

Virtual reality for the human-centred design of assistive devices

Prof. Ph.D Emilia Villani^{a,*}, Ph.D. Edmar Thomaz da Silva^a and M.Sc Ivan de Souza Rehder^a

^aInstituto Tecnológico de Aeronáutica, Praça Marechal Eduardo Gomes, 50, São José dos Campos, SP, 12228-900, Brazil

ARTICLE INFO

Keywords:

virtual reality
human factors
human machine interface

ABSTRACT

Society has developed technology to create autonomous vehicles and to connect different devices and machinery to exchange data and optimize production efficiency. With this technology, soon, it will be possible to achieve better methods to guide blind and visually impaired (BVI) users in their daily activities. We believe that the available products in the market have several limitations and do not satisfy BVI users and that one of the reasons behind this problem is that they are not members of the development team or are not consulted by these.

The purpose of paper is to use virtual reality (VR) to test and evaluate different designs of BVI products. Also to verify if BVI and non-BVI users have the same mental demand and situation awareness when using assistive products. The idea is to use VR as a testing ground where a BVI user can try different assistive solutions in different scenarios. To illustrate the proposed method, a case study of navigation of BVI users inside a medical clinic is performed.

The scenes were made using Unity3D and the VR device was the Tobii Eye Tracking VR. Based on the current situation in the virtual environment, inputs are provided to the user using aural commands and haptics devices. To assess the mental workload, physiological sensors, from TEA Captiv T-Sens, are used. Among them, are an electrocardiogram sensor (ECG), to gather heart-rate and heart-rate variance data, and a galvanic skin response sensor (GSR), to collect skin conductance. Besides these sensors, the users are also expected to answer mental workload assessment tests and situation awareness questionnaires.

Among the proposed method's expected benefits are the flexibility and agility to create different scenarios, and also the possibility to test all of them in the same physical room.

1. Introduction

According to the World Health Organisation (WHO) on the , there are at least 2.2 billion people with some visual impairment degree (wor (2019). Among them, 43,3 million are classified as blind and 295 million have moderate or severe vision impairment. In order to be fully integrated into our society, they rely on assistive devices, such as canes, braille speakers, among others Bourne, Steinmetz, Flaxman, Briant, Taylor, Resnikoff, Casson, Abdoli, Abu-Gharbieh, Afshin et al. (2021).

Although a range of products has already been proposed, incorporating different features, they do not entirely fulfill their aim. Among the problems, of the solutions available in the market, are the lack of practicality and portability, invasive and requiring too much effort to learn Lozano, Kaczmarek and Santello (2009).

The difficulty of using or learn how to use a device could be avoided if concepts from Human Factors, or Ergonomics, were analysed during the product's development, using appropriate methods. The early application of these methods and tests could be a gamechanger for the success of the product's user experience Wolf, Binder, Miehl and Wartzack (2019).

Motivated by the dissatisfaction of blind people with the currently available products, this paper starts from the hypothesis that a human-factors-centred design of assistive

devices for blind and visually impaired people (BVIs) requires the involvement of BVIs in the design process in order to evaluate the product under design. The user has to test the product under development to provide feedback for the design team to improve the product.


In order to approach this problem, this work proposes using virtual reality (VR) as a tool for creating virtual environments, where proof of concepts or prototypes of assistive devices could be tested by BVIs. VR can be used to create specific, immersive and interactive situations that could help the user to learn and train Farrell (2018), and the developers to create more user-friendly products.

In a virtual environment, as long as the BVI is wearing a locating system, s/he can navigate the environment. Any information about the scenario, such as the position of objects and their distances to the user, is known and could be extracted from the virtual platform. As a consequence the designer can test different ways of translating this information into inputs before actually implementing a prototype of the assistive device, providing a flexible, safe and easy way to have it evaluated by different users.

The use of virtual reality for design purposes is not new. The cabin design process is often said to be complex because it involves several stakeholders, each with his/her own set of preferences and requirements. Moerland-Masic, Reimer, Bock, Meller and Nagel (2021) proposed to anticipate the involvement of the final users based on co-design. In their proposal, the users can influence the product's development from the beginning. However, for the involvement to happen, a communication channel needed to be established, and it was done using virtual reality. The use case showed some

*Corresponding author

**Principal corresponding author

 evillani@ita.br (E. Villani); ivan@ita.org (I.d.S. Rehder)

ORCID(s):

benefits and disadvantages of using virtual reality. The virtual reality helped to bring the client closer to the design team, allowing them to draw quick sketches in brainstorming gatherings. It was associated with a steep learning curve for the designers. Among the disadvantages, it was considered a high-cost tool, and its use for a long time was associated with nausea.

Motivated by the popularization of virtual reality technology, Siu, Sinclair, Kovacs, Ofek, Holz and Cutrell (2020) developed a white cane to be used by BVI users in a virtual environment. Their purpose was to make virtual reality applications available for BVI users. In order to evaluate their proposal, the authors performed an experiment where the participants had to play a “scavenger hunt” using an HTC Vive system. Among the relevant findings of Siu et al. (2020) is that not all the participants reacted the same to a particular stimulus. The vibration of the cane was considered confusing by some participants, while others were familiar with it. Another interesting observation was that, similar to what happens in the real world, it was easier for the participants to navigate in larger areas than in tight spaces. Moreover, the authors observed that the participants focused their attention on the primary task, without freely exploring the environment, which might have impacted the low time to achieve the goal and the low number of obstacle hits.

Kirner, Kirner, Wataya and Valente (2011) raised two questions, “How can blind people learn 3D concepts aiming to be able to convert explored 3D environments into pictures?” and “How can we develop a spatial audio tutor with augmented reality technology to make easy the understanding of 3D concepts by blind people?” and used not using virtual reality technology but augmented reality to answer them. They developed a augmented reality application to be a tutor for BVI users. The application used allowed BVI users to play audio streams that were associated with spatial positions. The users learned 3D concepts and also were able to perceive, understand and produce embossed pictures representing real and imaginary 3D scenes. Also they were able to understand descriptions of 3D scenes described by non-BVI people. The authors believe that this application can be evolved to explain other concepts such as colors, transparency, shades, etc.

Bradley and Dunlop published two works (2002; 2005) about how BVI navigates and how much it is similar or different to how a sighted person navigates. The first work of Bradley and Dunlop was published in 2002 and discussed which type of information BVI uses to navigate in an environment and how it compares to sighted people. The second they compared the perceived workload of BVI participants and sighted participants when they navigate using user-tailored information created with the results of the previous experiments Bradley and Dunlop (2005). The results showed that BVI users reached landmarks significantly quicker when given the information made for that group, but still longer than sighted users. Also it showed that BVI participants systematically have a higher workload than sighted participants and that BVI users did have a higher workload when guided

by orientations provided by sighted people, as well as the sighted participants did with orientations from BVI.

Mental workload is one of the main concepts studied in Human Factors Stanton, Hedge, Brookhuis, Salas and Hendrick (2004). The mental workload is similar to the physical workload but refers to the mental capacity necessary to perform a task. Each human being has a finite mental capacity. When the mental demand is higher than the operator’s capacity, the person needs to adapt to finish the task, or the overall performance of the task is compromised. Otherwise, if the mental workload is too low, the operator may get bored and easily distracted and could also fail or not process the task’s information. The mental workload is not a quantitative resource or something that one can directly measure, but several different techniques have been proposed in the literature to infer it and they can be: techniques based on task performance, techniques based on physiological measures and techniques based on subjective questionnaires.

The term “situation awareness” was first proposed for the Aeronautics domain and today is considered a key factor for designing complex and dynamic systems from other domains, such as automotive, medical and nuclear Endsley (1995). It can be defined as “the perception of the elements within a volume of time and space (Level 1), the comprehension of their meaning (Level 2), and the projection of their status in the near future (Level 3)” Sanders and McCormick (1998). It is an essential factor to make sure that the user will be capable to make important decisions correctly and achieve high-performance Endsley (1988, 2018). As it is for the mental workload, situation awareness is not a quantitative subject. The most common way to measure it is using subjective techniques, among which one of the most famous is the Situation Awareness Global Assessment Technique (SAGAT). It was proposed by Endsley (1988) and is based on how the information is processed inside the user’s mind.

Co-design, or collaborative design, refers to a design process in which individuals of the design team have different backgrounds or bring different experiences, which can be essential for the product under design. It is based on good communication and information sharing among the team Chiu (2002).

This paper’s main goal is the use of virtual reality as a tool for evaluating proofs of concept of assistive devices for blind and visually impaired people from a human-factors perspective. The purpose is to provide a flexible and easily configured way of testing different concepts of assistive devices in order to support an agile and user-centered development.

This goal is related to the following research questions, which are investigated in this work:

- Is it possible to evaluate and compare concepts of assistive devices from a human factors perspective in a virtual environment? What are the main limitations of the use of a virtual reality environment?
- Do non-BVI users, when deprived of their vision, similarly evaluate assistive devices as BVI users?

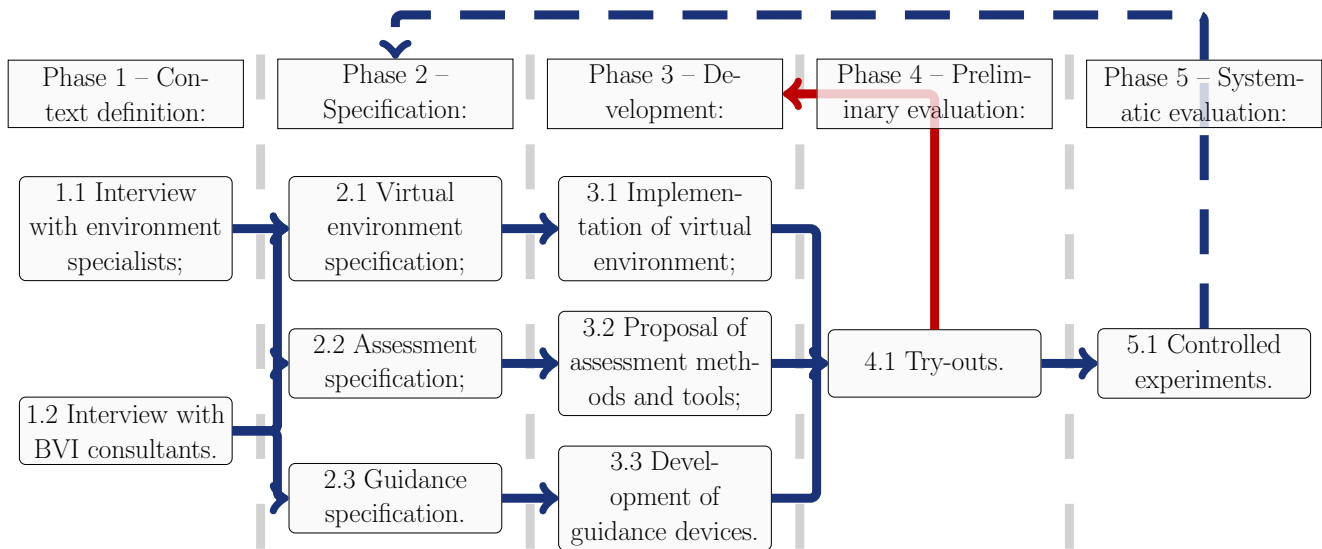


Figure 1: Method's diagram

The concepts of assistive devices presented as part of this work are used only as examples for investigating the research questions presented. The challenges related to their full development up to high Technology Readiness Levels (TRLs), as well as their feasibility as commercial products, are out of the scope of this work.

Structure of the text

The next Section of this paper are organized as follows.

Section 2 details the proposal of this paper describing how virtual reality could be used to integrate BVI users into the design process of assistive design. It illustrates the proposed method by applying it to evaluate three different assistive devices (audio guide, virtual cane and haptic belt), as well as their mixed-use, in the environment of a hospital reception.

Section 3 describes the experiment designed to evaluate the paper's proposal and analyses the results in order to investigate the research questions of Section ??

Finally, section 5 summarizes the main conclusions of this work and discusses future work.

2. Material and methods

This chapter describes the method proposed in this work for evaluating assistive devices using virtual reality. The method is organized into 5 phases as illustrated in Figure 1.

The first phase is the context definition. It consists of defining the main features of the environment in which the assistive device will be used, based on interviews from hospital's specialists from São José dos Campos. It also includes a step for understanding the limitations of the current assistive devices and defining the main features of assistive devices to be designed. This last step is based on

interviews with two BVI users, one that is blind since 13 years old and another that has Usher's disease.

In the second phase, the information collected through the interviews of Phase 1 is used to make critical decisions about the virtual environment where the evaluation of the assistive devices will be carried out. It is also used to define which human factors should be assessed. Finally, it contributes to define the guidance devices that should be implemented in the assistive devices.

The third phase is dedicated to developing the virtual environment, the evaluation tools and techniques, and the first proof of concept of the assistive devices, which should be integrated into the virtual environment for testing.

Virtual Environment

The virtual reality environment was developed using the Unity3D platform and the implementation should be followed by preparing the corresponding physical space to perform the test campaign. Following the recommendations from the BVI interviews, typical reception furniture was placed in the virtual environment as illustrated in Figure 2. Sounds were also used to increase the feeling of immersion and indicate to the BVI participant where the reception desk was located. The participant navigation through the reception was composed of 4 tasks, also illustrated in Figure 2:

1. Clean the hands at the sanitizer totem (COVID-19 procedures);
2. Go to the reception desk to receive a queue number;
3. Go to the waiting area and wait for the number calling;
4. Leave the room when called.

Figure ?? shows the virtual environment created in Unity3D and the corresponding real environment assembled at the CCM entrance hall.

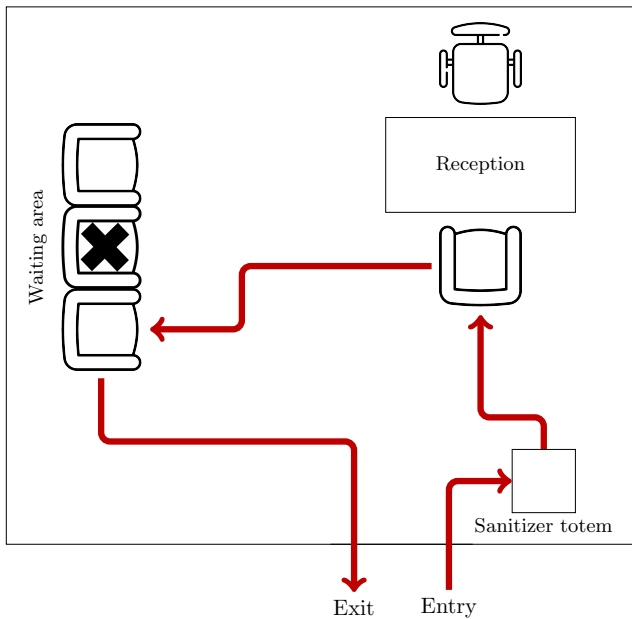
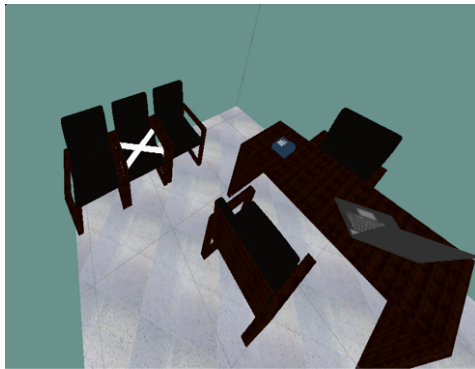


Figure 2: Scheduled task of the experiment and their order.



(a): Virtual environment screenshot



(b)

Figure 3: Environment comparison

Human Factors techniques

The assessment should evaluate workload, situation awareness and the user's impressions using:

1. Workload;

Following the recommendations from the review of literature, the workload is estimated using two approaches:

- Physiological measures obtained from an ECG (Electrocardiogram) sensor and a GSR (Galvanic Skin Response) sensor;
- NASA-TLX (National Aeronautics and Space Administration Task Load Index) subjective questionnaire.

2. Situation awareness;

A modified SAGAT (Situation Awareness Global Assessment Technique) questionnaire is used to evaluate the BVI situation awareness. As original idea from Endsley (1988), the proposed version is based on 3 levels of situation awareness:

Level 1. Perception

It aims to evaluate if the user can perceive the environment surrounding him/her.

Level 2. Comprehension

After the user answer about an detected object, he/she is asked to point to where the object is located.

Level 3. Projection

This level is measured after every question that asks the location of an object. He/she is then required to answer how far he/she supposes that this object is

3. Devices evaluation;

Finally, a questionnaire is proposed for evaluating the guidance devices. The questions were about the comfort, the sense of safety, the sense of confusion and on the precision that the manipulation of the device caused.

Assistive Devices

Four guidance methods were proposed to be evaluated in this work.

1. Audio guidance;

The first method is audio guidance. Basically, in the course of the experiment, the participant could give two different voice commands: "What is around me?" and "Where is (something)?" The answers of both commands was done with the interference of a member of the design team.

2. Vibration guidance with command – virtual cane;

When using a white cane, the user points it to check nearby obstacles in a specific direction. The virtual cane has a similar way of functioning, but instead of connecting the user to the nearby object through the cane, it vibrates when it detects an obstacle in the direction the user pointed it.

3. Vibration guidance without command – haptic belt

The belt has appended 8 vibration units that vibrate accordingly to the direction and distance of the closest object around the user. The main differences between the

virtual cane and the haptic belt is that the haptic belt checks 360° around the user. When objects are within a certain limit, it vibrates indicating to the user the direction of the closest object.

4. Mixture of audio and vibration guidance

This option is implemented making the three options available to the user: audio guidance, haptic belt and virtual cane.

The fourth phase provides a preliminary assessment of the devices through its unstructured experimentation by BVI consults. This preliminary assessment provides feedback for improving the device concept. The cycle of “try-out and improve device concept” can be repeated until the device concept is considered mature to be tested through a systematic set of controlled experiments.

The fifth phase consists of executing a campaign of controlled experiments, following the best practices of the DoE (Design of Experiments) discipline, and analysing the results. Concluded this phase, the results should provide information for the design team to decide between proceeding to the detailed design of the assistive devices or performing a new evaluation cycle.

In the case of this work, the proposed experiment consists of asking the participant to use five guidance methods: the audio guidance, haptic belt, virtual cane, mixed and, additionally, the device used daily by the BVI (e.g., white cane). Moreover, each participant should use each guidance method twice (“first visit” and “return visit”), in order to provide some information about how the guidance devices performs in new and known environments. In order to avoid the learning effect from one method to the other, five versions of the reception scene are developed, changing the position of the objects - one version to be used with each guidance method. The scene order is randomized for each participant.

Moreover, particularly in the case of this work, an additional round of tests is added to the experiment to investigate the differences between the evaluation performed by BVI users and sighted (non-BVI) users. For this purpose, the same experiment is repeated with a set of non-BVI users. The purpose is to investigate whether or not performing the analysis with non-BVI users could lead to different conclusions.

The experiment is organized in the following way:

- Briefing:
- Execution of the experiment:
 - Guidance method training;
 - First visit and SAGAT questionnaire;
 - NASA-TLX for the first visit;
 - Return visit and SAGAT questionnaire;
 - NASA-TLX for the return visit;
 - Questionnaire about the guidance method.
- Experiment’s conclusion

As a result of the experiment, the following data are collected:

- Answers to the NASA-TLX questionnaire;
- ECG and GSR signals;
- Answers to the SAGAT questionnaire;
- Answers to the guidance method questionnaire.

3. Results and discussion

The described experiment was performed with the following groups and has an approval of the brazilian ethics committee.:

- Blind group: composed of 4 participants with ages varying from 26 to 56, all male, three of them graduated and one with ongoing graduation.
- Sighted group: composed of 4 participants with ages varying from 22 to 31, three males and one woman, all graduated.

The data analysis follows the following sequence:

- Analysis of subjective questionnaires:
 - NASA-TLX:
 - Adapted SAGAT:
 - Guidance method’s questionnaire:
- Analysis of physiological sensors:
 - ECG (Electrocardiogram):

Two features are extracted from the ECG signal, heart rate (BPM) and heart rate variance (SDNN).
 - GSR (Galvanic Skin Response):

3.1. Evaluation of assistive device from a human factors’ perspective in a virtual environment

3.1.1. NASA-TLX

The NASA-TLX provides two relevant pieces of information to the workload analysis. The first is the score attributed to the “mental demand” dimension and the second is the average obtained from NASA-TLX’s six dimensions. The two analyses are presented in the next subsections.

Analysis of the mental demand scale

Figure 4 presents a boxplot of the mental demand score grouped by the methods. This figure shows that there may be two groups: one associated with lower demand, composed of base and audio, and another with higher demand, composed of haptic belt, virtual cane and mixture. It indicates that maybe a guidance method that uses vibration as input is not intuitive. Figure 5 presents a boxplot of the mental demand

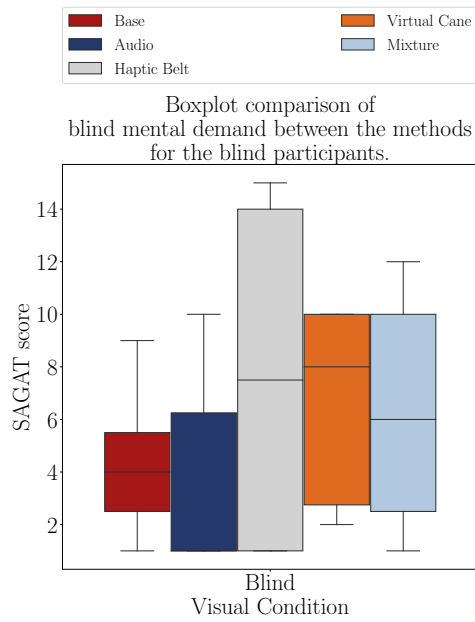


Figure 4: Boxplot of the mental demand of the blind participants grouped by the methods.

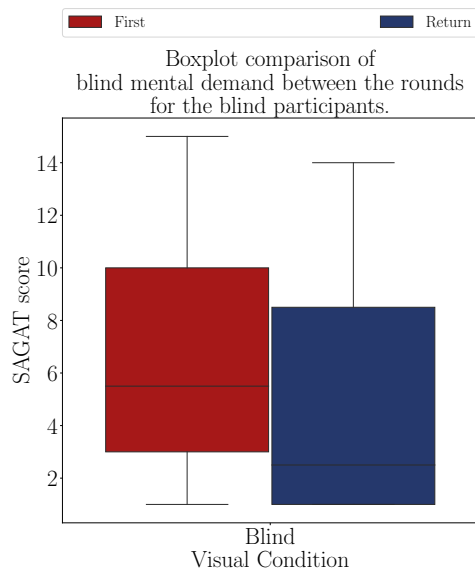


Figure 5: Boxplot of the mental demand of the blind participants grouped by the rounds.

grouped by the rounds, confirming the general tendency to reduce the required "mental demand".

The results of ANOVA are presented in Table 1. A p-value of 0.05 is commonly adopted as a threshold to confirm the hypothesis. According to this criterion, neither method or round have a significant influence on the mental demand.

Analysis of the NASA-TLX score

Figure 6 presents the boxplot with the NASA-TLX global score grouped by the methods. Similar to what happened for the "mental demand", it is possible to split the

Table 1

ANOVA p-value for mental demand – blind participants

Source	P-Value
Methods	0.170
Rounds	0.075
Interaction	0.993

Table 2

Anova p-value for the NASA-TLX score – blinded users.

Source	P-Value
Methods	0.029**
Rounds	0.022**
Interaction	0.814

methods into two different groups: base and audio, which require a lower level of workload, and another group, which requires a higher level. Figure 7 presents a boxplot with the NASA-TLX global score grouped by the rounds, showing that the two groups are still different.

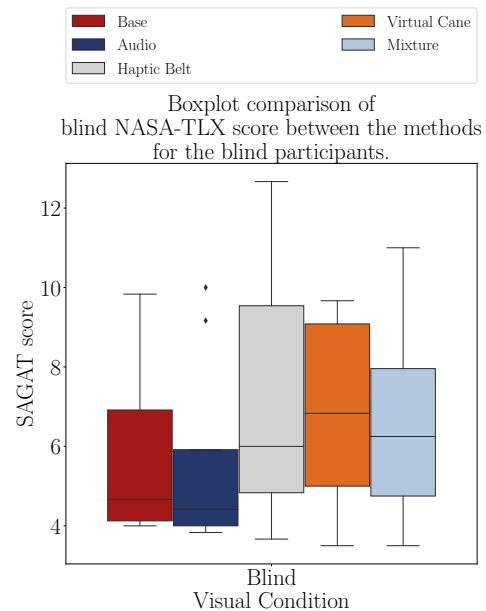


Figure 6: Boxplot of the mental demand of the blind participants grouped by the methods.

The sample residuals are not homogeneous meaning that the participants have different variability among them and that impacts the ANOVA.

Table 2 brings the p-value resulting from ANOVA. In this case, both the methods and the rounds were appointed as significant variables that influence the mean value of the NASA-TLX global score.

Finally a pairwise Fisher LSD test comparing each pair of guidance methods. The results show that only audio is similar to the base. All the other methods are different from each other.

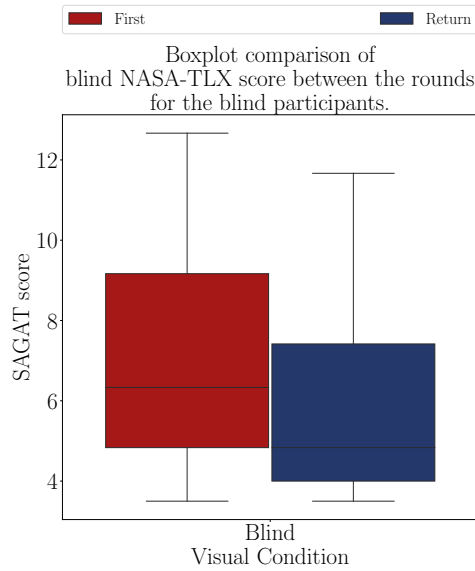


Figure 7: Boxplot of the mental demand of the blind participants grouped by the rounds.

3.1.2. Adapted SAGAT

For each question of the SAGAT questionnaire, the participant could score 1 point or a fraction of it. The closer to the value 1, higher is the situation awareness of the user.

Figure 8 brings the boxplot of the SAGAT score grouped by the guidance methods. It shows that the methods can be divided into two groups. The first one is composed of base, haptic belt and the mixture. This group received scores higher than the second group, composed of audio and virtual cane. Figure 9 shows the boxplot of the data grouped by round and confirms the general improvement of situation awareness from the first to the return round.

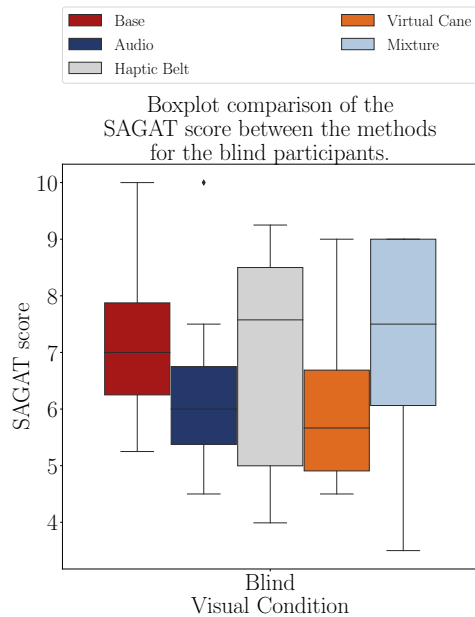


Figure 8: Boxplot of the SAGAT score of the blind participants grouped by the methods.

Table 3

Anova p-value for the SAGAT score – blinded users.

Source	P-Value
Methods	0.277
Rounds	0.002**
Interaction	0.834

Table 4

Anova p-value for the questionnaire score – blinded users.

Source	P-Value
Method	0.001**

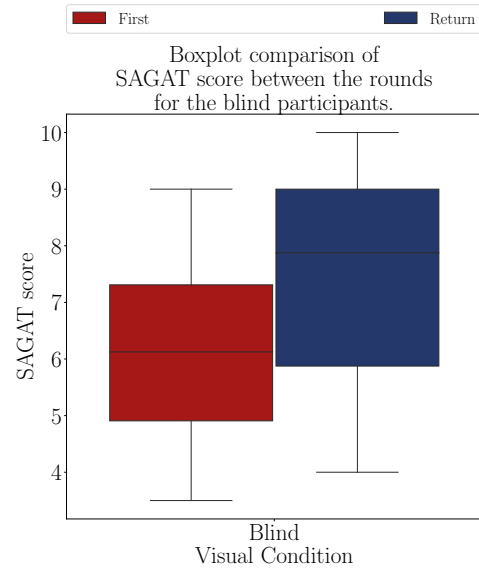


Figure 9: Boxplot of the SAGAT score of the blind participants grouped by the rounds.

Table 3 shows the ANOVA test p-value of the SAGAT score. It indicates that the round is a significant variable that influences the value of the SAGAT score. The same cannot be said for the method, which has no significant influence.

3.1.3. Guidance method's questionnaire.

The data from the questionnaire for evaluating the user experience with each guidance method is also analysed. The higher the score, the more satisfied the user is with the method. It is essential to observe that this analysis does not include the base method as the questions are specific about each method and the base may vary among the participants. Also, there is no distinction between first and return rounds. Each questionnaire is answered only once for each method. Figure 10 brings the questionnaire boxplot, which clearly shows the difference between two groups: haptic belt and virtual cane, and audio and mixture.

The results of ANOVA are presented in Table 4 and it shows that the method, with a p-value of 0.001, is indeed a significant variable that affects the user's satisfaction.

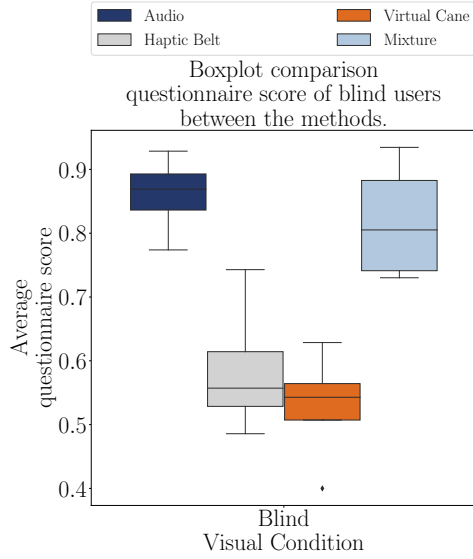


Figure 10: Boxplot of the questionnaire score of the blind participants grouped by the methods.

In order to complement the ANOVA analysis, the pairwise comparison of the methods obtained from the Fisher LSD shows that audio and mixture are equivalent from the perspective of user satisfaction. All the other comparisons indicate there is a difference between the methods.

Additional to the scores, the participants also expressed their dissatisfaction with the answers to the open questions of the questionnaire, where they commented that the haptic belt and the virtual cane are confusing, are not precise enough, and are very different from what they are used to.

3.1.4. Electrocardiogram (ECG) data

After the experiment, the ECG signal processing is organized in the following steps:

- Filtering and removing outliers. Since the participants moved during the whole experience, the sensors also captured some noise data.
- Normalization between -1 and 1;
- Peak detection and evaluation – if the results were not of good quality, the peak detection method's parameters were adjusted to improve it;
- Calculation of BPM using Kubius HRV Standard;
- Calculation of SDNN using Kubius HRV Standard.

At the beginning of each experiment, a baseline was collected to establish a comparison between the relaxed state of the participant and the scenes' induced state. However, the results were not consistent. During the experiment, it was expected that the heart rate would increase compared to the baseline because the participants were at rest. However, for most of the participants, it decreased, indicating a systematic problem may have occurred. Due to this fact, the analysis is based only on absolute values.

Analysis of the heartbeat frequency (BPM)

Figures 11 and 12 brings the corresponding boxplot, grouped by method and round. In both cases, it is not possible to observe significant differences among the methods or rounds.

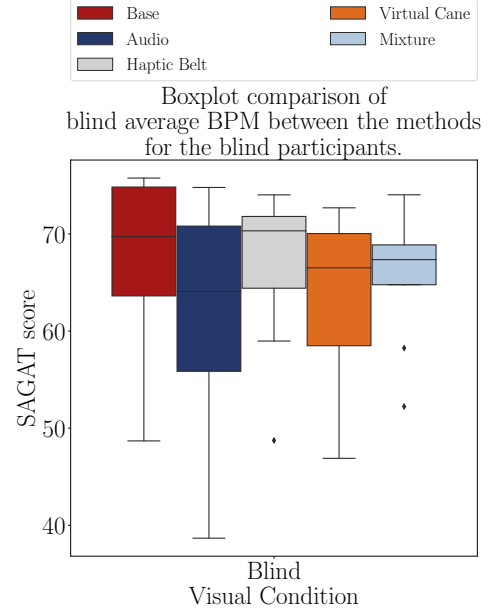


Figure 11: Boxplot of the BPM of the blind participants grouped by the methods.

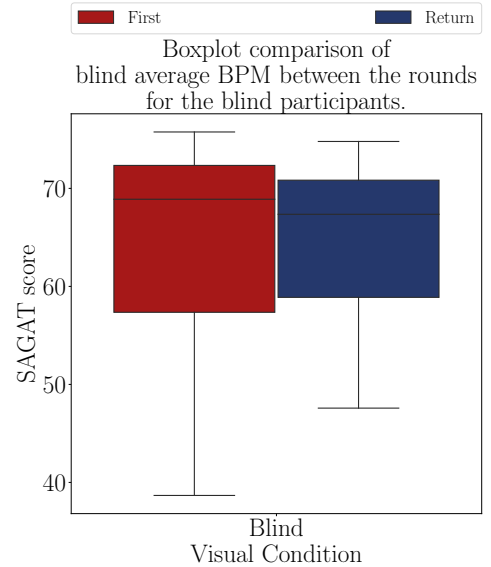


Figure 12: Boxplot of the BPM of the blind participants grouped by the rounds.

The participants do not have a similar variance, which jeopardize the results of ANOVA. Considering this limitation, Table 5 brings the p-value obtained by ANOVA, which confirmed the previous analysis, as it does not indicate a

Table 5

Anova p-value for the BPM – blinded users.

Source	P-Value
Methods	0.100
Rounds	0.371
Interaction	0.894

Table 6

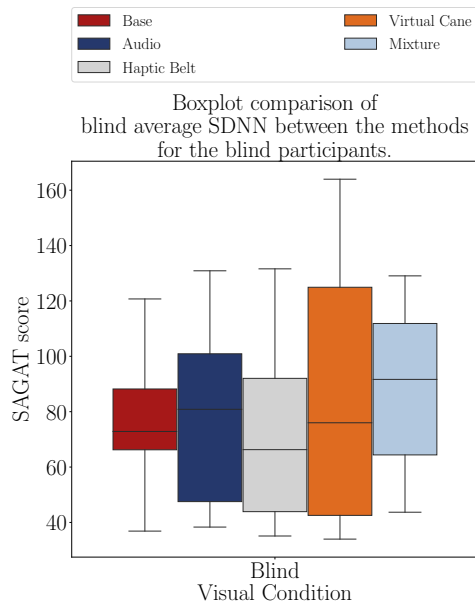
Anova p-value for the average SDNN – blinded users.

Source	P-Value
Methods	0.486
Rounds	0.223
Interaction	0.473

significant influence of either the guidance methods or the rounds in the participants' heart rate.

Analysis of the heartbeat variance (SDNN)

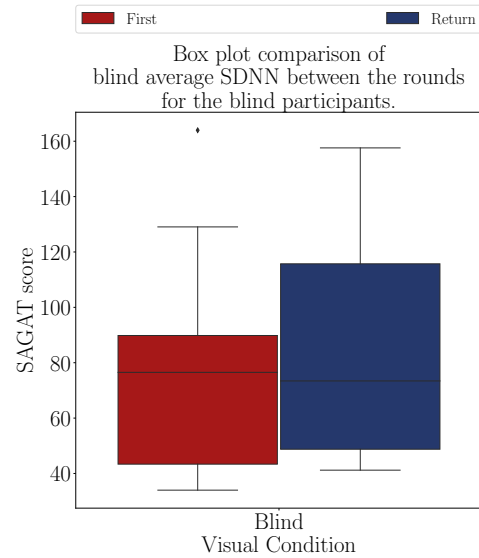
Figure 13 and Figure 14 bring the SDNN barplot grouped by the methods and the rounds. There is a slight tendency among the participants to increase the heartbeat in the return round.

**Figure 13:** Boxplot of the SDNN of the blind participants grouped by the methods.

The ANOVA results are presented in Table 6 and do not confirm any influence of the methods nor the rounds on the ECG heart rate variance.

3.1.5. Galvanic skin response and temperature data;

The GSR analysis is based on the signal's average level. Each experiment's round is compared to the participant baseline collected before the experiment. The GSR sensor

**Figure 14:** Boxplot of the SDNN of the blind participants grouped by the rounds.

was worn on the left hand for right-handed participant and on the right hand for left-handed participants. One of the blind participants had the GSR sensor removed during the experiment because it was not appropriately fixed.

Figure 15 presents the boxplot of the percentual variation in the skin conductance for each method. The base method has the lowest variation among all methods. Also, the introduction of vibration increases the method variance. Figure 16 presents the GSR grouped by the rounds. In this case, there is no apparent difference between the rounds.

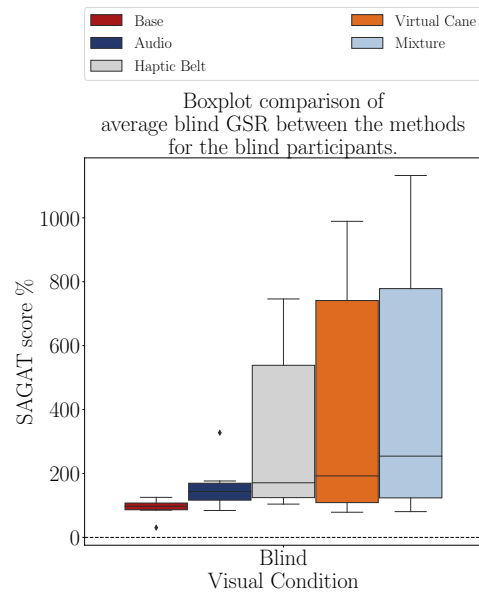
**Figure 15:** Boxplot of the GSR of the blind participants grouped by the methods.

Table 7 shows the ANOVA test p-value for the GSR percentual variance. Although the p-value for the method is

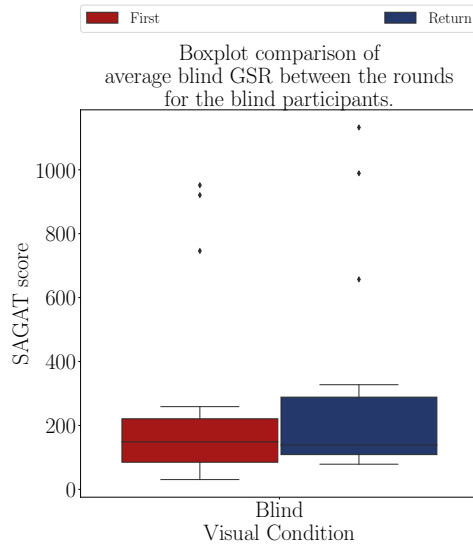


Figure 16: Boxplot of the GSR of the blind participants grouped by the rounds.

Table 7

Anova p-value for the GSR – blinded users.

Source	P-Value
Methods	0.051
Rounds	0.722
Interaction	0.996

not below the threshold of 0.05, it is close to it, indicating that probably the GSR is affected by it.

3.2. Comparison between BVI users and sighted users

This section investigates the second research question of this work: “do non-BVI users, when deprived of their vision, similarly evaluate assistive devices as BVI users?”.

To do so, the analysis performed in the previous section is now repeated with the data obtained from sighted participants. However, the data corresponding to the “base” method is omitted, as the daily method used by sighted people is based on their vision.

3.2.1. NASA-TLX

Analysis of the mental demand scale

Figures 17 and 18 presents the box plot for both groups, organized by the methods and the rounds. The mental demand is systematically higher for sighted people, which is expected. However, while blind participants considered the audio method less demanding, sighted participants preferred to the virtual cane. For both groups, we observe a decrease in the mental demand.

Table 8 brings the results of ANOVA. Unlike the blind participants, in the case of sighted ones, the p-value for the methods is below the threshold of 0.05, confirming it as a significant variable for the mental demand. In the case of

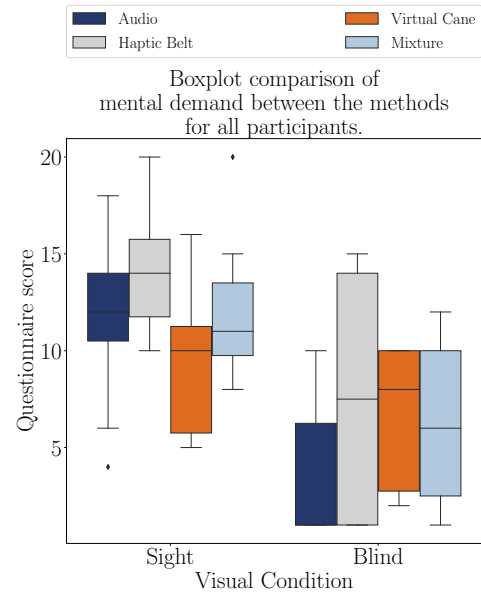


Figure 17: Boxplot of the mental demand of the participants grouped by the methods.

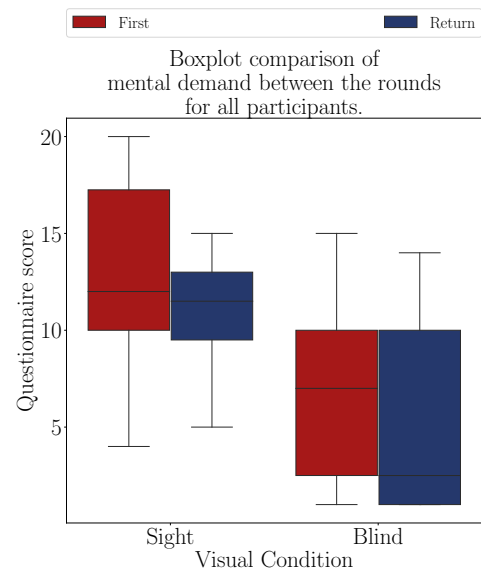


Figure 18: Boxplot of the mental demand of the participants grouped by the rounds.

the rounds, the data from both sighted and blind participants resulted in the exact p-value of 0.075, which is close to the traditional threshold of 0.05 but slightly higher.

Analysis of the NASA-TLX score

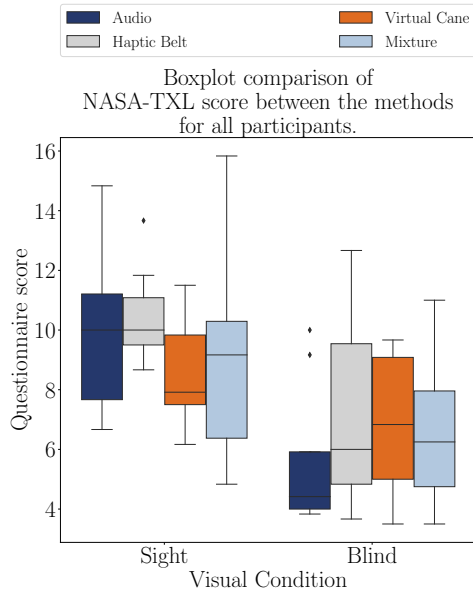
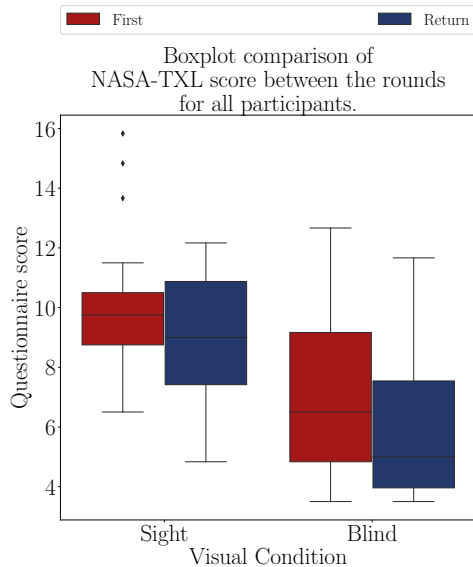
Figures 19 and 20 present the boxplots of the NASA-TLX global score. Again, it is possible to see that sighted people usually give higher workload scores than blind ones. The influence of the round is approximately the same. However, the order of preference of the methods is different.

The p-values for both groups are presented in Table 9. It confirms the influence of the round for both sighted and

Table 8

Anova p-value for the mental demand average on each method'

(a): Blind participants		(b): Sight participants	
Source	P-Value	Source	P-Value
Methods	0.170	Methods	0.049**
Rounds	0.075	Rounds	0.075
Interaction	0.993	Interaction	0.990

**Figure 19:** Boxplot of the NASA-TLX score of the participants grouped by the methods.**Figure 20:** Boxplot of the NASA-TLX score of the participants grouped by the rounds.

blind people. In the case of the methods, the p-value of blind is lower than the threshold of 0.5, while that of sighted is slightly higher.

Table 9

Anova p-value for the NASA-TLX score on each method

(a): Blind participants		(b): Sight participants	
Source	P-Value	Source	P-Value
Methods	0.029**	Methods	0.086
Rounds	0.022**	Rounds	0.034**
Interaction	0.814	Interaction	0.688

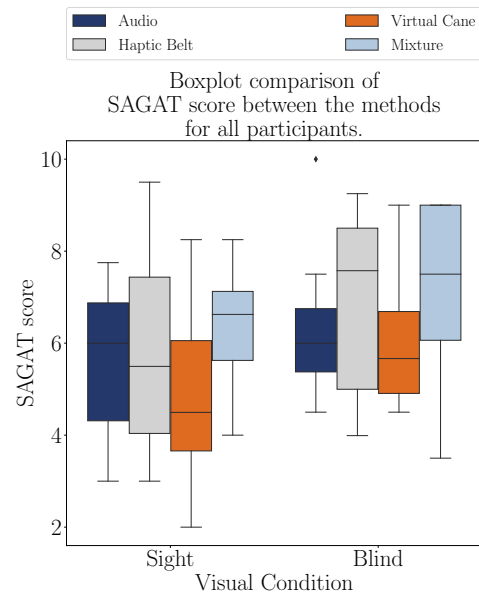
Table 10

Anova p-value for the SAGAT score on each method

(a): Blind participants		(b): Sight participants	
Source	P-Value	Source	P-Value
Methods	0.277	Methods	0.035**
Rounds	0.002**	Rounds	0.095
Interaction	0.834	Interaction	0.578

3.2.2. Adapted SAGAT

Figures 21 and 22 bring the boxplots. According to Figure 21, both groups presented a higher situation awareness with 'mixture' and 'haptic'. On the other hand, Figure 22 confirms that the difference between the rounds is more significant for blind participants.

**Figure 21:** Boxplot of the Sagat score of the participants grouped by the methods.

The variance of the residuals is not equal among the participants. Table 10 brings the p-value from ANOVA. While for the blind participants, the rounds are a significant factor and the methods are not, for the sighted participants the result is the opposite, showing a significant influence of the methods and not of the rounds.

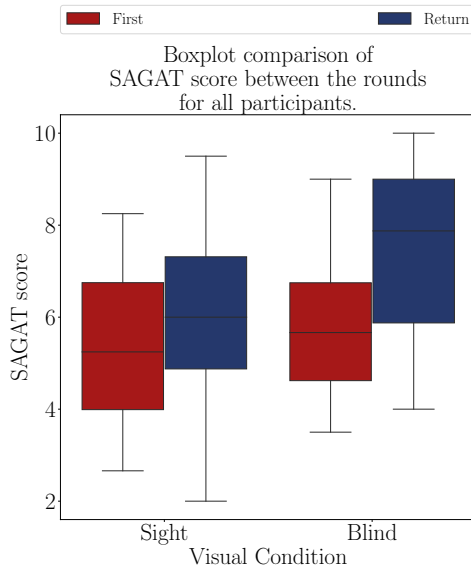


Figure 22: Boxplot of the Sagat score of the participants grouped by the rounds.

3.2.3. Guidance method's questionnaire.

The Figure 23 presents the box plot with the distribution of the scores. It is possible to see that there is some similarity between the two groups, except for the virtual cane method, which has a broader distribution for the sighted users. Also, it seems that the audio and mixture have similar acceptance for sighted and blind users.

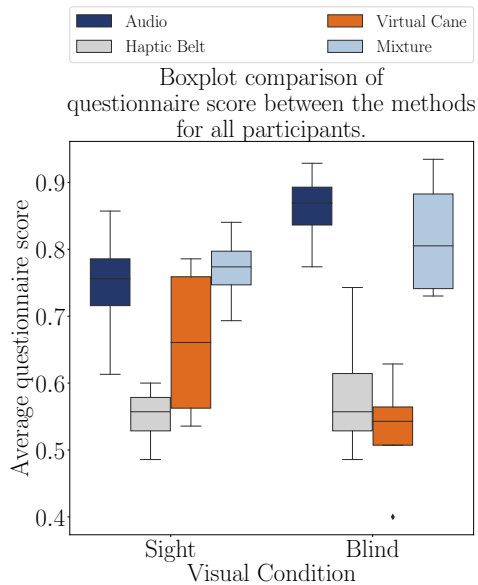


Figure 23: Boxplot of the questionnaire score of the the participants grouped by the methods.

The result of ANOVA is presented in Table 11 and indicates that the method is an effective variable for the sighted participants, as it is for the blind ones.

Table 11

Anova p-value for the questionnaire score on each method

(a): Blind participants.		(b): Sighted participants.	
Source	P-Value	Source	P-Value
Method	0.001**	Method	0.016**

Table 12 presents the conclusion of A pairwise Fisher LSD test between all the guidance methods for both groups shows that the results are coincident between the two groups.

3.2.4. Electrocardiogram (ECG) data

Analysis of the heartbeat frequency (BPM)

If the variation between the First and the Return round is positive, it means that the user had an increase on his/her mental workload and vice-versa. Comparing the two groups, the audio method is associated with a slightly lower heart rate for blind people, but the opposite happens for sighted participants. Moreover, data from blind participants have a significant variance. This significant variance can also be observed in the boxplot of Figures 24 and 25.

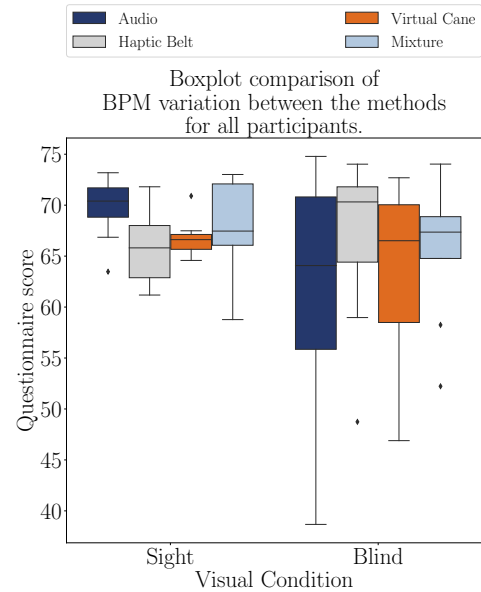


Figure 24: Boxplot of the average BPM of the participants grouped by the methods.

Table 13 brings the results from ANOVA, which are similar for both sighted and blind participants.

Analysis of the heartbeat variance (SDNN)

Figures 26 and 27 shows the boxplots for both groups. Both pictures show that the SDNN of the sighted users was higher than that of the blind users, indicating that sighted users had a lower mental workload than the blind users.

Table 12

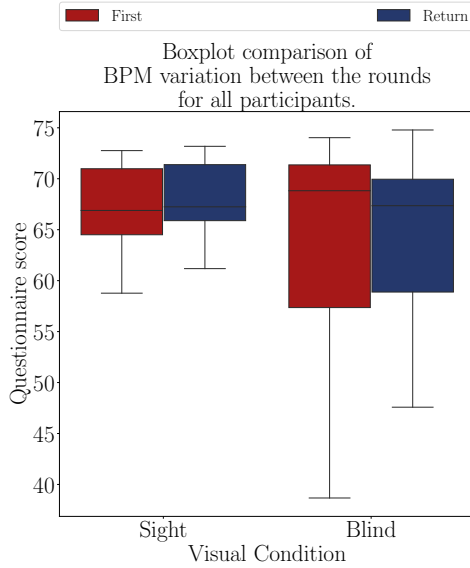
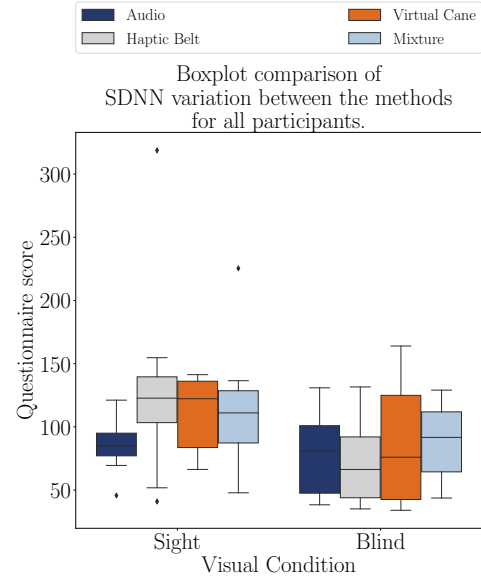
Anova p-value for the mental demand average on each method'

(a): Blind participants.

Method			Analysis	
Audio	X	Haptic Belt	$H_1 : \mu_{Audio} \neq \mu_{Haptic Belt}$	**
Audio	X	Virtual Cane	$H_1 : \mu_{Audio} \neq \mu_{VirtualCane}$	**
Audio	X	Mixture	$H_0 : \mu_{Audio} = \mu_{Mixture}$	
Haptic Belt	X	Virtual Cane	$H_1 : \mu_{Haptic Belt} \neq \mu_{VirtualCane}$	**
Haptic Belt	X	Mixture	$H_1 : \mu_{Haptic Belt} \neq \mu_{Mixture}$	**
Virtual Cane	X	Mixture	$H_1 : \mu_{VirtualCane} \neq \mu_{Mixture}$	**

(b): Sight participants.

Method			Analysis	
Audio	X	Haptic Belt	$H_1 : \mu_{Audio} \neq \mu_{Haptic Belt}$	**
Audio	X	Virtual Cane	$H_1 : \mu_{Audio} \neq \mu_{VirtualCane}$	**
Audio	X	Mixture	$H_0 : \mu_{Audio} = \mu_{Mixture}$	
Haptic Belt	X	Virtual Cane	$H_1 : \mu_{Haptic Belt} \neq \mu_{VirtualCane}$	**
Haptic Belt	X	Mixture	$H_1 : \mu_{Haptic Belt} \neq \mu_{Mixture}$	**
Virtual Cane	X	Mixture	$H_1 : \mu_{VirtualCane} \neq \mu_{Mixture}$	**

**Figure 25:** Boxplot of the average BPM of the participants grouped by the rounds.**Figure 26:** Boxplot of the average SDNN of the participants grouped by the methods.**Table 13**

Anova p-value for the BPM on each method.

(a): Blind participants		(b): Sight participants	
Source	P-Value	Source	P-Value
Methods	0.100	Methods	0.166
Rounds	0.371	Rounds	0.308
Interaction	0.894	Interaction	0.631

Table 14 shows the ANOVA test p-values. For both groups, none of the factors have a significant influence on the SDNN value.

3.2.5. Galvanic skin response and temperature data;

If the variation between the round and the Baseline is positive, it means that the user had an increase on his/her Mental Workload or stress. While the GSR varied for the blind participants, increasing for methods with vibration, the same does not happen for sighted participants. Also, the variance of GSR data for blind participants is significantly higher than that of sighted ones. The same conclusion can be drawn from the boxplots in Figures 26 and 27.

The results from ANOVA are presented in Table 15. In the case of blind participants, the p-value for the method is just slightly over the threshold, indicating a possible influence of the method. The same does not happen with

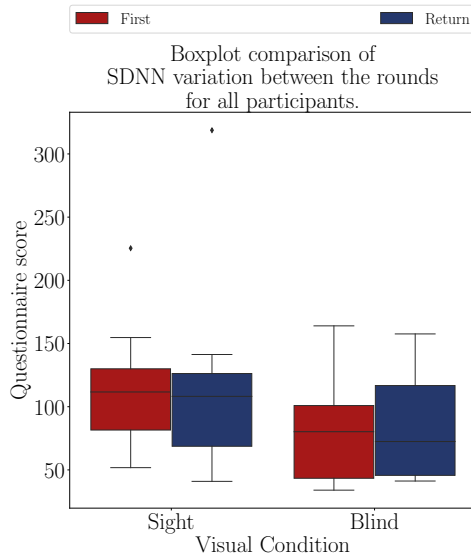


Figure 27: Boxplot of the average SDNN of the participants grouped by the rounds.

Table 14

Anova p-value for the average SDNN on each method.'

(a): Blind participants		(b): Sight participants	
Source	P-Value	Source	P-Value
Methods	0.486	Methods	0.189
Rounds	0.223	Rounds	0.969
Interaction	0.473	Interaction	0.455

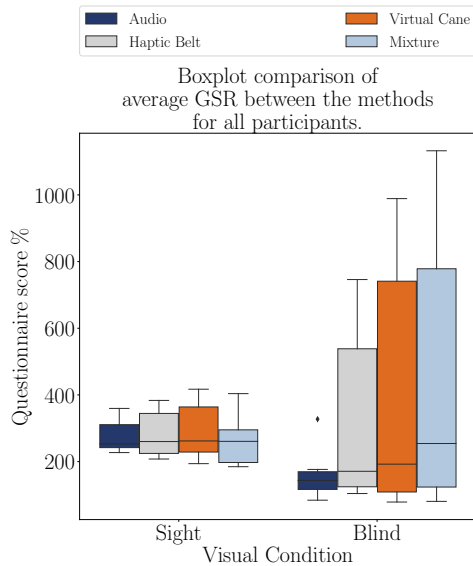


Figure 28: Boxplot of the average GSR of the participants grouped by method.

sighted participants, where the p-value of the method factor is the highest and well above the 0.05 threshold.

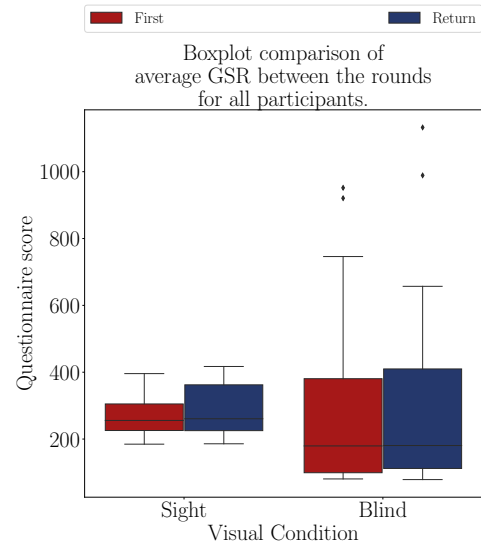


Figure 29: Boxplot of the average GSR of the participants grouped by round.

Table 15

Anova p-value for the skin conductance average on each method

(a): Blind participants		(b): Sight participants	
Source	P-Value	Source	P-Value
Methods	0.051	Methods	0.802
Rounds	0.722	Rounds	0.354
Interaction	0.996	Interaction	0.686

4. Discussion

Final Remarks

To summarize the conclusion obtained from the analysis of the data from blind participants, the audio method showed a lower score both for NASA-TLX mental demand and NASA-TLX global score. In contrast, the methods that include vibration achieved higher scores. This probably happened because the participants are already used to using sound to guide themselves, especially environmental sounds. The environment sounds used in the scenes were always the same (telephone ringing, laptop keyboard sounds, exterior noise, door opening and closing). The participants likely felt more relaxed when they only had to focus on the sounds around him/her. This is reinforced by the fact that, during the experiment with the audio method, half of the participants did not ask for any information, or the audio command option was used only a few times.

The fact that the haptic devices caused a higher workload is probably due to the fact that the users had to learn and get used to them. Besides, for being just conceptual, their precision was not as good as they were expecting. That explains why their results were not as good as the base or audio methods. The NASA-TLX results are correctly related to the satisfaction questionnaires, which scored them as the unsatisfied devices.

As expected, most of the variables from subjective questionnaires (NASA-TLX and SAGAT) show some influence of the rounds. On the other hand, the results from the physiological sensors did not show a clear tendency.

The statistical analysis based on ANOVA tests confirmed some of the observations from the bar and box plots. However, in many cases, the residual distributions were not homogenous and the statistical analysis was affected by the small number of samples.

All the blind participants showed great enthusiasm before, during and after the experiment. They also made several recommendations for both the virtual environment and the devices, such as:

- The speakers of the HMD are not good enough to give them the precise location of the sound origin
- The HMD is too large and covers half of the participant's face. It gives them a strange sensation, since some of them use the air or the wind feeling on the face to give them hints about the location of walls or other high obstacles;
- The precision of the vibration for both the haptic belt and the virtual cane needs to be improved. It is not enough for them to use the devices. This problem is related to how the HMD sets the position of the user in the virtual environment.
- The vibration from the haptic belt was not intense enough.

Final Remarks 2

The comparison between the results from the blind participants and the sighted participants showed that there are significant differences in the evaluation performed by each group.

The sighted users evaluated the mental demand and other dimensions of NASA-TLX higher than blind ones. Also, blind participants were more familiar with audio methods and therefore gave a lower score to its mental demand. In the case of sighted participants, the method that received the lowest score was the virtual cane.

The adapted SAGAT questionnaire showed a more significant influence of the round factor for blind participants, which significantly improved their situation awareness on the return round. In the case of sighted users, the difference between the rounds was not so striking. Also, the score achieved by sighted participants was lower than that of blind users, which was expected.

Another difference is that, for blind participants, it was possible to observe a difference between the methods that use vibration and those that do not. This difference was not clear for sighted participants.

Besides these results, the sighted participants also gave feedback about the experiment. They felt considerably insecure when walking, even when hand-guided by another person. On the other hand, blind participants were already

used to bumping their bodies when exploring new spaces. The sighted participants did not want that to happen and approached the furniture with caution. Similar to the blind participants, they also noticed the lack of precision of the haptic devices, but they did rely on them to navigate.

5. Conclusion

This work proposes using virtual reality to create an environment where concepts of assistive devices could be evaluated at the early stages of development by blind and visual impaired (BVI) people.

In order to systematize this proposal, this work presents a method composed of five phases that guide the development of the virtual environment in parallel with the design of the assistive devices and the proposal of assessment methods.

In order to illustrate the proposed method and investigate two research questions related to this work, it is described an application of the method for the evaluation of four different solutions of assistive devices in the environment of a hospital reception. For this example, it proposes as an assessment method the use of subjective questionnaires and physiological sensors.

In order to evaluate situation awareness, this work proposes an adapted version of the SAGAT (Situation Awareness Global Assessment Technique) questionnaire, which was initially introduced for evaluating the situation awareness of air traffic controllers.

Based on the results from the hospital reception example, the two research questions are discussed.

Is it possible to evaluate and compare concepts of assistive devices from a human factors' perspective in a virtual environment? What are the main limitations of the use of a virtual reality environment?

The example presented in this work showed that it is possible to evaluate both situation awareness and workload using experiments performed in a virtual environment. The tests performed in the virtual environment made possible the comparison of the assistive devices both qualitatively and quantitatively.

However, some limitations were identified during the development of this work, regarding both the virtual environment and the assessment techniques.

One of the most recurrent observations was the unsatisfactory quality of the sound system. According to blind participants, the headphone of the VIVE HMD does not provide sounds with a quality good enough for them to locate the source of a sound. A regular comment was, "I feel like the sound origin is inside my head". This limitation may be solved by placing a sound source in the real environment and use the HMD only for localizing the participant in the virtual environment.

Another limitation is the actual position of the furniture. After a first round, the furniture was not precisely aligned

with its virtual model. A future solution for this problem would be to use a locator on each piece of furniture.

Among the main limitations identified during this work, it is worth mentioning the failure to detect collisions in the virtual environment, which could be solved by integrating sensors that monitor the position of each arm and leg of the user.

Regarding the assessment methods for evaluating human factors, the physiological sensors did not show any systematic difference among the methods under analysis. This result may be due the noise in the sensors' data, compromising its quality. Another problem is that the low number of participants may compromise the statistical analysis, due to the large variability among users.

Do non-BVI users, when deprived of their vision, similarly evaluate assistive devices as BVI users?

Comparing the results of the experiments performed with blind and sighted participants, some differences were observed. Among the most important, is the relative evaluation of audio and haptic devices. Due to their enhanced sensitivity to sounds, BVI users tend to evaluate audio solutions better than non-BVI. Also, the effect of repeating a task in the same environment, i.e., performing different rounds of the same experiment, may differ between sighted and blind users.

Generally, the results reinforce the importance of having BVI users involved in the design of assistive devices from the early stages of the specification of requirements. Maybe the differences between the two groups would have been lesser if the non-BVI users were trained to navigate without sight before the experiment.

6. Future works and suggestions

The following topics are of interest for future research:

- Perform a comparison between an evaluation campaign executed in a real environment and the same campaign executed in a virtual environment, to assess the main differences brought by the use of virtual reality;
- Further improve the virtual reality environment by providing better sound solutions, using different sources of sound in the real environment instead of using the sound from the HMD;
- Develop a solution for automatic collision detection in the virtual system to introduce performance metrics in the assessment.
- Repeat the experimental campaign with large sample of both BVI and non-BVI users, in order to improve the statistical analysis.
- Investigate the sources of noise of the physiological sensors and improve their data acquisition.

References

- , 2019. Technical Report.
- Bourne, R., Steinmetz, J.D., Flaxman, S., Briant, P.S., Taylor, H.R., Resnikoff, S., Casson, R.J., Abdoli, A., Abu-Gharbieh, E., Afshin, A., et al., 2021. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the global burden of disease study. *The Lancet global health* 9, e130–e143.
- Bradley, N.A., Dunlop, M.D., 2002. Investigating context-aware clues to assist navigation for visually impaired people, in: *Proceedings of Workshop on Building Bridges: Interdisciplinary Context-Sensitive Computing*, University of Glasgow.
- Bradley, N.A., Dunlop, M.D., 2005. An experimental investigation into wayfinding directions for visually impaired people. *Personal and Ubiquitous Computing* 9, 395–403.
- Chiu, M.L., 2002. An organizational view of design communication in design collaboration. *Design studies* 23, 187–210.
- Endsley, M.R., 1988. Design and evaluation for situation awareness enhancement, in: *Proceedings of the Human Factors Society annual meeting*, Sage Publications Sage CA: Los Angeles, CA. pp. 97–101.
- Endsley, M.R., 1995. Measurement of situation awareness in dynamic systems. *Human factors* 37, 65–84.
- Endsley, M.R., 2018. Automation and situation awareness, in: *Automation and human performance: Theory and applications*. CRC Press, pp. 163–181.
- Farrell, W.A., 2018. Learning becomes doing: Applying augmented and virtual reality to improve performance. *Performance Improvement* 57, 19–28.
- Kirner, C., Kirner, T.G., Wataya, R.S., Valente, J.A., 2011. Using augmented reality to support the understanding of three-dimensional concepts by blind people.
- Lozano, C.A., Kaczmarek, K.A., Santello, M., 2009. Electrotactile stimulation on the tongue: Intensity perception, discrimination, and cross-modality estimation. *Somatosensory & motor research* 26, 50–63.
- Moerland-Masic, I., Reimer, F., Bock, T.M., Meller, F., Nagel, B., 2021. Application of vr technology in the aircraft cabin design process. *CEAS Aeronautical Journal*, 1–10.
- Sanders, M.S., McCormick, E.J., 1998. Human factors in engineering and design. *Industrial Robot: An International Journal*.
- Siu, A.F., Sinclair, M., Kovacs, R., Ofek, E., Holz, C., Cutrell, E., 2020. Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds, in: *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13.
- Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H.W., 2004. *Handbook of human factors and ergonomics methods*. CRC press.
- Wolf, A., Binder, N., Miehl, J., Wartzack, S., 2019. Towards virtual assessment of human factors: A concept for data driven prediction and analysis of physical user-product interactions, in: *Proceedings of the Design Society: International Conference on Engineering Design*, Cambridge University Press. pp. 4029–4038.