaluated by means of equation (4). We used a Friedman's Test and repeated measures ANOVA, to assess the dependence in the identification of obstacles and location of objects with the beep, voice and gesture interfaces. The identification of empty room was modeled as a dichotomous variable. The result was assessed with a Mantel-Haenszel test of general association. The results are shown in the **Table 19**

Table 19
Results: Identification of obstacles (I1), location objects (I2) and identification empty room (Is)

Variable	Sensitive Beep	Sensitive Voice	Sensitive Gesture
I1	Q=0.287, DF=2, p-value=0.752	F=4.434, DF=2, p-value=0.0418 Mauchly's W =0.63, p=0.4; Shapiro-Wilk: pV4=0,26149; pV5=0,167; pV6=0,48819; Bartlett's statistic=2.85, p-value=0.23	Q=0.85, DF=3, p=0.837
I2	Q=2, DF=2, p-value=0.368	Q=1.4, DF=2, p-value=0.497	Q=4.515, DF=3, p-value=0.211
Is	M^2=2, DF=2, p-value=0.36	M^2=4.33, DF=2 p-value=0.11	M^2=2.4, DF=3, p-value=0.494

The results did not shown statistically significant difference, except with the voice interface in the identification of obstacles. Pairwise comparison between sensitive voice interfaces was conducted by applying a Tukey pairwise comparison test. V5 showed a statistically significant difference with V4. In particular, there is a 10.7% improvement in the identification of obstacles compared to V4 (p-value=0.033). There is no statistically significant difference between V4 and V6 (p-value= 0.30), nor between V5 and V6 (p-value=0.37).

The absence of dependence in the case of the location of objects and in the identification of the empty room was particularly intriguing because effectiveness of obstacle detection during an exploration did show dependence with the voice interfaces (See section 3.1.2); so, the reasons why the identification dimension does not preserve this dependence were studied more carefully. We observed that the identification of walls (**Table 17**, column 3) reached the highest percentage in relation to the rest of the obstacles; then we decided to compare the number of walls detected with respect to the number of walls identified (See Fig. 4)

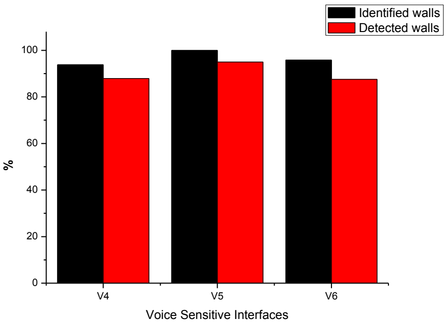

Fig. 4 Detection of walls vs. Identification of walls for voice sensitive interfaces

Both the detection and identification statuses of walls were modeled as dichotomous variables. The dependences of both statuses with the voice interfaces were assessed each with a Mantel-Haenszel test of general association, which resulted significant in the case of detection status (M^2=6, DF =2, p-value=0.049); but not significant for identification status (M^2=2, DF=2, p-value=0.368). The result shows that 75% of fully correct identifications are independent of the number of walls detected during exploration.

3.2.2 Quality of cognitive map

The quality of the spatial information collected during a proximity exploration was assessed with equation (5). The higher the value of the score obtained, the better the quality of the cognitive map. Fig. 5 summarizes the mean overall quality of cognitive maps scores for the sensitive interfaces under study.

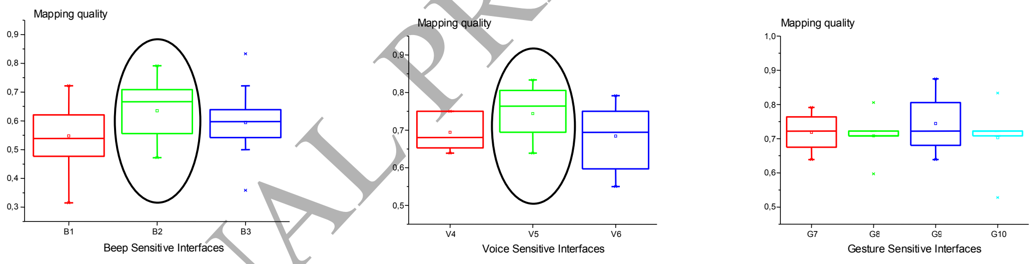

Fig. 5 Quality of cognitive map using beep, voice and gesture sensitive interfaces

The dependence of the quality of cognitive map with the three interfaces was assessed with repeated measures ANOVA, after checking assumptions for this kind of test. The results are shown in the **Table 20**, **Table 21** and **Table 22**.

Table 20
Statistic analysis on quality: beep interface

Test	B1	B2	B3
Shapiro-Wilk	p=0.927	p=0.057	p=0.436
Mauchly	W=0.812, p=0.17		
Bartlett	Bartlett's statistic=0.679, df=2, p-value=0.71		
ANOVA	F=8.02, df=2, p-value=0.0013		
Tukey	B2-B1: p-value=0.00085	B2 improved a 8.9% in cognitive mapping compared to B1	
	B1-B3: p-value=0.097	There was not found to statistically significant	
	B2-B3: p-value=0.162	There was not found to statistically significant	

Table 21
Statistic analysis on quality: voice interface

Test	V4	V5	V6
Shapiro-Wilk	p=0.39	p=0.88	p=0.73
Mauchly	W=0.6, p=0.279		
Bartlett	Bartlett's statistic=1.002, df=2, p-value=0.605		
ANOVA	F=4.1, df=2, p-value=0.043		
Tukey	V5-V6: p-value=0.0492	V5 improved a 6% in cognitive mapping compared to V6	
	V5-V4: p-value=0.107	There was not found to statistically significant	
	V4-V6: p-value=0.892	There was not found to statistically significant	

Table 22

Statistic analysis on quality: gesture interface

Test	G7	G8	G9	G10
Shapiro-Wilk	p=0.85	p=0.49	p=0.802	p=0.27
Mauchly	W=0.3913 p=0.768			
Bartlett	Bartlett's statistic=1.34, df=3, p-value=0.718			
ANOVA	F=0.328, df=3, p-value=0.8049 There was not found to statistically significant			

3.3 Cognitive map usefulness

The usefulness of a cognitive map was evaluated with the success in the completion of the absolute, relative and orientation tasks within the real scenario.

Success was modeled as a dichotomous variable for each type of task. The dependence of success for each sensitive interface was assessed with a Mantel-Haenszel test of general association. The results are shown in the **Table 23**.

Table 23
Fulfillment of the tasks

Task	Sensitive Beep	Sensitive Voice	Sensitive Gesture
Absolute	M^2=11.39,df=2,p-value=0.57	M^2=22.4,df=2,p-value=0.049	M^2=18.48,df=3, p-value=0.071
Relative	M^2=19.40,df=2,p-value=0.11	M^2=20.38,df=2,p-value=0.08	M^2=16.81,df=3, p-value=0.011
Orientation	M^2=17.89,df=2,p-value=0.16	M^2=19.58,df=2,p-value=0.10	M^2=19.80, df=3, p-value=0.051

The dependence of success with the beep sensitive interface did not result statistically significant for any of the tasks.

In the case of absolute tasks the voice interfaces resulted statistically significant. The pairwise comparison between the types of voice-sensitive interfaces was performed by applying a Tukey test. The sensitive interface V5 entails a 24% (p-value=0.048) improvement at carrying out the absolute task compared to sensitive interface V4 and a 17.5% (p-value= 0.049) compared to sensitive interface V6. There was not a significant statistical difference between V4 and V6 (p-value=0.67).

In the case of relative tasks the sensitive interface gestures resulted statistically significant. The pairwise comparison between the types of gesture interfaces was performed by applying a Tukey test. The sensitive interface G9 entailed a 28.57% (p-value=0.018) improvement at carrying out the task, compared to sensitive interface G7 and a 16.6% (p-value=0.003) compared to sensitive interface G10. There is no statistically significant difference between G8 and G9, (p-value = 0.23). The sound of the touch could help to participants in the better identification the objects and structures indoor. In the case of orientation tasks the gesture sensitive interface resulted not statistically significant.

4 DISCUSSION

It seems this is one of the first works that compares the use of sensitive interfaces during proximity exploration in virtual reality applications for blind people.

At the end of this study a greater efficiency associated to proximity exploration was evidenced when the voice interface V5 and the gesture interface G7 were used. The voice interface V5 reduced the duration of the exploration time by 39.72% (3.1.1) compared to V6, and the gesture interface G7 reduced it by 34.28% (3.1.1) compared to G9. This evidence supports our initial hypothesis of greater efficiency associated with higher rates of voice reproduction and simple gesture patterns. The sensitive voice interfaces had two variants in voice reproduction: the speed and the terminology used to say that an action was performed or an object was present. We assume that the whole combination of speed and terminology produced the final result and not separately; however, the individual effect of each variant could be studied later. Here it is important to mention that reproduction speeds higher than twice the normal speed did not show greater efficiency, but in fact has the opposite effect. This was due to the fact that an increase higher than double the speed deteriorates the quality of the message received. The beep interfaces did not show statistically significant difference in the reduction of the exploration time; however with the beep interface B2 the lowest exploration time was obtained.

Evidence also confirmed a greater efficiency due to the use of sensitive interface V5 during the proximity exploration. This interface showed an increase in obstacle recognition of 5.9% (3.1.2). However, despite showing better effectiveness, the error dimension in the location of objects and in the recognition of the empty room showed no significant differences. There was deterioration in the conversion rate of detected walls versus identified walls, this does not happen with doors and windows. The beep, voice and gesture interfaces showed a conversion rate over 100% as well. Our interpretation is that blind people tend to rely on their preconceived ideas regarding the usual structure of a room (Sun et al., 2008); hence they reported the correct number of walls regardless how many of them actually detected during the exploration.

Sensitive beep interfaces did not show any statistically significant difference during the exploration process, however these were statistically significant with the quality of the cognitive map, and despite the limited sample-power there is a rather strong statistically significant difference (p-value=0.0013). The sensitive interface B2 entailed an improvement of 8.9% compared to B1 (p-value = 0.00085). We interpret this as the fact that the variation in frequency to generate warnings during the proximity exploration has a positive impact on the correct identification and location of objects. In addition B2 provided less exploration time and greater percentages of obstacles detected. It also corroborates our hypothesis that the mid and lower frequencies reproduced with the B2 interface were fully accepted and differentiated by the participants.

A dependence of the quality of the cognitive map was observed with the V5 interface (3.2.2) during the proximity exploration; supporting the hypothesis previously observed. V5 interface allowed blind people to identify a greater number of obstacles while spending less time in the exploration process; also although the identification error is reasonably low most objects were not correctly placed inside the model. The use of sensitive interfaces for the proximity exploration may require longer training periods to better capture the nature of the information regarding obstacle location; also other feedback configurations could be used to reduce this error.

The success in the absolute tasks shows a dependence with the V5 sensitive interface (p = 0.049) and for the relative tasks there is a dependency with the G9 sensitive interface (p = 0.011) that did not show dependence in the detection nor identification of obstacles; we believe that this fact is due to the user associating the sequential position of the obstacles with a sound instead of the gesture. There was no dependence on the success of the orientation tasks with any of the interfaces proposed in the research. Orientation tasks implied physically walking around the room to update the cognitive map of the blind user. Some users recalled the starting point which resulted in a greater number of successful orientation tasks and was not related to use of their cognitive map. There is a tendency in the success at the relative tasks depending on the voice-sensitive interface V5 (p = 0.08), and in the success at the absolute and orientation tasks depending on the gesture-sensitive interface G9 (p = 0.071 and 0.051), this should be studied further with a greater number of users.

In a proximity exploration study using a multi-sensory virtual environment for people who are blind, (Orly Lahav and Mioduser, 2008) reported mean exploration times of 10 min for a room of similar size and complexity to our three spaces. These average times of exploration are similar to those found in our study (beep=10 min, voice=14 min, gesture= 12min). The main difference in the two studies was proximity exploration mechanisms. The previous research had participants identifying obstacles by learning their shape via a haptic feedback editor using the vibration to identify objects. In our study obstacle functionality was directly reported by the feedback interface via voice messages and vibration was used to warn of avatar’s movement in the virtual world. In our study the vibration became transparent for several users during proximity exploration, because their interest stayed focused on the voice interface (Shimomura et al., 2010).

Obstacles identification results in our study are similar or better than those reported (O. Lahav and Mioduser, 2008): objects (84% vs. beep=81%, voice=89%, gesture=90%), doors (46% vs. beep=79%, voice=83%, gesture=80%), windows (25% vs. beep=80%, voice= 79%, gesture=89%), and columns (48% vs. beep=62 %, voice= 83%, gesture=87%); they do not report figures for wall detection. Regarding the selection of the model, we have a lower percentage (95% vs. beep = 65%, Voice = 64%, gesture = 68.75%). The main difference in model-choice stage between the two studies was that the former one offered three options with different shapes and sizes while in our study shape and sizes remain the same and changes were only observable in the disposition of doors and windows. They also report better location of miniatures within the models (46% vs. beep=4%, voice=16%, gesture=26%) and similar success in real tasks: absolute (81% vs. beep=51%, voice=77%, gesture=87%), relative (71% vs. beep=54%, voice=75%, gesture=87%). Other previous studies could not be compared because different spaces and methodologies were used. Coming from a within-subjects design, the results are adequate to eliminate the individual differences between the participants as a source of error. The sample size used in this research was similar in magnitude to other previous studies based on proximity exploration. The low sampling power increases the probability of making a mistake by not rejecting the null hypothesis. However, a study with a larger sample and a design between subjects would be appropriate to improve the sampling power and overcome the restrictions imposed by the limited availability of physical spaces and members of the research team.

These results correspond to a low sampling power that increases the probability of extreme results, however, the type I error is still low. We believe that a study with a larger number of users and another type of design should be carried out to corroborate the results.

5 LIMITATIONS

The parameters associated with voice interfaces were the speed of voice reproduction and the terminology used to define distances and objects. We demonstrated that one of the combinations tested was relevant in this study; however, the use of certain terms or speed could affect the results. We believe that it is possible that the final results may be affected by the influence or predominance of any of these parameters.

The influence of vibration in this study was not considered, since all applications used vibration to keep the user informed about the movement of the avatar.

Participants tested nine applications during a work day and also trained with three others applications. Three participants did not complete the last two experiments. One of them was tired and the others were sharing their concerns and expectations with the team for longer than expected. The fatigue of the participants in the execution of the tasks could affect the results obtained.

V4 and G7 interfaces were fixed to test the beep interfaces, V4 and B1 were fixed to test the gesture interfaces, and B1 and G7 were fixed to test the voice interfaces. Since this settings were necessary to the configuration of the study. There is a possibility that some users may have more exposed to certain interfaces, which could also affect the results. The research design was within-subjects cross-sectional to reduce these effects.

The values obtained in the study may not be optimal, but they serve as the basis for the next studies, which are developed in this line of research.

6 CONCLUSIONS

Virtual visits in advance for the blind require training with a mobile to become familiar with the interfaces implemented. These visits allowed the participants to gain spatial knowledge of an interior unknown space before visiting it.

The users' previous knowledge (Oliveira et al., 2011) appears to affect the identification of structural elements of the model. Although prior each experience there was training using a test application, a change in the interaction with the gestures interface required greater effort on the participants. Strokes or wrong touches were observed on the screen when trying to initiate an action or while obtaining information. It seems that the duration of the training periods was insufficient.

We hypothesized that the voice messages reproduced at a higher speed, the use of low and medium frequency beeps to warn about objects and the simple gestures on the mobile screen improve the effectiveness and efficiency of the exploration process. Likewise we hypothesized that a cognitive map of good quality would be associated with the correct identification of obstacles and a reasonable accuracy of the location of obstacles within the model. Finally, we hypothesized that the utility of the cognitive map would be associated with success to reach the objective points in the real environment.

Based on our study we reached the following conclusions:

A greater efficiency during exploration with VRA was achieved when using double the speed to reproduce voice messages, and to inform about obstacles that entered and left the security zones of the avatar (V5). Efficiency was also improved when the user gave a one-touch gesture on the mobile screen (G7) to start the exploration or to stop the avatar while walking through the virtual environment. The frequencies of beeps of 200Hz, 500Hz and 1KHz to inform about low, medium and high objects obtained the shortest duration; however these did not show any statistically significant difference.

The V5 sensitive interface is also associated with greater effectiveness during the exploration stage, greater identification of the obstacles, better quality of the cognitive map and greater success in carrying out the absolute tasks.

The location of the objects in the model did not show good results during the cognitive mapping, nor was it statistically significant for any of the implemented interfaces. Then, other alternatives, such as vibration variation, could be considered in the next study to reinforce the information about the location of the objects.

The B2 sensitive interface was associated with a better cognitive map. The user made a better identification of the objects and made fewer errors in the location of these when the application generated different frequencies to differentiate one from another type of object.

The G9 sensitive interface was associated with greater success in performing the relative task, a trend is observed with the other tasks, further studies are necessary to corroborate this result.

The sample size in our study and the within-subjects nature makes variables susceptible to adverse carry-over effects. A study with a larger sample and a between-subjects design would be appropriate to address previously mentioned drawbacks.

ACKNOWLEDGMENTS

This research was supported by Cátedra Indra-Fundación Adecco (e-Glance project). We thank to the ONCE Foundation for its collaboration with the participants and the facilities where the tests were conducted. Special thanks to blind volunteers for spending their time and effort in this study.

BIOGRAPHY

Akhavian, R., Behzadan, A.H., 2016. Smartphone-based construction workers' activity recognition and classification. Autom. Constr. https://doi.org/10.1016/j.autcon.2016.08.015

Bourbakis, N., Makrogiannis, S.K., Dakopoulos, D., 2013. A System-Prototype Representing 3D Space via Alternative-Sensing for Visually Impaired Navigation. IEEE Sens. J. 13, 2535–2547. https://doi.org/10.1109/JSEN.2013.2253092

Cobo, A., Guerron, N.E., Martín, C., Serrano, J.J., del Pozo, F., 2017. Computers in Human Behavior Differences between blind people's cognitive maps after proximity and distant exploration of virtual environments 77, 294–308. https://doi.org/10.1016/j.chb.2017.09.007

De, F., Renna, F., Attolico, G., Distante, A., 2011. Hapto-Acoustic Interaction Metaphors in 3D Virtual Environments for Non-Visual Settings, in: Virtual Reality. InTech. https://doi.org/10.5772/13116

De Felice, F.F., Renna, F., Attolico, G.G., Distante, A., 2007. A haptic/acoustic application to allow blind the access to spatial information. Proc. - Second Jt. EuroHaptics Conf. Symp. Haptic Interfaces Virtual Environ. Teleoperator Syst. World Haptics 2007 310–315. https://doi.org/10.1109/WHC.2007.6

Espinoza, M., Sánchez, J., De Borba Campos, M., 2014. Videogaming interaction for mental model construction in learners who are blind. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8514 LNCS, 525–536. https://doi.org/10.1007/978-3-319-07440-5_48

Fabio, D., Floriana, R., Giovanni, a, 2011. Hapto-Acoustic Interaction Metaphors in 3D Virtual Environments for Non-Visual Settings. Virtual Real.

Fernandes, H., Conceição, N., Paredes, H., Pereira, A., Araújo, P., Barroso, J., 2012. Providing accessibility to blind people using GIS. Univers. Access Inf. Soc. 11, 399–407. https://doi.org/10.1007/s10209-011-0255-7

Giraud, S., Brock, A.M., Macé, M.J.M., Jouffrais, C., 2017. Map learning with a 3D printed interactive small-scale model: Improvement of space and text memorization in visually impaired students. Front. Psychol. 8, 1–10. https://doi.org/10.3389/fpsyg.2017.00930

Gomez, J.D., Bologna, G., Pun, T., 2012. Spatial awareness and intelligibility for the blind, in: Proceedings of the 2012 ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA ’12. ACM Press, New York, New York, USA, p. 1529. https://doi.org/10.1145/2212776.2223667

Guerreiro, J., Ahmetovic, D., Kitani, K.M., Asakawa, C., 2017. Virtual Navigation for Blind People: Building Sequential Representations of the Real-World. Assets’17 280–289. https://doi.org/10.1145/3132525.3132545

Guerrón, N., Cobo, A., Serrano, J.J., 2018. METHODOLOGY FOR BUILDING VIRTUAL REALITY MOBILE APPLICATIONS FOR BLIND PEOPLE ON ADVANCED VISITS TO UNKNOWN INTERIOR SPACES, in: IADIS 2018 International Conference. pp. 3–15.

Johnson, L.A., Higgins, C.M., 2006. A Navigation Aid for the Blind Using Tactile-Visual Sensory Substitution, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 6289–6292. https://doi.org/10.1109/IEMBS.2006.259473

Kajimoto, H., Suzuki, M., Kanno, Y., 2014. Tactile vision substitution system for palm using electro-tactile display and smartphone. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8619, 469–471. https://doi.org/10.1007/978-3-662-44196-1

Kane, S.K., Wobbrock, J.O., Ladner, R.E., 2011. Usable Gestures for Blind People : Understanding Preference and Performance. Gesture 413–422.

Kitchin, R., 2001. Cognitive Maps. Int. Encycl. Soc. Behav. Sci. https://doi.org/https://doi.org/10.1016/B0-08-043076-7/02531-6

Koutny, R., Miesenberger, K., 2016. Computers Helping People with Special Needs 9759, 253–256. https://doi.org/10.1007/978-3-319-41267-2

Kwon, J., Lee, Y., 2016. Serious games for the job training of persons with developmental disabilities. Comput. Educ. 9, 328–339.

Lahav, O., Mioduser, D., 2008. Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind 24, 1139–1155. https://doi.org/10.1016/j.chb.2007.04.003

Lahav, O., Mioduser, D., 2008. Haptic-feedback support for In the case of absolute tasks of unknown spaces by people who are blind. Int. J. Hum. Comput. Stud. 66, 23–35. https://doi.org/10.1016/j.ijhcs.2007.08.001

Lahav, O., Schloerb, D.W., Kumar, S., Srinivasan, M.A., 2012. A Virtual Environment for People Who Are Blind - A Usability Study. J. Assist. Technol. 6. https://doi.org/10.1108/17549451211214346

Lahav, O., Schloerb, D.W., Srinivasan, M.A., 2015. Rehabilitation Program Integrating Virtual Environment to Improve Orientation and Mobility Skills for People Who Are Blind. Elsiever Comput. 80, 1–14. https://doi.org/10.1016/j.compedu.2014.08.003

Leuthold, S., Bargas-Avila, J.A., Opwis, K., 2008. Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users. Int. J. Hum. Comput. Stud. 66, 257–270. https://doi.org/10.1016/J.IJHCS.2007.10.006

Majerova, H., 2017. The Person in a Situation of Visual Impairment and its Perception and Imagination from the Qualitative Viewpoint. Procedia - Soc. Behav. Sci. 237, 751–757. https://doi.org/10.1016/j.sbspro.2017.02.117

Malecki, E.J., 2017. Socio-Economic Planning Sciences Real people , virtual places , and the spaces in between *. Socioecon. Plann. Sci. 58, 3–12. https://doi.org/10.1016/j.seps.2016.10.008

Martinez, M., Constantinescu, A., Schauerte, B., Koester, D., Stiefelhagen, R., 2014. Cognitive evaluation of haptic and audio feedback in short range navigation tasks, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer International Publishing, pp. 128–135. https://doi.org/10.1007/978-3-319-08599-9_20

Merabet, L.B., Jaime, S., 2016. Universal Access in Human-Computer Interaction. Users and Context Diversity 9739, 595–606. https://doi.org/10.1007/978-3-319-40238-3

Miesenberger, K., Fels, D., Archambault, D., Peñaz, P., Zagler, W., 2014. Computers Helping People with Special Needs, in: Springer (Ed.), 14th International Conference, ICCHP 2014. Paris, p. 640. https://doi.org/10.1007/978-3-319-08599-9

Moreno, M., Shahrabadi, S., José, J., du Buf, J.M.H., Rodrigues, J.M.F., 2012. Realtime Local Navigation for the Blind: Detection of Lateral Doors and Sound Interface. Procedia Comput. Sci. 14, 74–82. https://doi.org/10.1016/j.procs.2012.10.009

Nincarean, D., Bilal, M., Dayana, N., Halim, A., Abdul, H., 2013. Mobile Augmented Reality : the potential for education 103, 657–664. https://doi.org/10.1016/j.sbspro.2013.10.385

Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., Gonçalves, D., 2011. Blind People and Mobile Touch-based Text-Entry: Acknowledging the Need for Different Flavors. Proc. 13th Int. ACM SIGACCESS Conf. Comput. Access. - ASSETS ’11 179. https://doi.org/10.1145/2049536.2049569

Papadopoulos, K., Koustriava, E., 2017. Computers in Human Behavior Cognitive maps of individuals with blindness for familiar and unfamiliar spaces : Construction through audio-tactile maps and walked experience. Comput. Human Behav. 75, 376–384. https://doi.org/10.1016/j.chb.2017.04.057

Pasqualotto, A., Jane, M., Jansari, A.S., Proulx, M.J., 2013. Visual experience facilitates allocentric spatial representation. Behav. Brain Res. 236, 175–179. https://doi.org/10.1016/j.bbr.2012.08.042

Picinali, L., Afonso, A., Denis, M., Katz, B.F.G., Denis;, M., Katz, B.F.G., 2014. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. Elsiever Int. J. Human-Computer Stud. 72, 393–407. https://doi.org/10.1016/j.ijhcs.2013.12.008

Proulx, M.J., Brown, D.J., Pasqualotto, A., Meijer, P., 2014. Multisensory perceptual learning and sensory substitution. Neurosci. Biobehav. Rev. 41, 16–25. https://doi.org/10.1016/j.neubiorev.2012.11.017

Robison, M.K., Unsworth, N., 2017. Working memory capacity , strategic allocation of study time , and value-directed remembering. J. Mem. Lang. 93, 231–244. https://doi.org/10.1016/j.jml.2016.10.007

Rodriguez-Sanchez, M.C., Moreno-Alvarez, M.A., Martin, E., Borromeo, S., Hernandez-Tamames, J.A., 2014. Accessible smartphones for blind users: A case study for a wayfinding system. Expert Syst. Appl. 41, 7210–7222. https://doi.org/10.1016/j.eswa.2014.05.031

Sanabria, L.B., 2007. Mapeo cognitivo y exploración háptica para comprender la disposición del 45–65.

Sánchez, J., 2010. Una Metodologia para Desarrollar y Evaluar la Usabilidad de Entornos Virtuales Basados en Audio para el Aprendizaje y la cognición de Usuario Ciegos. RIED - Rev. Iberoam. Educ. a Distancia 13, 265–293.

Sánchez, J., de Borba Campos, M., Espinoza, M., Merabet, L.B., 2014. Audio Haptic Videogaming for Developing Wayfinding Skills in Learners W ho are Blind. IUI - Users Motion 199–207. https://doi.org/dx.doi.org/10.1145/2557500.2557519

Shimomura, Y., Hvannberg, E.T., Hafsteinsson, H., 2010. Accessibility of audio and tactile interfaces for young blind people performing everyday tasks. Univers. Access Inf. Soc. 9, 297–310. https://doi.org/10.1007/s10209-009-0183-y

Spinosa, M.A.N.E., Ngar, S.I.U., Cha, E.S.O., Espinosa, M.A., Ungar, S., Ochaíta Esperanza, Blades Mark, Spencer Christopher, 1998. COMPARING METHODS FOR INTRODUCING BLIND AND VISUALLY IMPAIRED PEOPLE TO UNFAMILIAR URBAN ENVIRONMENTS. Elsevier. J. Environ. Psychol. 18, 277–287. https://doi.org/http://dx.doi.org/10.1006/jevp.1998.0097

Štorek, D., Rund, F., Baráth, T., Vítek, S., 2013. Virtual Auditory Space for Visually Impaired - Methods for Testing Virtual Sound Source Localization. Proc. ICAD 2013 33–36.

Sun, P.C., Tsai, R.J., Finger, G., Chen, Y.Y., Yeh, D., 2008. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Comput. Educ. 50, 1183–1202. https://doi.org/10.1016/j.compedu.2006.11.007

Tönnis, M., Plecher, D.A., Klinker, G., 2013. Computers & Graphics Representing information – Classifying the Augmented Reality presentation space $. Comput. Graph. 37, 997–1011. https://doi.org/10.1016/j.cag.2013.09.002

Unity 3D, 2017. Unity 3D [WWW Document]. URL https://unity3d.com

Yatani, K., Banovic, N., Truong, K.N., 2012. SpaceSense: representing geographical information to visually impaired people using spatial tactile feedback. Proc. CHI 2012 415–424. https://doi.org/10.1145/2207676.2207734

Zhao, Y., Bennett, C.L., Benko, H., Cutrell, E., Holz, C., Morris, M.R., Sinclair, M., 2018. Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. https://doi.org/10.1145/3173574.3173690

Graphical abstract