Virtual reality for the human-centred design of assistive devices

Prof. Ph.D Emilia Villani[a], Ph.D Edmar Thomaz da Silva[a], M.Sc Ivan Rehder[a]

[a]*ITA, Praça Marechal Eduardo Gomes, 50, 12228-900, São José dos Campos - SP, Brazil.*

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.

Keywords: asda,asdfa,asdfasdf,asdfasdf

1. Introduction

According to the World Health Organisation (WHO), 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 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 et al. (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

Email addresses: evillani@ita.br (Prof. Ph.D Emilia Villani), evillani@ita.br (Ph.D Edmar Thomaz da Silva), ivan@ita.br (M.Sc Ivan Rehder)

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 et al. (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 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 et al. (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 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 et al. (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 et al. (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 associeated 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 et al. (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 essencial 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 essencial 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?

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 **??** summarizes the main conclusions of this work and discusses future work.

2. Methodology

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;

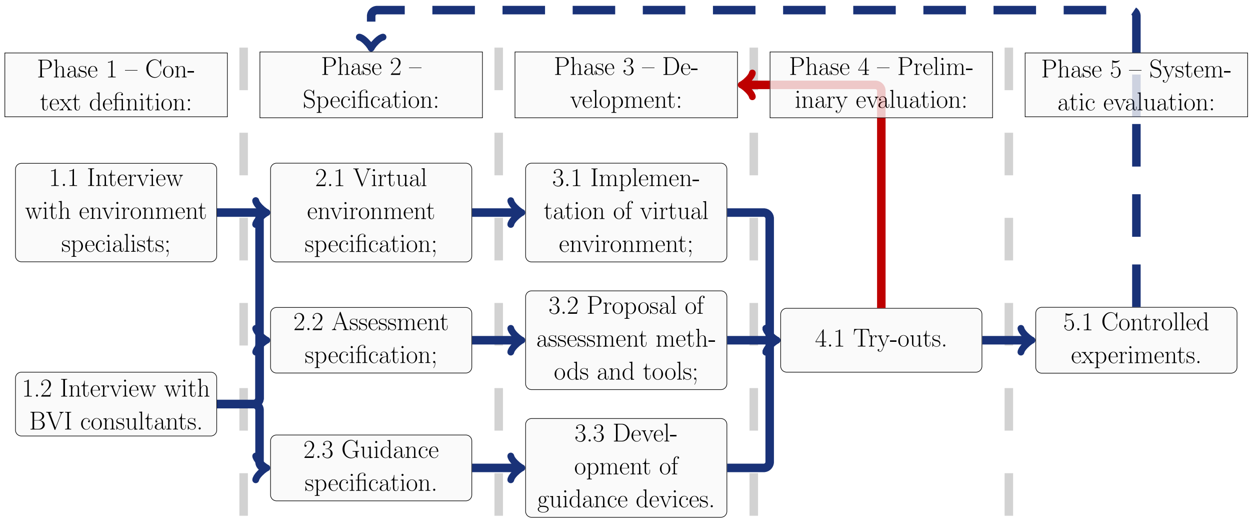


Figura 1: Method's diagram

4. Leave the room when called.

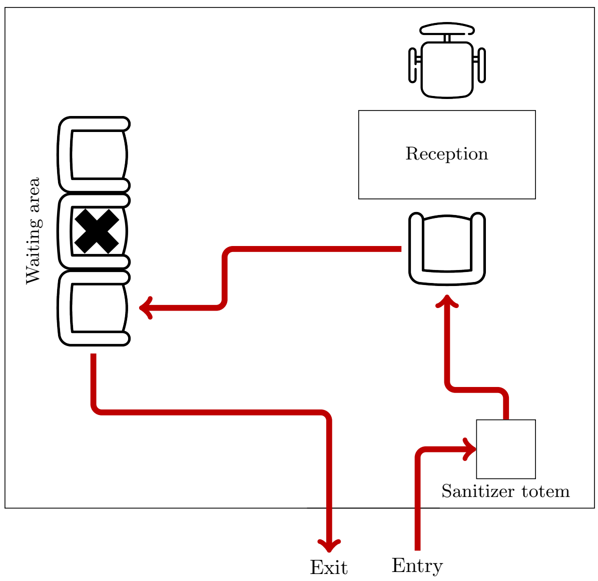


Figura 2: Scheduled task of the experiment and their order.

Figure 3 shows the virtual environment created in Unity3D and the corresponding real environment assembled at the CCM entrance hall.

Human Factors techniques

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

A) 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.

B) 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.

(a) Virtual environment screenshot

(b) Real environment photo

Figura 3: Environment comparisson

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

C) 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.

A) 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.

B) 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.

C) 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.

D) 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.
- 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

The purpose of the experiment discussed in this chapter is to investigate the two research questions proposed for this work:

- Is it possible to evaluate and compare concepts of an assistive device 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 device as BVI users?

For this purpose, the experiment described in Section 2 was performed with the following groups:

- 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.

In order to answer the two research questions, this chapter is organized in the following way. Section 4 is dedicated to the first question and brings an analysis performed only with data from blind participants. Then, Section 5 repeats the same analysis now with data from sighted participants and compares the results with those obtained from blind participants in order to answer the second research question.

In both sections, the data analysis follows the following sequence:

- Analysis of subjective questionnaires:
  - NASA-TLX (National Aeronautics and Space Administration Task Load Index): it aims at assessing the workload perceived by the user in six dimensions, including 'mental demand'. It is expected a decrease in the mental workload between the 'first' to the 'return' round. It is also expected that some guidance methods would differ regarding the required mental workload.
  - Adapted SAGAT (Situation Awareness Global Assessment Technique): it aims at assessing the situation awareness and the user's mental map. It is expected that the SAGAT score would increase from the 'first' to the 'return' round. It is also expected that some guidance methods would differ regarding the required situation awareness provided to the user.
  - Guidance method's questionnaire: It assess the user experience with each method. It is also expected that some guidance methods would differ regarding the score received in this questionnaire.
- Analysis of physiological sensors:
  - ECG (Electrocardiogram): it aims at assessing the user workload. Two features are extracted from the ECG signal, heart rate (BPM) and heart rate variance (SDNN). The heart rate is expected to decrease slightly from the 'first' to the 'return' round, while the heart rate variance is expected to increase slightly.
  - GSR (Galvanic Skin Response): it aims at assessing the user workload and stress. It is expected that the GSR average would increase at every 'first' round and then a slight decrease in the 'return' round.

The solution to the collision detection problem would be the acquisition of additional sensors to monitor the user's hands, arms and legs. However, this solution was not feasible within the time frame available for this work. As a result, the performance was removed from the list of factors to be evaluated. Developing an automatic solution for detecting collision in the virtual world is recommended as future work.

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 experiment is repeated with a set of non-BVI users and the same data is collected. The purpose is to investigate whether or not performing the analysis with non-BVI users could lead to different conclusions.

The experiment has an approval of the ethics commitee.

4. Evaluation of assistive device from a human factors' perspective in a virtual environment

4.1. Subjective data

4.2. 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

Table 1 presents the "mental demand"score of each blind participant to each guidance method. The higher the value, the higher the mental demand is. The base method refers to the guidance method that the person uses in his/her daily life (e.g., white cane).

Tabela 1: Score of NASA-TLX mental demand for the blind participants.

Participant	Round	Base	Audio	Haptic Belt	Virtual Cane	Mixture
001C	First	3	1	14	3	6
	Return	1	1	10	2	6
002C	First	5	1	1	10	12
	Return	1	1	1	10	3
003C	First	5	5	5	8	1
	Return	3	1	1	2	1
004C	First	9	10	15	10	10
	Return	7	10	14	8	10

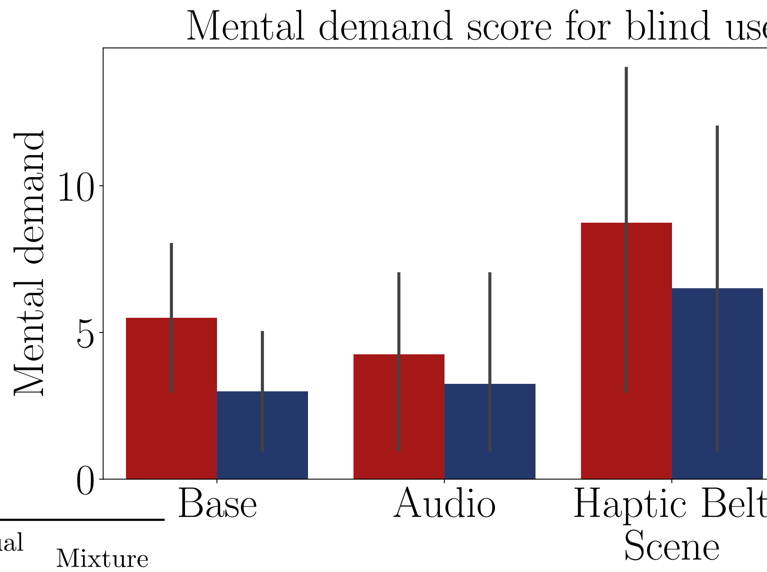


Figura 4: Mean and standard deviation of mental demand of blind participants for each method.

The mean value obtained for each guidance method is illustrated in Figure 4. It shows a systematic reduction in the perceived mental workload between the rounds for all methods, confirming that the participants get familiar with the devices after the first use. It also shows that although the haptic belt obtained the most considerable mean, it also had the most significant variation, showing that the effort required from the user may vary significantly.

Figure 5 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 6 presents a boxplot of the mental demand grouped by the rounds, confirming the general tendency to reduce the required "mental demand".

In order to support the statistical analysis, Figures 7 and 8 presents the QQ-plot and the residual plot of the "mental demand"data, confirming that the data follow a normal distribution and the residues are homogenous.

Figures 7 and 8 show the distribution and variance of Table 1. These figures show that the data are normally distributed and that the methods have a similar variance. Table 2 shows the ANOVA test p-values of the mental demand of the "blind" sample between the guidance methods. The methods' and the rounds' p-values indicate that there is no influence from them in the mental demand. The interaction between the methods and the round also does not influence the mental demand.

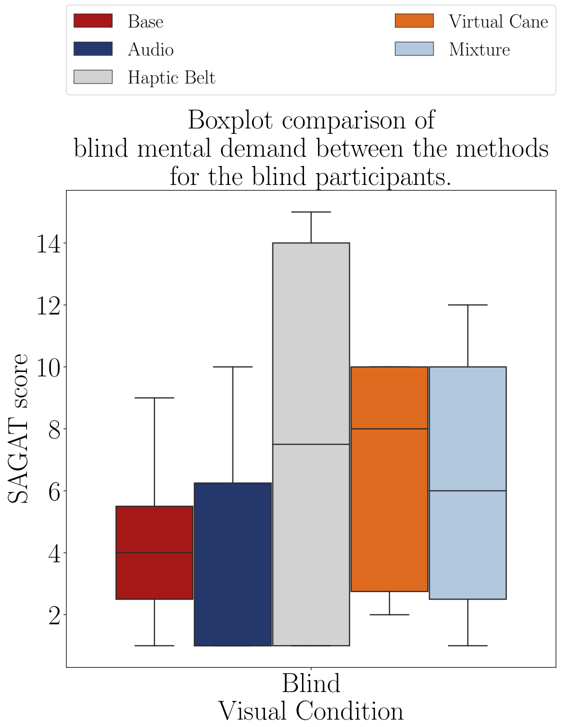


Figura 5: Boxplot of the mental demand of the blind participants grouped by the methods.

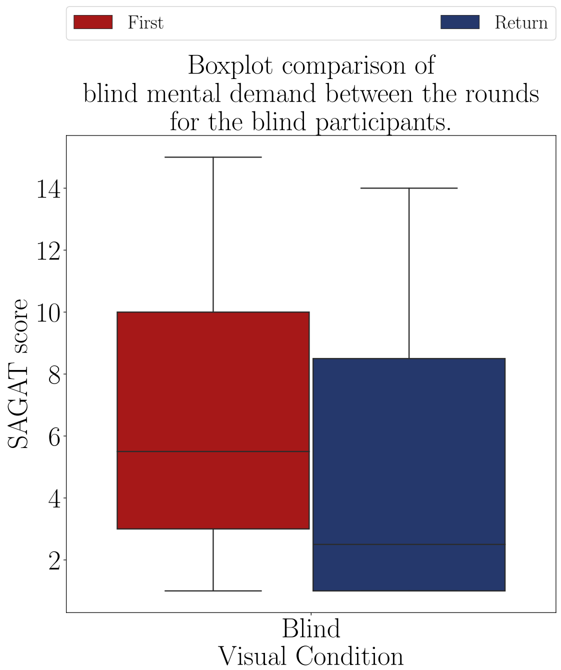


Figura 6: Boxplot of the mental demand of the blind participants grouped by the rounds.

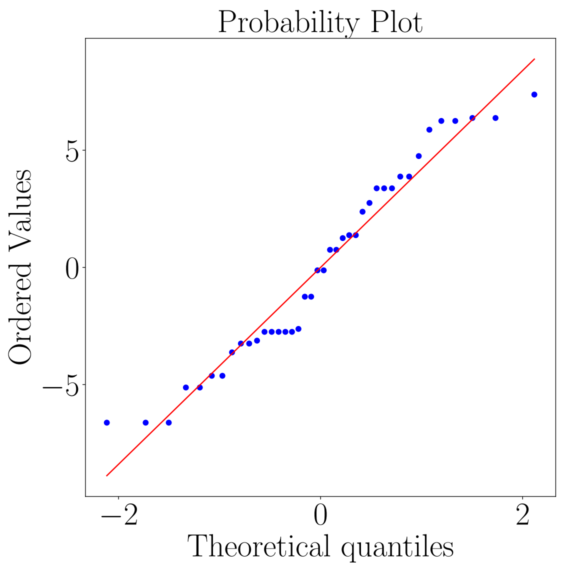


Figura 7: QQ plot of the mental demand of the blind participants on each method.

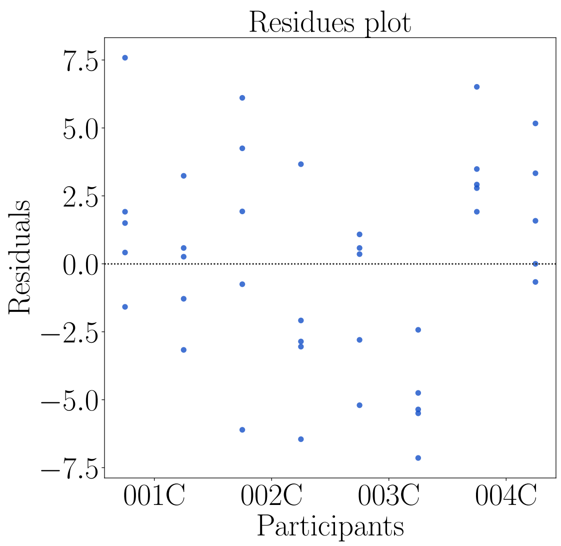


Figura 8: Residual plot of the mental demand score the blind participants on each method.

Following, the statistical model of Equation 5.1 is used for the analysis of variance (ANOVA). The ANOVA analyses the influence of the used method, the rounds and the interaction of those two in the analysed variable:

$$y_{ijk} = \mu + \tau_i + \beta_j + \omega_k + (\tau\beta_{ij}) + e \qquad (1)$$

where:

- y_{ij} - output variable for method i, round j and participant k;
- μ - mean of all the observations;
- τ_i - variance from method i;
- β_j - variance from round j;
- ϖ_k - variance from participant k, which is treated as a block;
- $\tau\beta_{ij}$ - combined variance from the interaction between method i and round j;
- e - residual error.

The results of ANOVA are presented in Table 2. ANOVA tests the hypothesis that the means of independent data groups are equal or not. In the literature, a p-value of 0.05 is commonly adopted as a threshold to confirm the hypothesis. A p-value < 0.05 indicates that the means of the groups are statistically different with 95% of confidence. According to this criterion, neither method or round have a significant influence on the mental demand.

However, due to the low number of participants, the threshold of 0.1 could also be considered. In this case, it indicates, with 90% confidence, that the mean of the first and return rounds are different. For the guidance method, the p-value of 0.170 is close to the threshold but slightly higher, suggesting that the means may be different. However, this hypothesis is not statistically confirmed with the current data.

Tabela 2: ANOVA p-value for mental demand – blind participants

Source	P-Value
Methods	0.170
Rounds	0.075
Interaction	0.993

In order to conclude the analysis of the NASA-TLX mental demand, Table 3 brings the average difference between the mental demand of the first and return rounds. If the variation is positive, it means that the user had an increase on his/her mental demand. Unexpectedly, it shows that the most significant variation is obtained to the base, i.e., the guidance method the participant uses and, therefore, should not present a significant variation. The methods with the lower variation was audio, probably because it already had a shallow score in the first round.

Tabela 3: Mental demand variation grouped by participant and visual condition

Visual Condition	Base	Audio	Haptic Belt	Virtual Cane	Mixture
Blind	-2.5	-1.0	-2.2	-2.2	-2.2

Analysis of the NASA-TLX score

This section repeats the analysis steps of the previous section but now considers the mean value of all dimensions of NASA-TLX, referred to in this text as the global score. Again, the higher the value of the average, higher the mental workload perceived by the user. Table 4 presents the global score of each blind participant.

Tabela 4: NASA-TLX score felled by the blinded participants.

Participant	Round	Base	Audio	Haptic Belt	Virtual Cane	Mix
001C	First	4.833	4.000	8.833	5.167	6
	Return	4.167	4.000	6.667	4.500	6
002C	First	6.333	4.833	4.833	9.000	7
	Return	4.500	4.833	4.833	7.000	5
003C	First	4.000	4.000	5.333	6.667	3
	Return	4.000	3.833	3.667	3.500	3
004C	First	9.833	10.000	12.667	9.667	11
	Return	8.667	9.167	11.667	9.333	10

Figure 9 brings the corresponding barplot with the mean value and standard deviation for each guidance method and each round. In a qualitative comparison with Figure 4, the differences between the methods are confirmed but softened. It is possible to notice that the mean score of audio and base are still lower than that of the other methods. The differences between first and return rounds are also reduced. However, the standard deviation is also considerably reduced for all methods, and especially for the haptic belt.

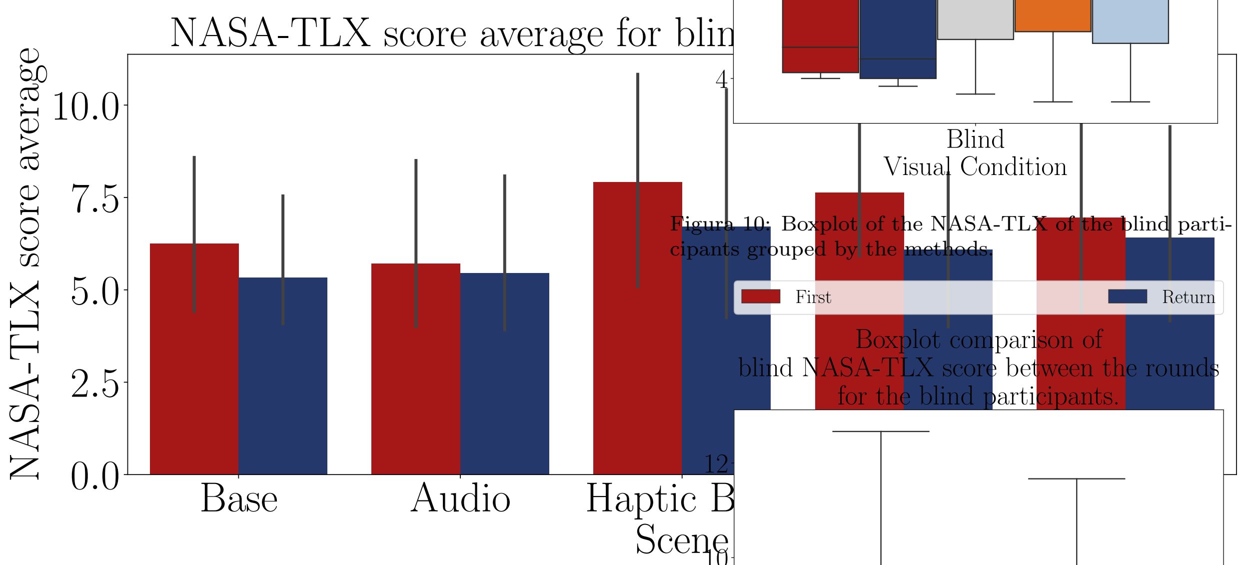


Figura 9: Barplot of the average NASA-TLX score of the blind participants on each method.

Figure 10 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 methods into two different groups: base and audio, which require a lower level of workload, and another group, which requires a higher level. Figure 11 presents a boxplot with the NASA-TLX global score grouped by the rounds, showing that the two groups are still different.

Figures 12 and 13 presents the QQ plot and residual distribution of the NASA-TLX global score,

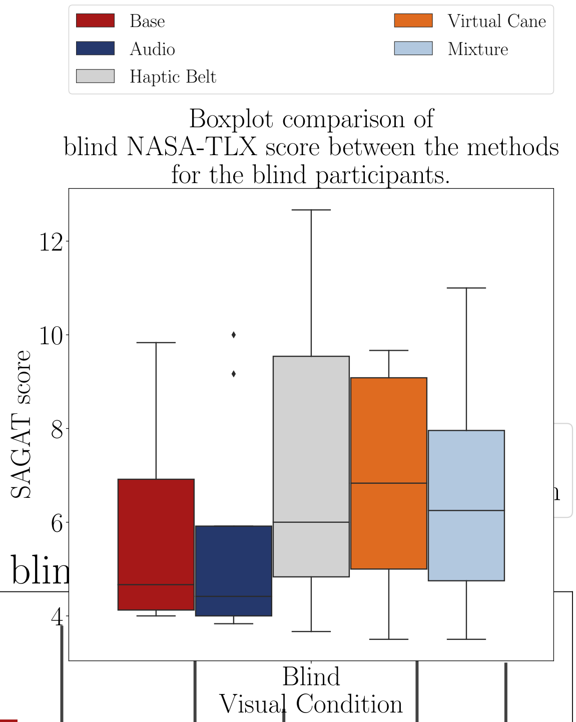


Figura 10: Boxplot of the NASA-TLX of the blind participants grouped by the methods.

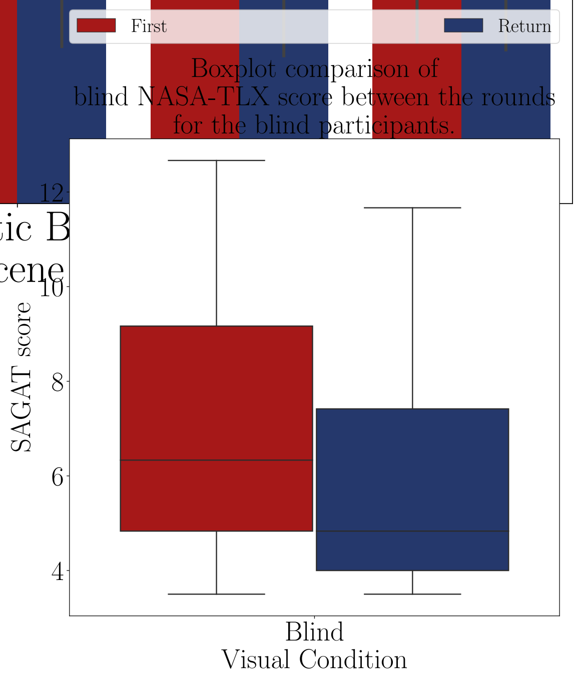


Figura 11: Boxplot of the NASA-TLX demand of the blind participants grouped by the rounds.

showing that the data are normally distributed. However, the residuals are not so homogeneous as in the previous case, showing that the participants have different variability among them.

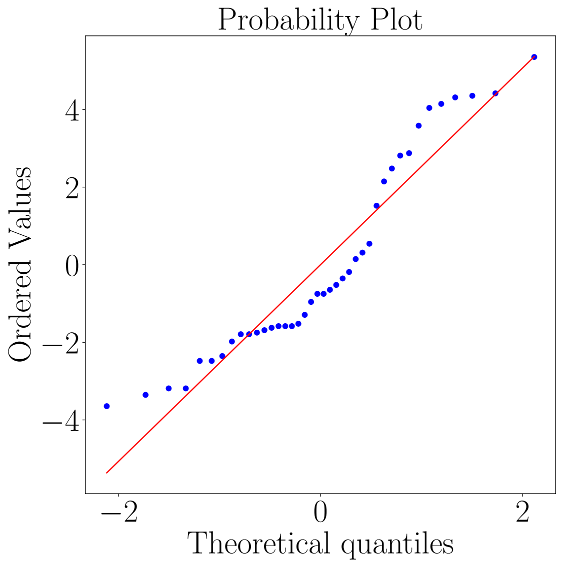


Figura 12: QQ plot of the NASA-TLX score of the blind participants on each method.

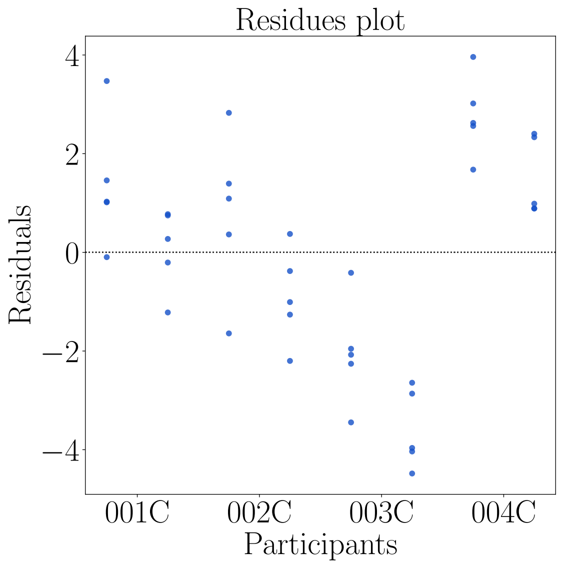


Figura 13: Residual plot of the NASA-TLX score the blind participants on each method.

Table 5 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.

Tabela 5: Anova p-value for the NASA-TLX score on each method for blinded users.

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

Finally, Table 6 presents the results of a pairwise Fisher LSD test comparing each pair of guidance methods. The results show that only audio is similar base. All the other methods are different from each other.

Tabela 6: Cross validation p-value for the NASA-TLX score on each method for blinded users.

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

Table 7 shows the difference in the NASA-TLX global score between the first and return rounds. As before, if the variation is positive, it means that the user had an increase on his/her Mental Workload. It shows that the audio difference is the lowest among all methods, while the highest difference is for the virtual cane.

Tabela 7: NASA-TLX score grouped by participant and visual Condition.

Visual Condition	Base	Audio	Haptic Belt	Virtual Cane	Mixture
Blind	-0.92	-0.25	-1.21	-1.54	-0.54

4.2.1. Adapted SAGAT

This section discusses the results of the adapted SAGAT questionnaire, which aims at assessing the participant's situation awareness and their mental map of the environment.

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. The total score achieved by each blind participant is presented in Table 8. Figure 14 illustrates the corresponding bar plot, indicating the mean and standard deviation for each guidance method and each round. This figure shows clearly that the participants improved their situation awareness in the return round, when they already had some information about the environment. Also, it is possible to observe that the worst situation awareness is obtained in the first round for the virtual cane. However, on the return round, the SAGAT mean score becomes equivalent to the audio method.

Tabela 8: SAGAT global score felled by the blinded participants.

Participant	Round	Base	Audio	Haptic Belt	Virtual Cane	Mixture
001C	First	6.25	5.50	5.33	5.83	3.500
	Return	6.25	6.50	8.50	5.50	5.500
002C	First	6.75	4.50	3.99	4.50	6.250
	Return	5.25	5.00	4.00	6.50	8.500
003C	First	7.25	7.50	7.49	4.66	9.000
	Return	10.00	10.00	8.50	9.00	9.000
004C	First	7.50	6.00	7.66	4.99	6.500
	Return	9.00	6.00	9.25	7.25	9.000

Figure 15 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 16 shows the boxplot of the data grouped by round and confirms the general improvement of situation

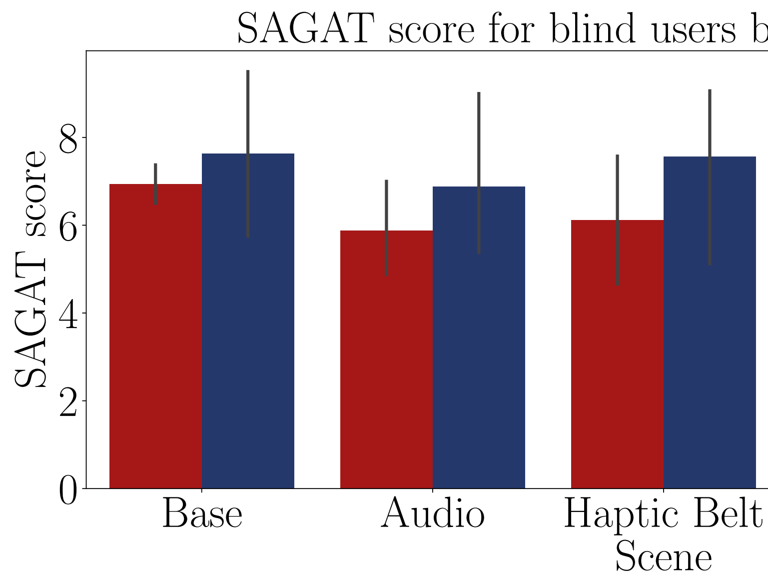


Figura 14: Barplot of the average SAGAT score of the blind participants on each method.

awareness from the first to the return round.

Proceeding to the statistical analysis of the data, Figures 17 and 18 present the QQ plot and the residual distribution, which confirms the normal distribution assumption and the homogeneity of variances.

Table 9 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.

Tabela 9: Anova p-value for the SAGAT score on each method for blinded users.

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

Finally, Table 10 presents the mean difference in the SAGAT score between the first and return

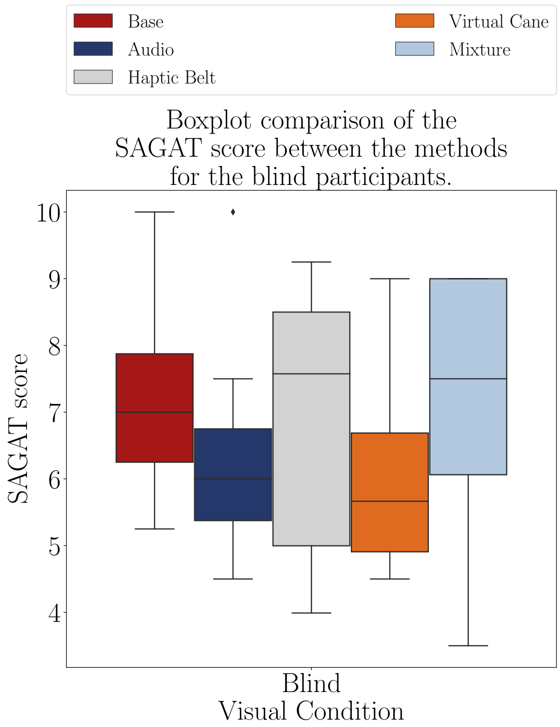


Figura 15: Boxplot of the SAGAT score of the blind participants grouped by the methods.

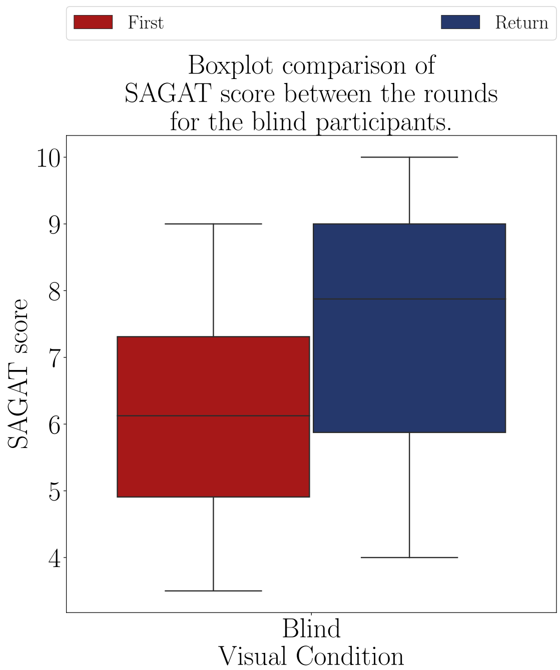


Figura 16: Boxplot of the SAGAT score of the blind participants grouped by the rounds.

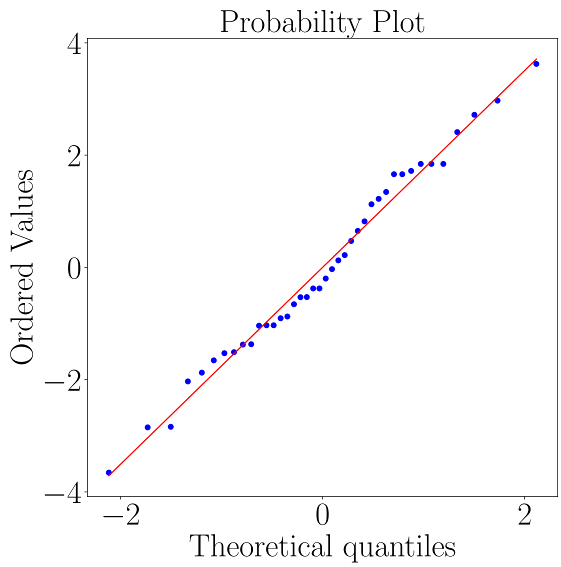


Figura 17: QQ plot of the SAGAT score of the blind participants on each method.

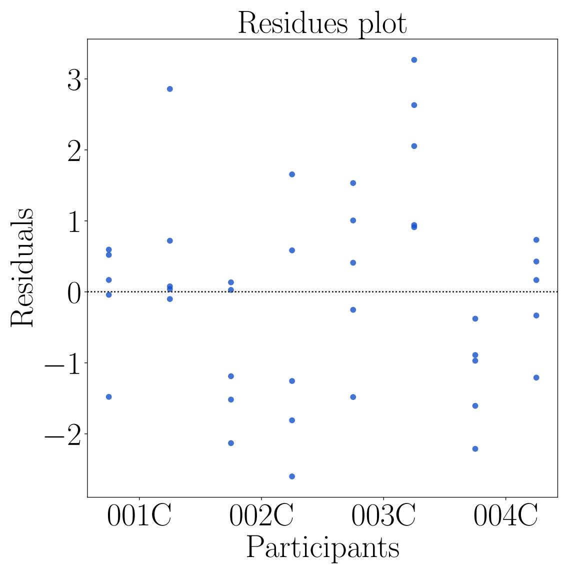


Figura 18: Residual plot of the SAGAT score the blind participants on each method.

rounds for each guidance method. If the variation is positive, it means that the user had an increase on his/her Situation Awareness. It shows that the base and audio methods have the lowest difference, while the highest was obtained for the virtual cane.

Tabela 10: Adapted Sagat global score variation grouped by participant and visual Condition

Visual Condition	Base	Audio	Haptic Belt	Virtual Cane	Mixture
Blind	8.93	15.66	23.49	44.30	32.90

4.2.2. 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.

Table 11 presents the score attributed to each method by each participant. The mean values are plotted in Figure 19 and show a dissatisfaction with the methods that only use vibration for communicating with the participant, i.e., the haptic belt and the virtual cane.

Tabela 11: Guidance method questionnaire score felled by the blinded participants.

Participant	Audio	Haptic Belt	Virtual Cane	Mixture
001C	0.774	0.543	0.629	0.865
002C	0.857	0.743	0.543	0.935
003C	0.929	0.571	0.543	0.745
004C	0.881	0.486	0.400	0.730

Figure 20 brings the questionnaire boxplot, which clearly shows the difference between two groups: haptic belt and virtual cane, and audio and mixture.

Figures 21 and 22 show that the data follows a normal distribution. However, the residual variance is not strictly homogenous among the participants.

The results of ANOVA are presented in Table 12

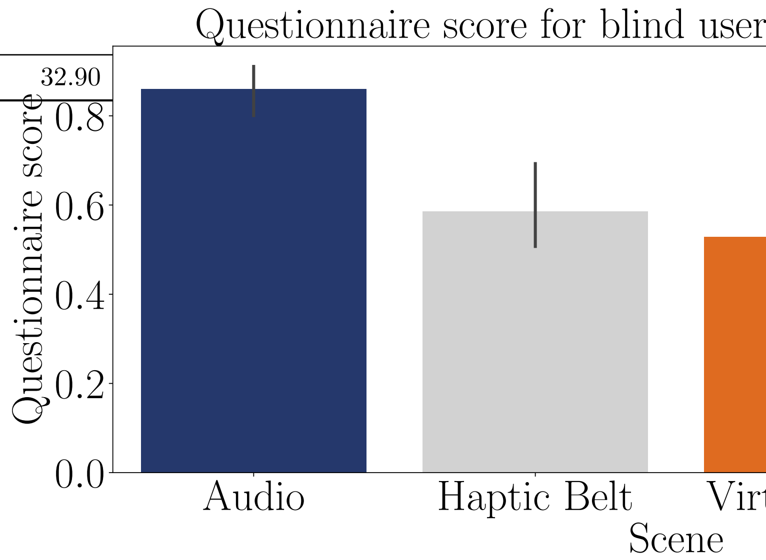


Figura 19: Barplot of the average questionaire score of the blind participants on each method.

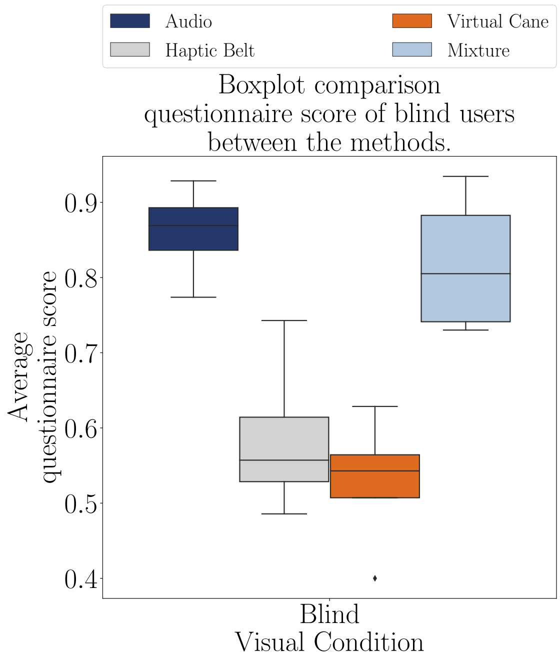


Figura 20: Boxplot of the questionaire score of the blind participants grouped by the methods.

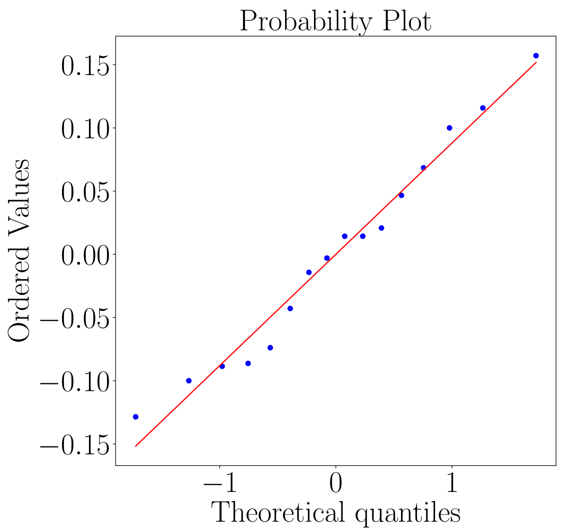


Figura 21: QQ plot of the questionnaire score of the blind participants on each method.

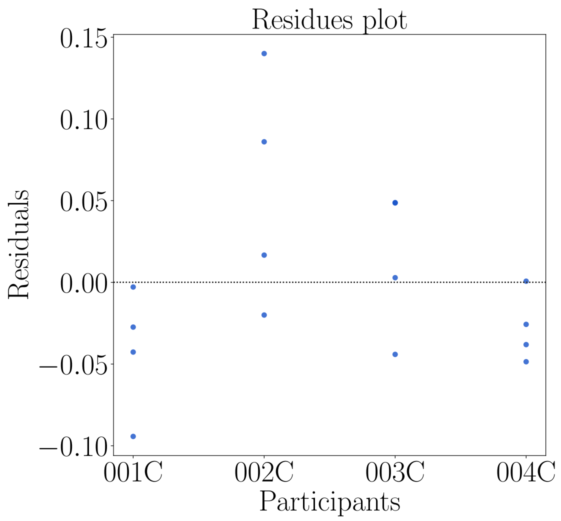


Figura 22: Residual plot of the questionnaire score the blind participants on each method.

and it shows that the method, with a p-value of 0.001, is indeed a significant variable that affects the user's satisfaction.

Tabela 12: Anova p-value for the questionnaire score on each method for blinded users.

Source	P-Value
Method	0.001**

In order to complement the ANOVA analysis, the pairwise comparison of the methods obtained from the Fisher LSD test is presented in Table 13. The results show that audio and mixture are equivalent from the perspective of user satisfaction. All the other comparisons indicate there is a difference between the methods.

Tabela 13: Cross validation p-value for the questionnaire score on each method for blinded users.

Method			Analysis
Audio	X	Haptic Belt	$H_1 : \mu_{Audio} \neq \mu_{HapticBelt}$
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	$H_1 : \mu_{HapticBelt} \neq \mu_{VirtualC}$
Haptic Belt	X	Mixture	$H_1 : \mu_{HapticBelt} \neq \mu_{Mixture}$
Virtual Cane	X	Mixture	$H_1 : \mu_{VirtualCane} \neq \mu_{Mixtur}$

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.

4.3. Physiological data

During the experiment, data from two physiological sensors were captured: ECG and GSR. As commonly found in the literature, these data are used to assess mental workload. The corresponding analysis is presented in this section.

Electrocardiogram (ECG) data

As previously stated, the ECG analysis is based on two variables: the heart rate (BPM – beats per minute) and heart rate variance (SDNN – standard deviation of NN intervals).

After the experiment, the ECG signal processing is organized in the following steps (Figure 23):

- 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.

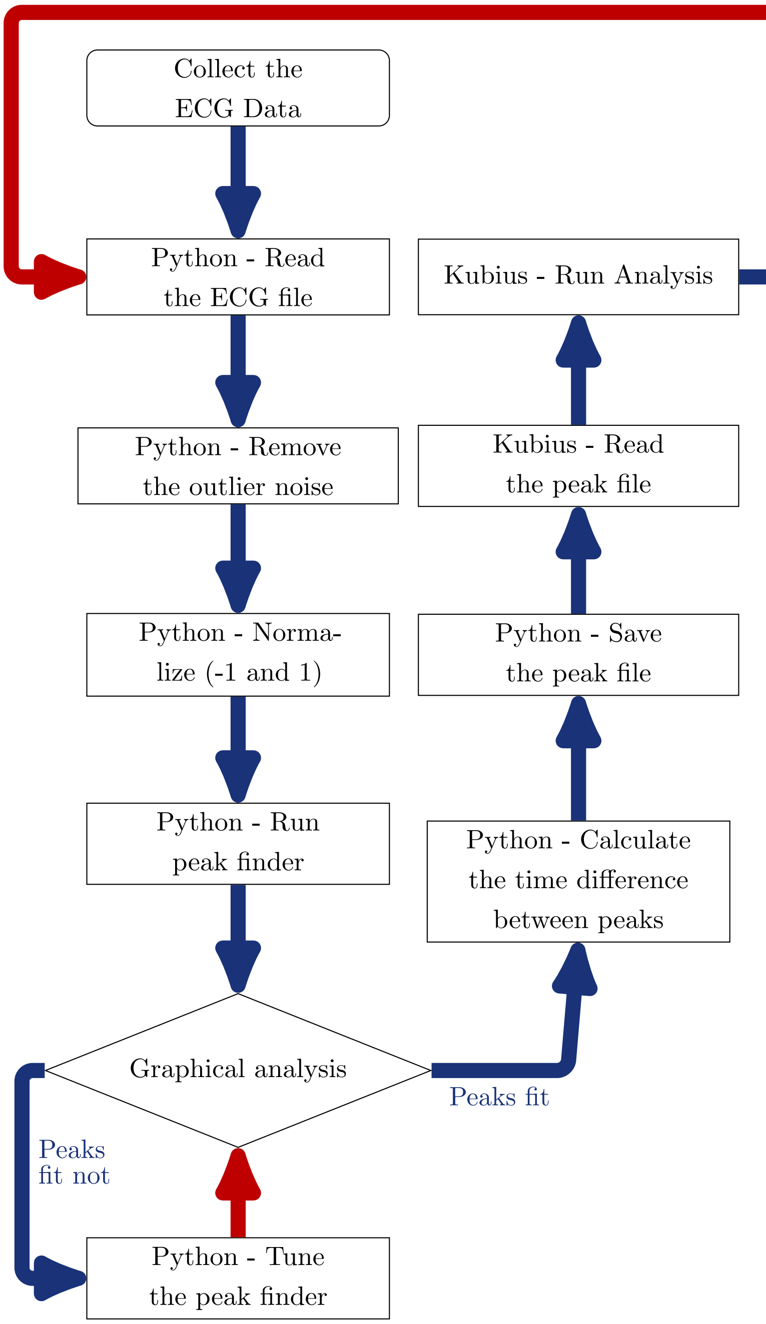


Figura 23: ECG's data treatment algorithim.

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)

Table 14 presents the heart rate of each blind participant for each guidance method. 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. It is possible to observe that there is no systematic difference between the methods. Also, there are significant differences among the participants, with some presenting values significantly lower than others.

Tabela 14: Average BPM felled by the blinded participants [BPM].

Participant	Round	Base	Audio	Haptic Belt	Virtual Cane	Mixture
001C	First	75.75	60.71	71.17	59.07	68.24
	Return	71.05	58.61	66.22	64.20	70.76
002C	First	48.69	38.67	48.74	46.89	52.23
	Return	52.46	47.58	58.97	56.75	58.25
003C	First	68.37	69.89	70.95	69.41	66.94
	Return	67.34	67.44	69.68	68.82	67.37
004C	First	75.09	73.55	73.70	71.94	74.03
	Return	74.74	74.79	74.02	72.69	67.34

Figure 24 presents the mean heart rate. It shows a slight increase in the heart rate between the rounds, except for the base method, indicating that the participants had a higher Mental Workload on the return round.

Figures 25 and 26 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.

Figures 27 and 28 bring the QQ Plot and residual distribution. The last one shows that the partici-

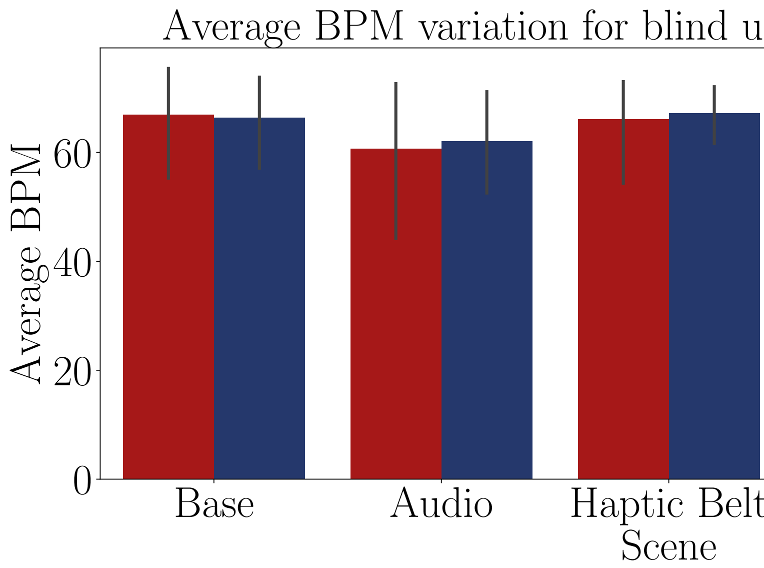


Figura 24: Barplot of the average BPM of the blind participants on each method.

pants do not have a similar variance, which may jeopardize the results of ANOVA. Considering this limitation, Table 14 brings the p-value obtained by ANOVA, which confirmed the previous analysis, as it does not indicate a significant influence of either the guidance methods or the rounds in the participants' heart rate.

Tabela 15: Anova p-value for the BPM on each method for blinded users.

Source	P-Value
Methods	0.100
Rounds	0.371
Interaction	0.894

Analysis of the heartbeat variance (SDNN)

Table 16 brings the value of the second variable extracted from the ECG: SDNN, the standard deviation of the interbeat interval. Different to the heart

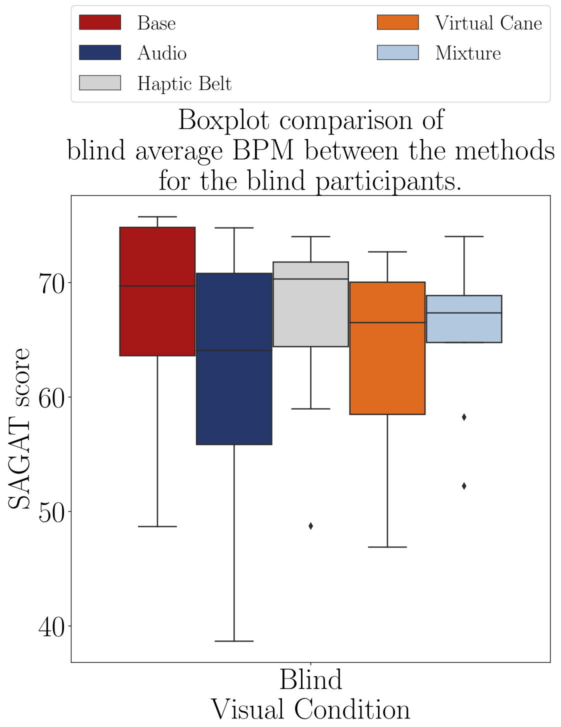


Figura 25: Boxplot of the BPM of the blind participants grouped by the methods.

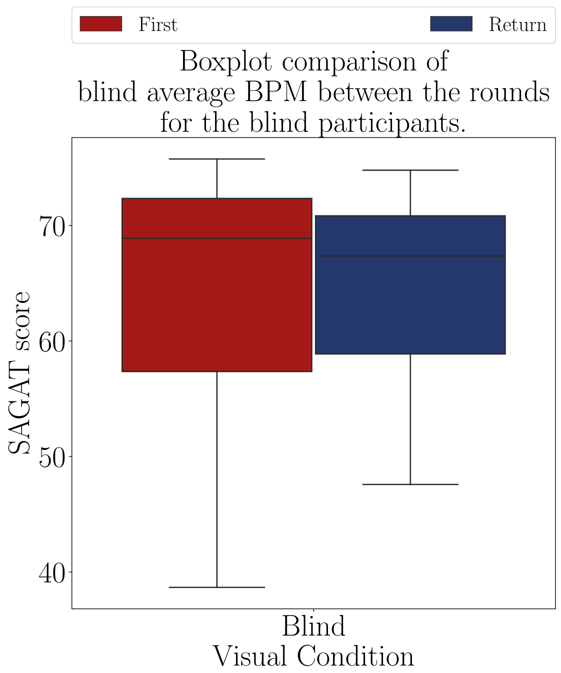


Figura 26: Boxplot of the BPM of the blind participants grouped by the rounds.

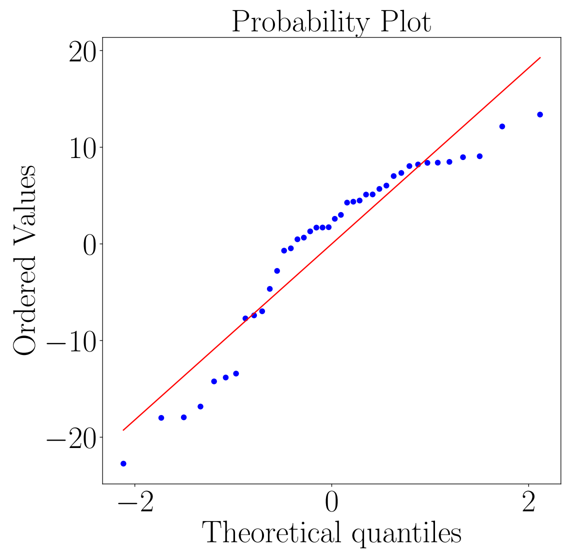


Figura 27: QQ plot of the BPM of the blind participants on each method.

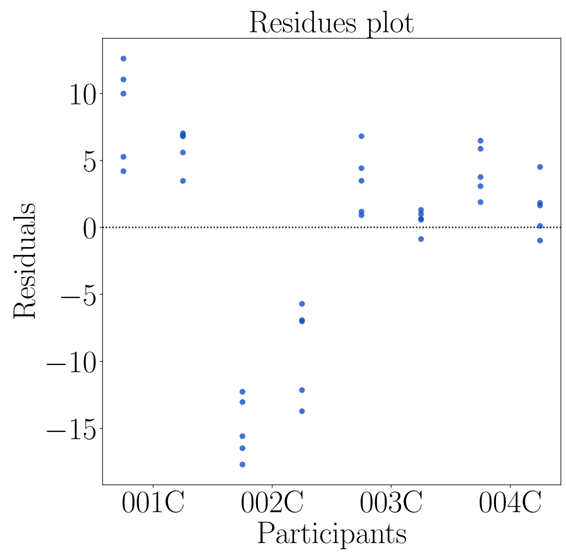


Figura 28: Residual plot of the BPM score the blind participants on each method.

rate frequency, if the variation between the First and the Return round is negative, it means that the user had an increase on his/her mental workload and vice-versa. Similar to what is observed for the BPM, it is not possible to draw a pattern from this data. The participants had increased or decreased their heart rate with different methods.

The barplot of Figure 29 shows the average SDNN

Tabela 16: Average SDNN of the blind participants during the each round and method [ms].

Participant	Round	Base	Audio	Haptic Belt	Virtual Cane	Mixture
001C	First	81.292	107.061	124.737	163.968	
	Return	120.719	130.885	131.590	157.589	
002C	First	73.761	98.863	81.140	33.977	
	Return	108.940	49.627	42.815	114.057	
003C	First	36.870	38.325	35.101	42.392	
	Return	52.750	41.196	44.256	42.602	
004C	First	70.728	86.827	62.560	85.900	
	Return	71.950	74.895	70.017	66.089	

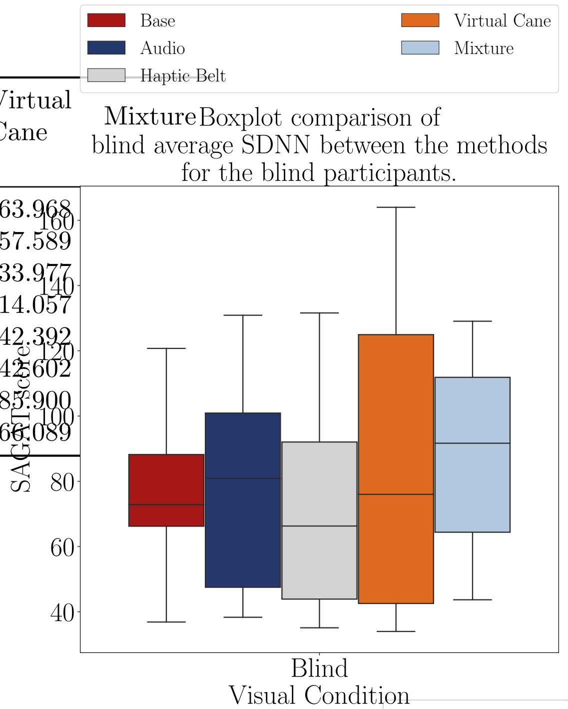


Figura 30: Boxplot of the SDNN of the blind participants grouped by the methods.

for each method. It is possible to notice that some methods are associated with an increase in the SDNN between the rounds, while others present a slight decrease.

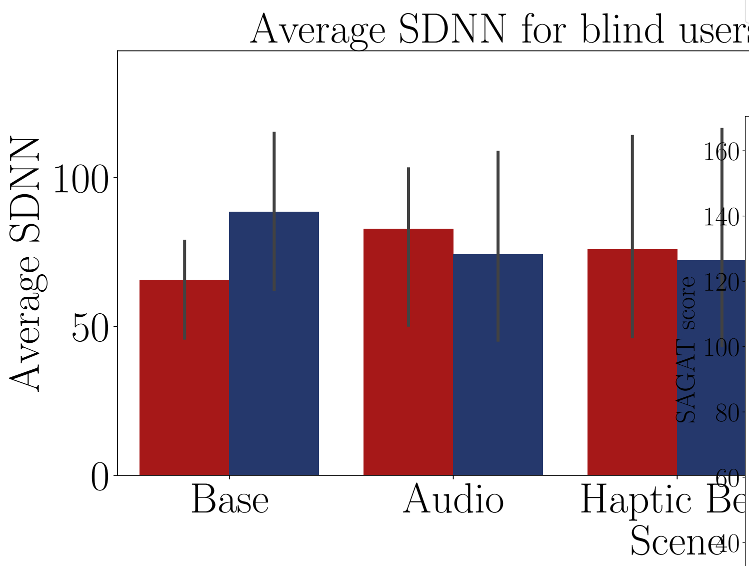


Figura 29: Barplot of the average SDNN of the blind participants on each method.

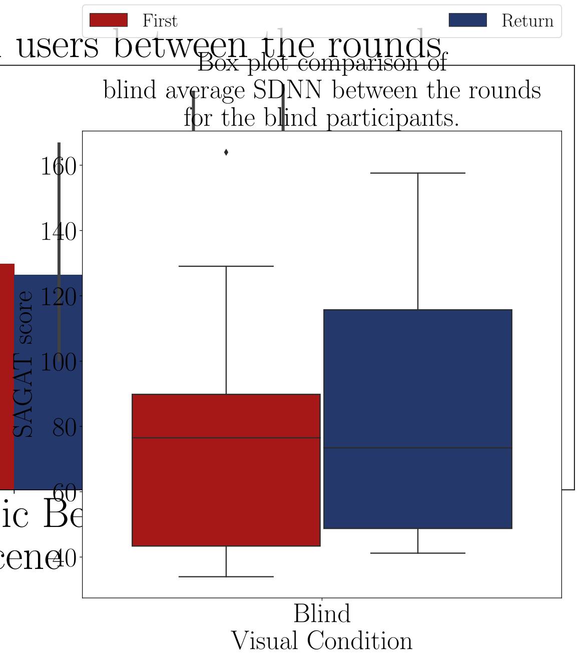


Figura 31: Boxplot of the SDNN of the blind participants grouped by the rounds.

Figure 30 and Figure 31 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.

Figures 32 and 33 bring the QQ Plot and residual distribution. In this case, the residual distribution is more uniform than in Figure 28. The ANOVA results are presented in Table 17 and do not confirm any influence of the methods nor the rounds on the ECG heart rate variance.

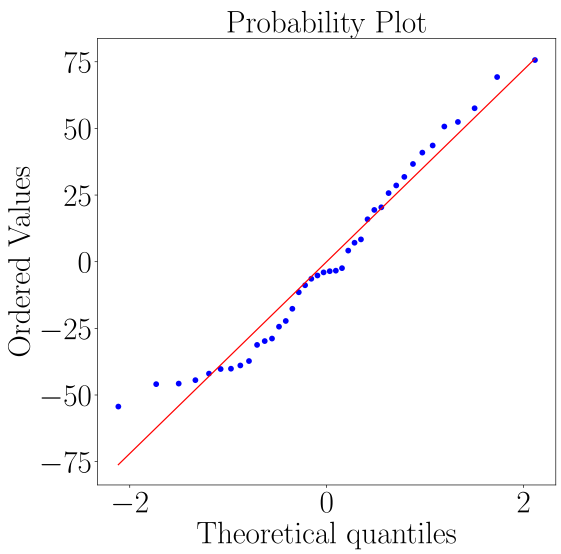


Figura 32: QQ plot of the SDNN of the blind participants on each method.

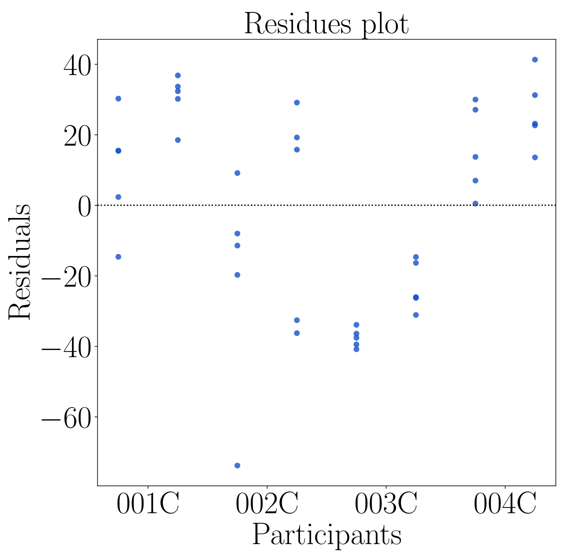


Figura 33: Residual plot of the SDNN of the blind participants on each method.

Tabela 17: Anova p-value for the average SDNN on each method for blinded users.

Source	P-Value
Methods	0.486
Rounds	0.223
Interaction	0.473

4.3.1. 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 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.

Table 18 presents the GSR average values for the three remaining participants. 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. For all the participants, the baseline was smaller than the values obtained during the experiment, as expected. Moreover, in most cases, the skin conductance has risen from the first to the return, indicating an increase in the mental workload.

Tabela 18: Average GSR felled by the blind participants [μS].

Participant	Round	Baseline	Base	Audio	Haptic Belt	Virt Can
001C	First	0.37	0.48	1.03	3.14	3
	Return		0.83	1.58	2.81	4
003C	First	0.30	0.56	0.56	0.62	0
	Return		0.62	0.63	0.65	0
004C	First	1.24	2.34	3.07	3.49	2
	Return		2.57	2.95	3.20	2

Table 19 brings the percentual increase in the GSR average compared to the baseline value. Figure 34 shows the corresponding barplot. The presence of a haptic device causes an increase in the skin conductance, hence its mental workload. Also, it is possible to observe the increase in GSR average between the two rounds, except for the haptic belt.

Tabela 19: Average GSR variation in relation to the baseline in each round of the blind participants [μS].

Participant	Round	Base	Audio	Haptic Belt	Virtual Cane
001C	First	30.58%	176.54%	746.10%	920.72
	Return	125.29%	327.42%	656.99%	988.93
003C	First	85.36%	84.23%	104.19%	182.35
	Return	105.34%	109.23%	112.95%	202.35
004C	First	89.62%	148.53%	182.84%	84.33
	Return	108.22%	138.64%	159.00%	78.73

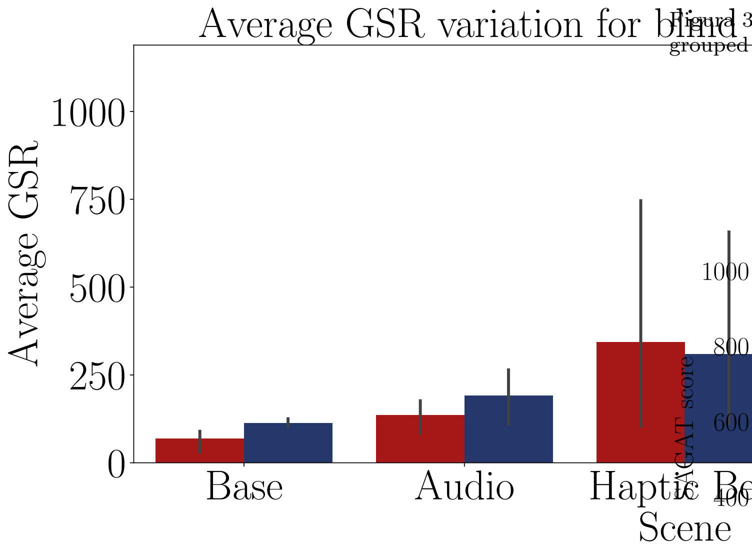


Figura 34: Barplot of the average SDNN of the blind participants on each method.

Figure 35 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 36 presents the GSR grouped by the rounds. In this case, there is no apparent difference between the rounds.

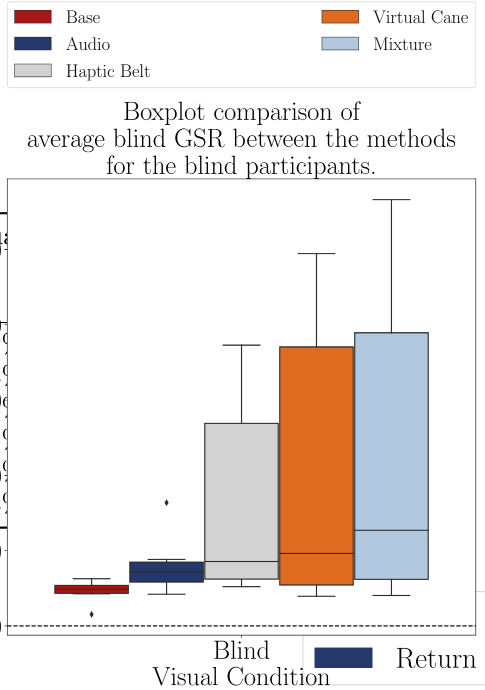


Figura 35: Boxplot of the GSR of the blind participants grouped by the methods.

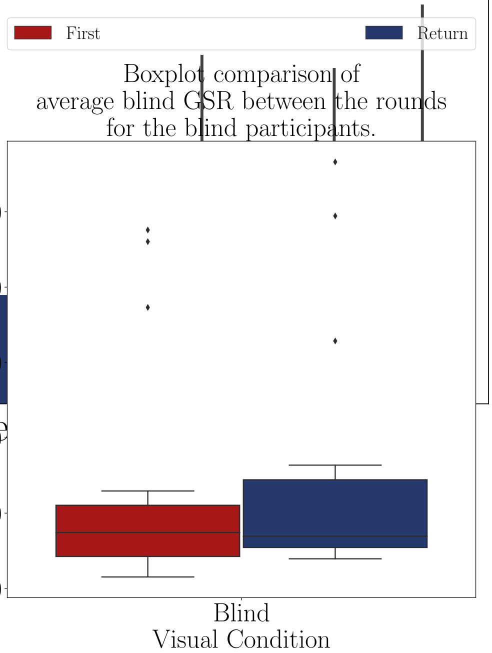


Figura 36: Boxplot of the GSR of the blind participants grouped by the rounds.

Figures 37 and 38 shows the QQ plot and the residual distribution. Table 20 shows the ANOVA test p-value for the GSR percentual variance. Although the p-value for the method is not below the threshold of 0.05, it is close to it, indicating that probably the GSR is affected by it.

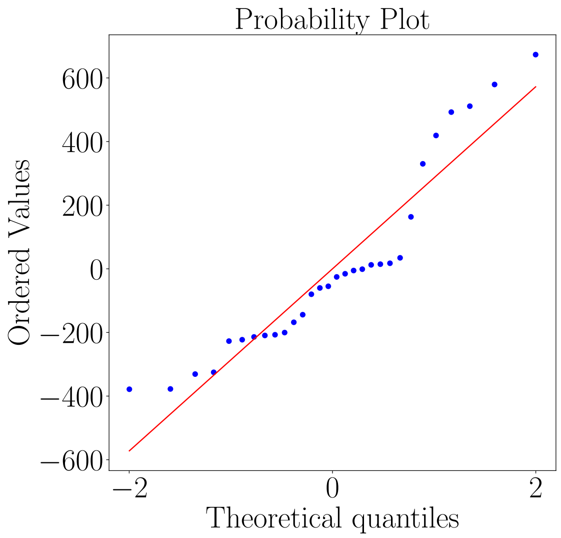


Figura 37: QQ plot of the SDNN of the blind participants on each method.

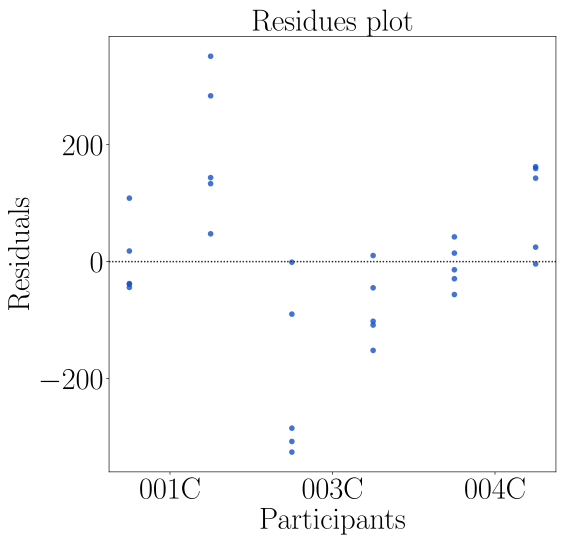


Figura 38: Residual plot of the SDNN of the blind participants on each method.

Tabela 20: Anova p-value for the mental demand average on each method for blinded users.

Source	P-Value
Methods	0.051
Rounds	0.722
Interaction	0.996

4.4. 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.

5. 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.

5.1. Subjective data

5.2. NASA-TLX

5.2.1. Analysis of the mental demand scale

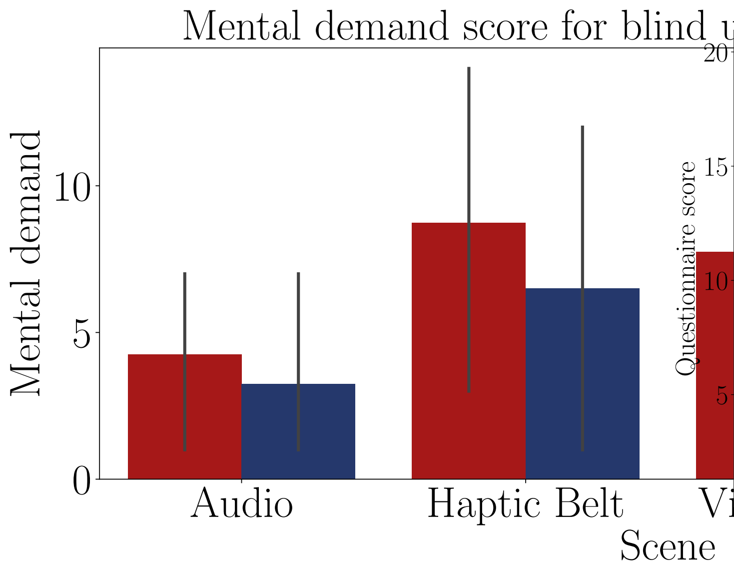
Table 21 presents the mental demand score of all participants, while the corresponding barplot is presented in Figure 39. As said before, the higher the value, the higher is the mental demand of the user. It is interesting to observe that sighted people gave a higher score to audio, as they are not so familiar with using sounds as source of guidance.

Tabela 21: Mental demand felled by the participants.

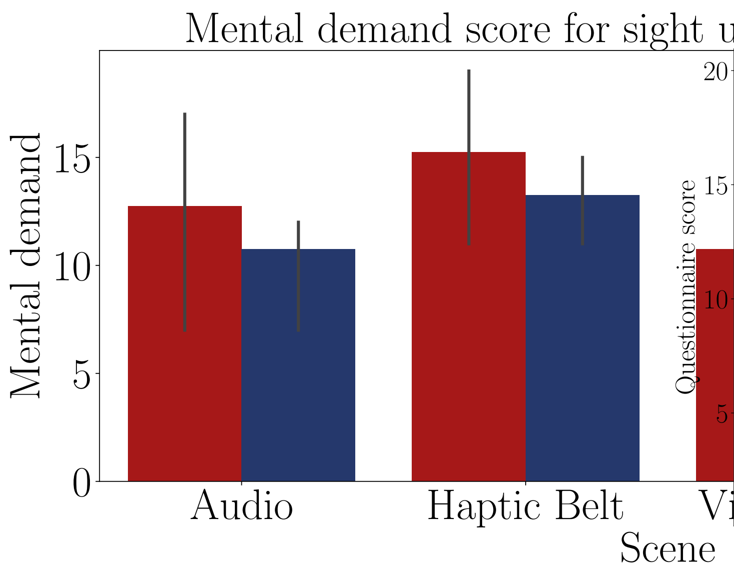
Participant	Visual Condition	Round	Audio	Haptic Belt	Virtual Cane
001C	Blind	First	1	14	3
		Return	1	10	2
002C	Blind	First	1	1	10
		Return	1	1	10
003C	Blind	First	5	5	8
		Return	1	1	2
004C	Blind	First	10	15	10
		Return	10	14	8
001	Sight	First	12	11	5
		Return	13	13	5
003	Sight	First	18	18	16
		Return	12	15	11
004	Sight	First	17	20	12
		Return	12	15	10
005	Sight	First	4	12	10
		Return	6	10	6

Figures 40 and 41 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 prefered to the virtual cane. For both groups, we observe a decrease in the mental demand.

Figures 42 and 43 show the QQ plot and residual distribution for the sighted data, confirming that



(a) Blind participants



(b) Sight participants

Figura 39: Barplot of the average mental demand on each method and each round.

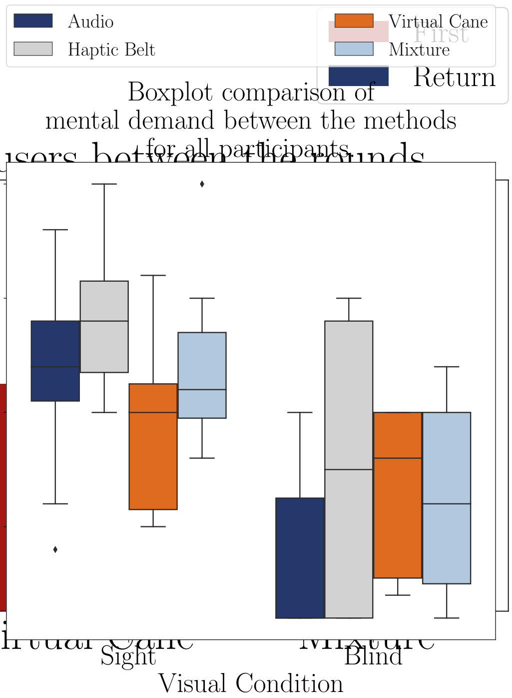


Figura 40: Boxplot of the mental demand of the participants grouped by the methods.

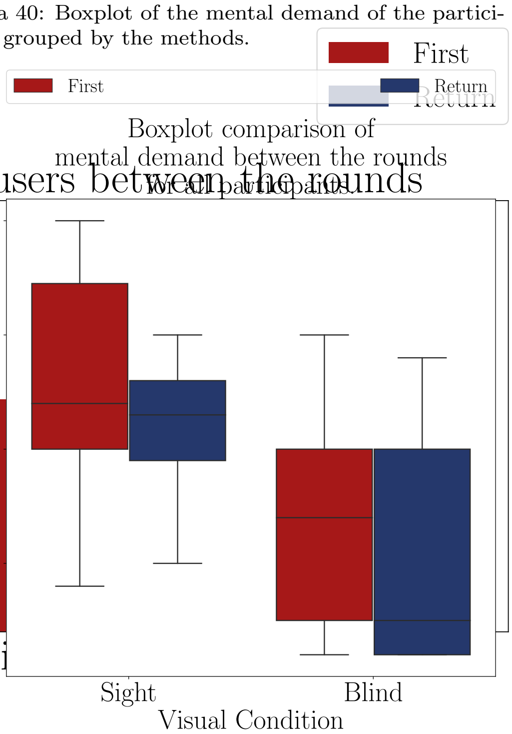


Figura 41: Boxplot of the mental demand of the participants grouped by the rounds.

the data is normally distributed and participants have similar variance. Table 22 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 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.

Tabela 22: Anova p-value for the mental demand average on each method'

(a) Blind participants

Source	P-Value
Methods	0.170
Rounds	0.075
Interaction	0.993

(b) Sight participants

Source	P-Value
Methods	0.049**
Rounds	0.075
Interaction	0.990

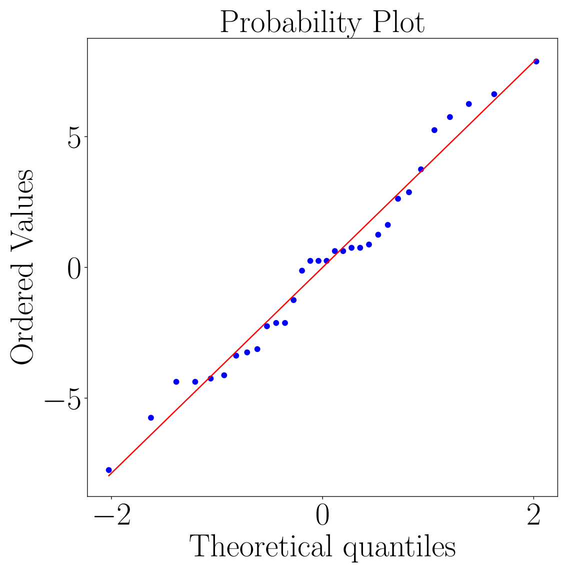


Figura 42: QQ plot of the mental demand of the sight participants on each method.

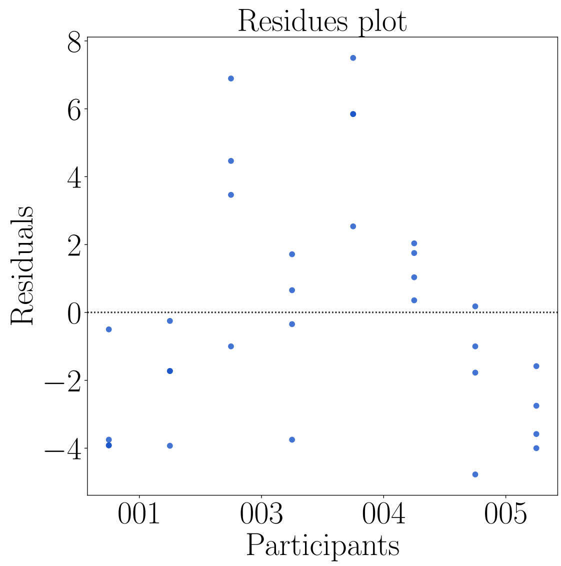


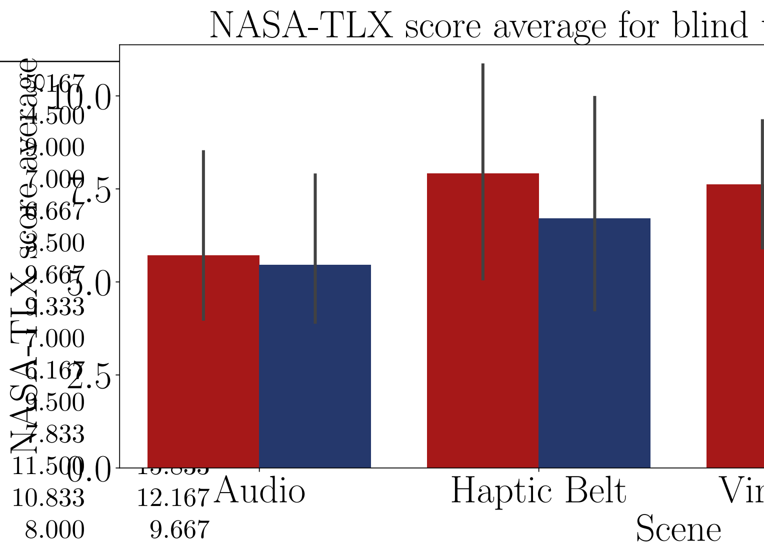
Figura 43: Residual plot of the mental demand score the sighted participants on each method.

5.2.2. Analysis of the NASA-TLX score

Table 23 brings the NASA-TLX global score of all participants, while the corresponding barplot is presented in Figure 44. As before, the higher the value, the higher is the Mental Workload of the user.

Tabela 23: NASA-TLX score felled by the participants.

Participant	Visual Condition	Round	Audio	Haptic Belt	Virtual Cane	Mixture
001C	Blind	First	4.000	8.833	[illegible]	
		Return	4.000	6.667	4.500	
002C	Blind	First	4.833	4.833	[illegible]	
		Return	4.833	4.833	7.000	
003C	Blind	First	4.000	5.333	[illegible]	
		Return	3.833	3.667	[illegible]	
004C	Blind	First	10.000	12.667	9.667	
		Return	9.167	11.667	[illegible]	
001	Sight	First	10.167	9.833	7.000	
		Return	11.000	10.833	6.167	
003	Sight	First	9.833	10.167	[illegible]	
		Return	6.667	9.667	7.833	
004	Sight	First	14.833	13.667	11.500	[illegible]
		Return	11.833	11.833	10.833	12.167
005	Sight	First	7.667	9.000	8.000	9.667
		Return	7.667	8.667	7.667	6.000

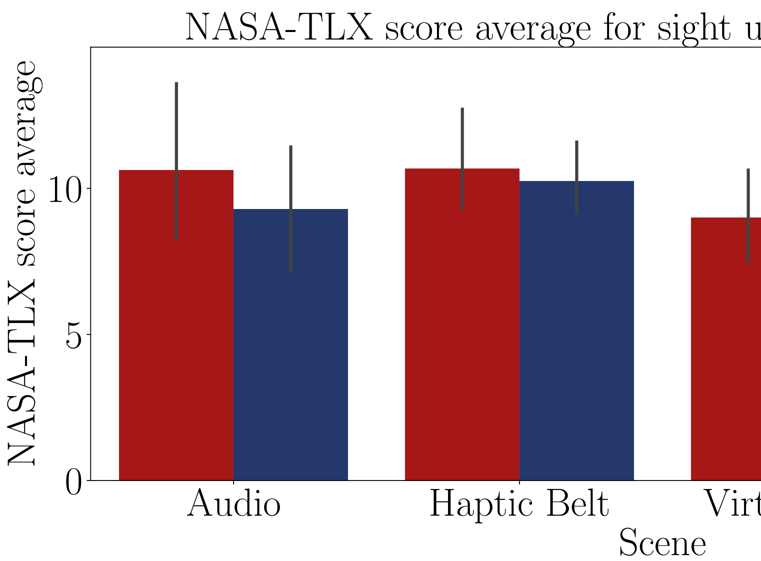


(a) Blind participants.

From Figure 44, it is possible to see that, similar to blind participants, sighted participants consider that the workload of the return round was lower than that of the first round. However, similar to what happened for the mental demand, sighted participants considered virtual cane as the methods with the lowest workload, while, for blind participants, it was the audio.

Figures 45 and 46 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.

Figures 47 and 48 bring the QQ plot and residual distribution of the data from sighted participants, showing that ANOVA can be used. The p-values for both groups are presented in Table 24. It confirms the influence of the round for both sighted and 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.



(b) Sight participants.

Figura 44: Barplot of the NASA-TLX score on each method and each round.

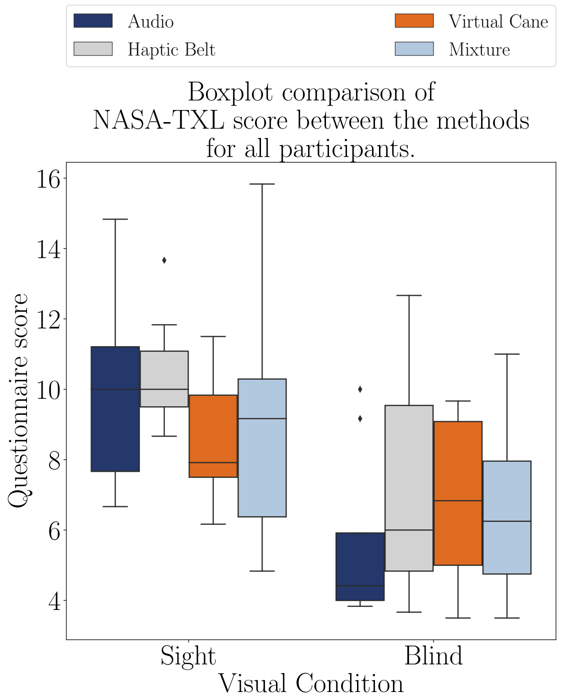


Figura 45: Boxplot of the NASA-TLX score of the participants grouped by the methods.

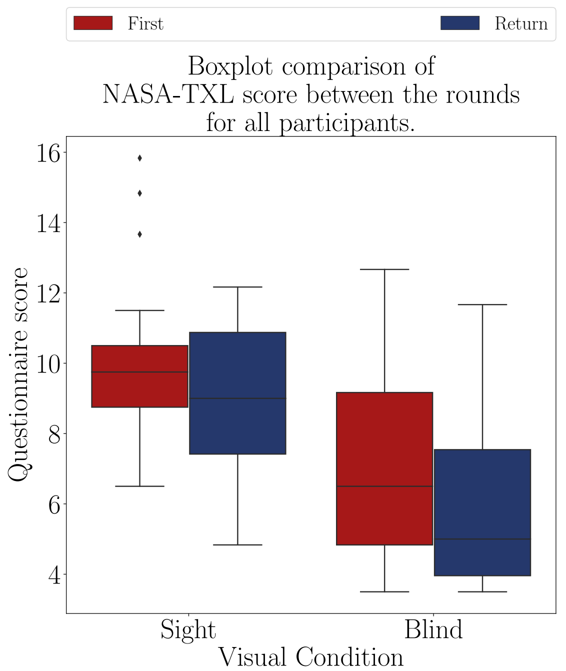


Figura 46: Boxplot of the NASA-TLX score of the participants grouped by the rounds.

Tabela 24: Anova p-value for the NASA-TLX score on each method

(a) Blind participants

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

(b) Sight participants

Source	P-Value
Methods	0.086
Rounds	0.034**
Interaction	0.688

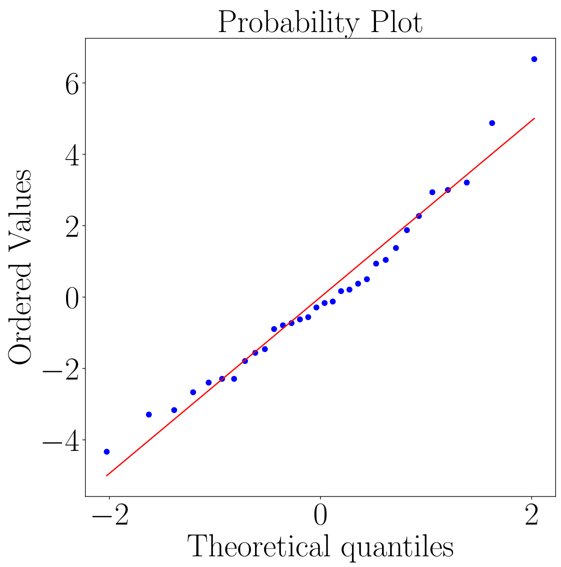


Figura 47: QQ plot of the NASA-TLX score of the sight participants on each method.

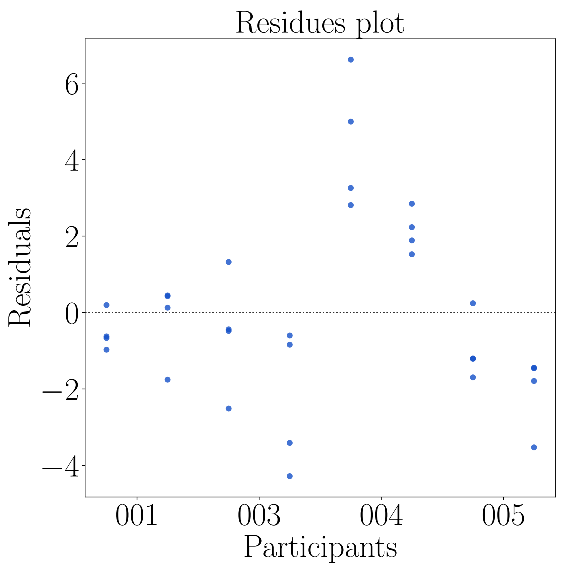


Figura 48: Residual plot of the NASA-TLX score the sight participants on each method.

5.2.3. *Adapted SAGAT*

Table 25 presents the SAGAT score of all participants. As said before, the higher the value, the higher is the Situation Awareness of the user. The corresponding barplot is presented in Figure 49.

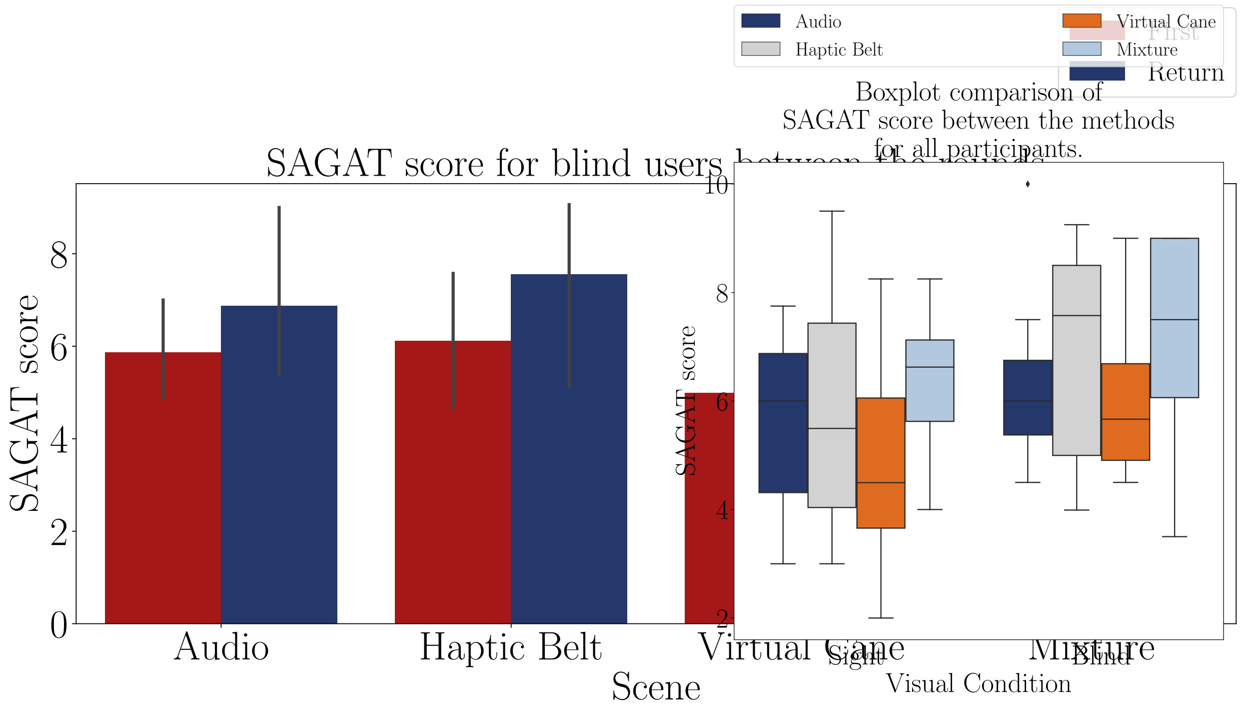
Figure 49. shows that the SAGAT score for sighted participants is, on average lower than that of blind participants, which is expected as they are not used to navigating without vision. Also, the increase in situation awareness from the first to the return round is lower. In the case of the mixture method, the SAGAT score did not improve at all. For both groups, the virtual cane was the method with the lowest score in the first round.

Tabela 25: SAGAT global score felled by the participants.

Participant	Visual Condition	Round	Audio	Haptic Belt	Virtual Cane
001C	Blind	First	5.500	5.330	5.830
		Return	6.500	8.500	5.500
002C	Blind	First	4.500	3.990	4.500
		Return	5.000	4.000	6.500
003C	Blind	First	7.500	7.490	4.660
		Return	10.000	8.500	9.000
004C	Blind	First	6.000	7.660	4.990
		Return	6.000	9.250	7.250
001	Sight	First	4.500	4.330	2.660
		Return	6.000	5.000	5.000
003	Sight	First	6.750	5.990	3.990
		Return	6.000	7.250	6.250
004	Sight	First	7.250	7.990	5.990
		Return	7.750	9.500	8.250
005	Sight	First	3.000	3.160	3.990
		Return	3.750	3.000	2.000

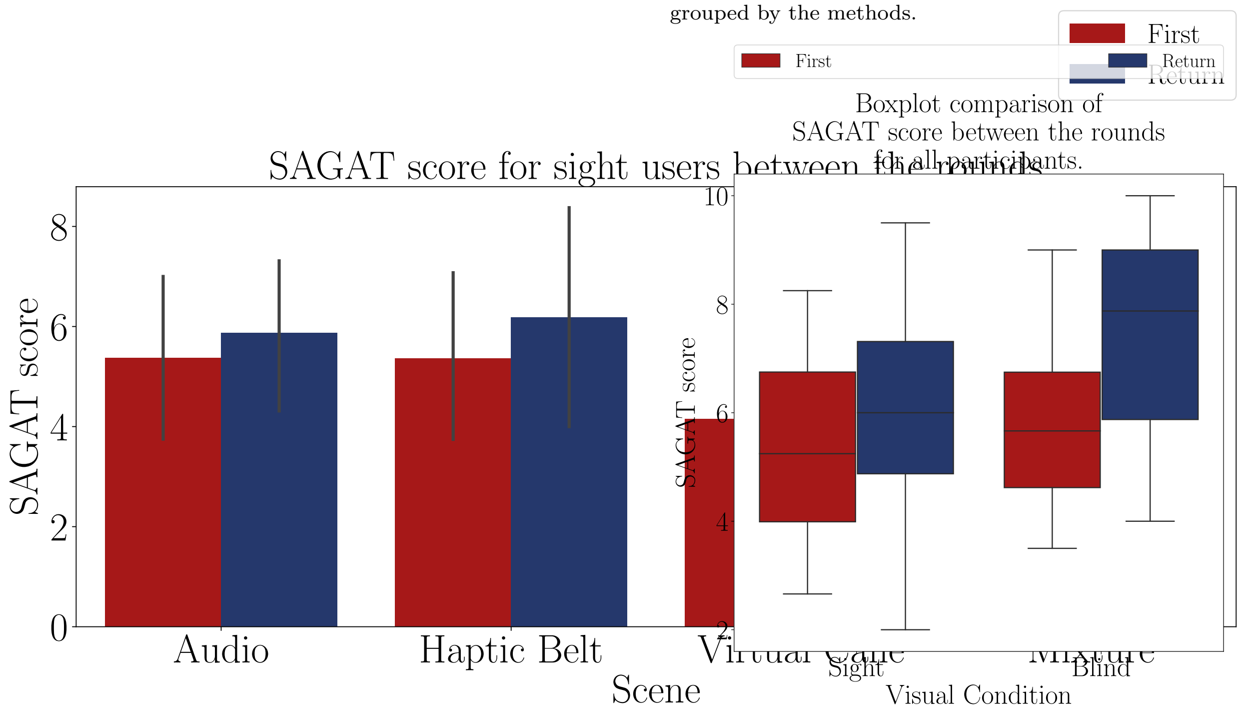
Figures 50 and 51 bring the boxplots. According to Figure 50, both groups presented a higher situation awareness with 'mixture' and 'haptic'. On the other hand, Figure 51 confirms that the difference between the rounds is more significant for blind participants.

Figures 52 and 53 brings the QQ plot and residual distribution. The variance of the residuals is not equal among the participants. Table 26 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.



(a) Blind participants.

Figura 50: Boxplot of the Sagat score of the participants grouped by the methods.



(b) Sight participants.

Figura 51: Boxplot of the Sagat score of the participants grouped by the rounds.

Figura 49: Barplot of the SAGAT score on each method and each round.

Tabela 26: Anova p-value for the SAGAT score on each method

(a) Blind participants

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

(b) Sight participants

Source	P-Value
Methods	0.035**
Rounds	0.095
Interaction	0.578

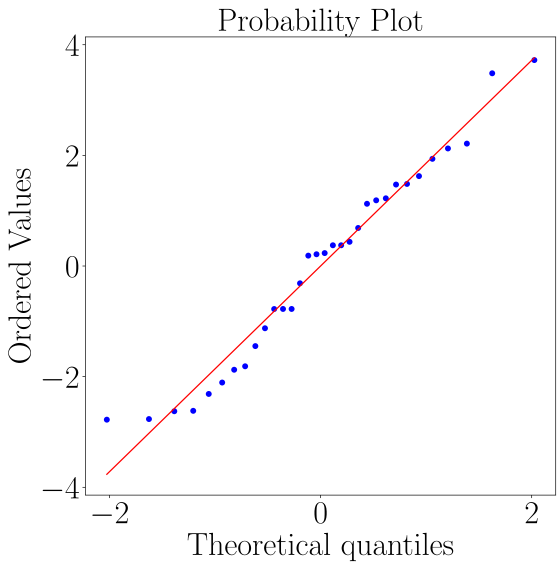


Figura 52: QQ plot of the mental demand of the sight participants on each method.

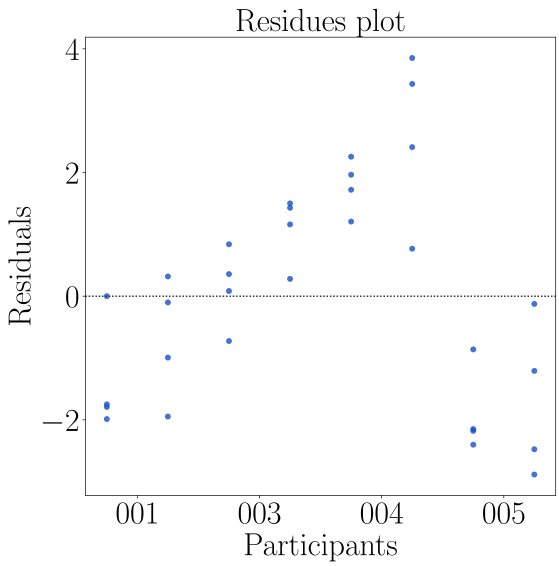


Figura 53: Residual plot of the mental demand score the sight participants on each method.

5.2.4. Guidance method's questionnaire.

As for the blind users, the sighted user also answered the Guidance questionnaire to give their thoughts about their experience with the guidance methods. Table 27 shows the score of both groups. As said before, the higher the value, the higher is the user satisfaction. The corresponding barplots are presented in Figure 54. Both groups prefer audio and mixture methods. The difference lies in the preference between the haptic belt and virtual cane. The blind users tend to prefer the first one, while the sighted users tend to prefer the last.

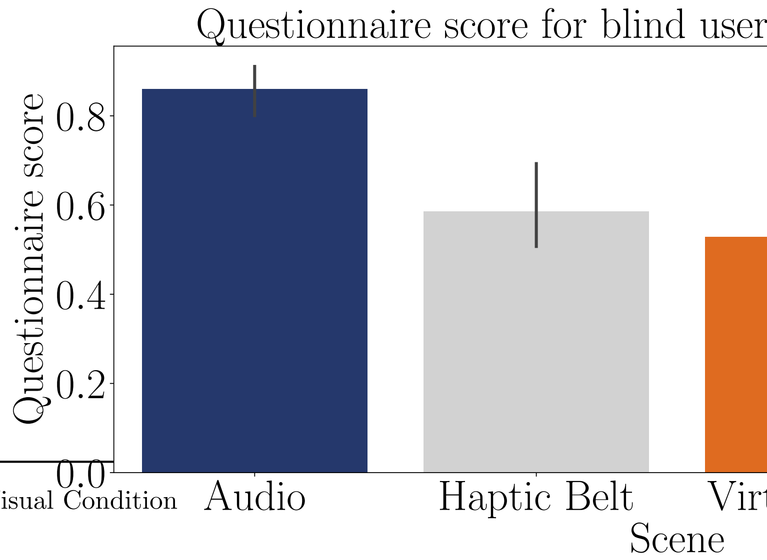
Tabela 27: Guidance method questionnaire score grouped by participant.

Participant	Audio	Haptic Belt	Virtual Cane	Mixture	Visual Condition
001C	0.77	0.54	0.63	0.87	Blind
002C	0.86	0.74	0.54	0.93	Blind
003C	0.93	0.57	0.54	0.74	Blind
004C	0.88	0.49	0.40	0.73	Blind
001	0.75	0.49	0.57	0.69	Sight
003	0.76	0.54	0.54	0.78	Sight
004	0.86	0.60	0.79	0.76	Sight
005	0.61	0.57	0.75	0.84	[illegible]

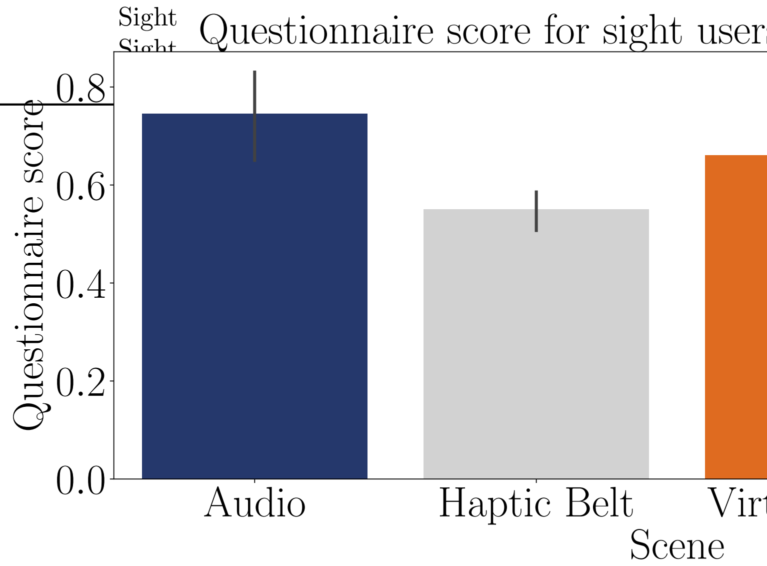
The Figure 55 presents the box plot with the distribution of the scores. It is possible to see that there is some similarity between the two groups, except for the virtual cane method, which has a broader distribution for the sighted users. Also, it seems that the audio and mixture have similar acceptance for sighted and blind users.

Figures 56 and 57 brings the QQ plot and residual distribution, which confirm that ANOVA can be applied. The result of ANOVA is presented in Table 28 and indicates that the method is an effective variable for the sighted participants, as it is for the blind ones.

Table 29 presents the conclusion of a pairwise Fisher LSD test between all the guidance methods for both groups, showing that the results are coinci-



(a) Blind participants.



(b) Sighted participants.

Figura 54: Barplot of the average questionaire score on each method.

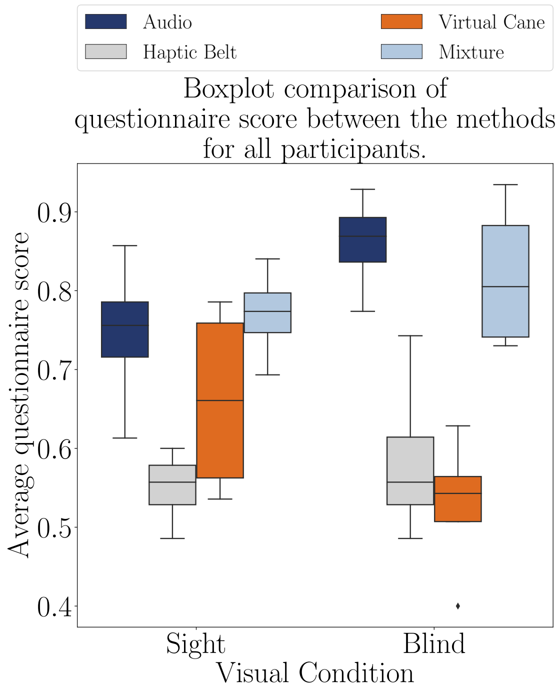


Figura 55: Boxplot of the questionaire score of the the participants grouped by the methods.

Tabela 28: Anova p-value for the questionnaire score on each method

(a) Blind participants.

Source	P-Value
Method	0.001**

(b) Sight participants.

Source	P-Value
Method	0.016**

dent.

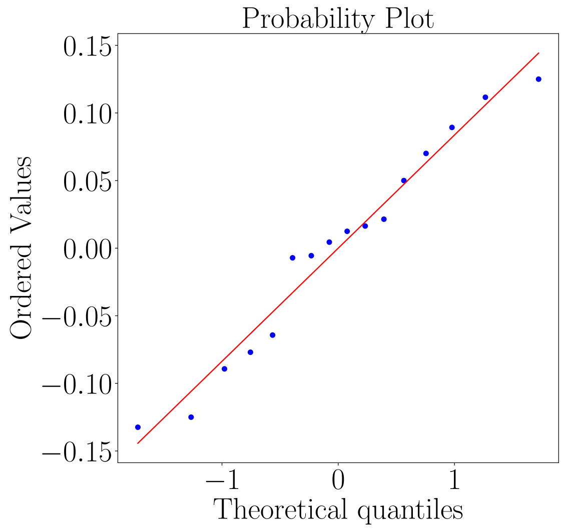


Figura 56: QQ plot of the questionnaire score of the sighted participants on each method.

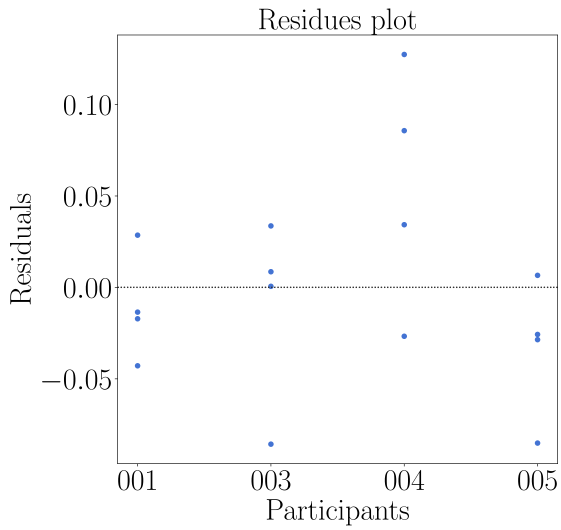


Figura 57: Residual plot of the questionnaire score the sighted participants on each method.

5.3. Physiological data

5.4. Electrocardiogram (ECG) data

Analysis of the heartbeat frequency (BPM)

Table 30 presents the average heart rate for both sighted and blind groups. If the variation between the First and the Return round is positive, it means that the user had an increase on his/her mental

Tabela 29: Anova p-value for the mental demand average on each method'

(a) Blind participants.

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

(b) Sight participants.

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

Tabela 30: ECG average BPM felled by the participants using the proposed methods [BPM].

Part	Visual Condition	Round	Audio	Haptic Belt	Virtual Cane	Mixtu
001C	Blind	First	60.71	71.17	59.07	68.
		Return	58.61	66.22	64.20	70.
002C	Blind	First	38.67	48.74	46.89	52.
		Return	47.58	58.97	56.75	58.
003C	Blind	First	69.89	70.95	69.41	66.
		Return	67.44	69.68	68.82	67.
004C	Blind	First	73.55	73.70	71.94	74.
		Return	74.79	74.02	72.69	67.
001	Sight	First	71.23	63.02	64.85	58.
		Return	73.18	61.18	66.78	66.
003	Sight	First	63.47	71.80	70.90	72.
		Return	72.75	71.23	67.49	73.
004	Sight	First	66.85	62.45	65.94	67.
		Return	69.48	65.65	64.58	71.
005	Sight	First	71.34	66.93	66.46	67.
		Return	69.57	65.97	67.00	65.

workload and vice-versa. The barplots are presented in Figure 58. 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 59 and 60.

Figures 61 and 62 show the QQ plot and residual distributions for the sighted participants of Table 30. These figures show that the data are normally distributed and that the methods have a similar variance. Table 31 brings the results from ANOVA, which are similar for both sighted and blind participants.

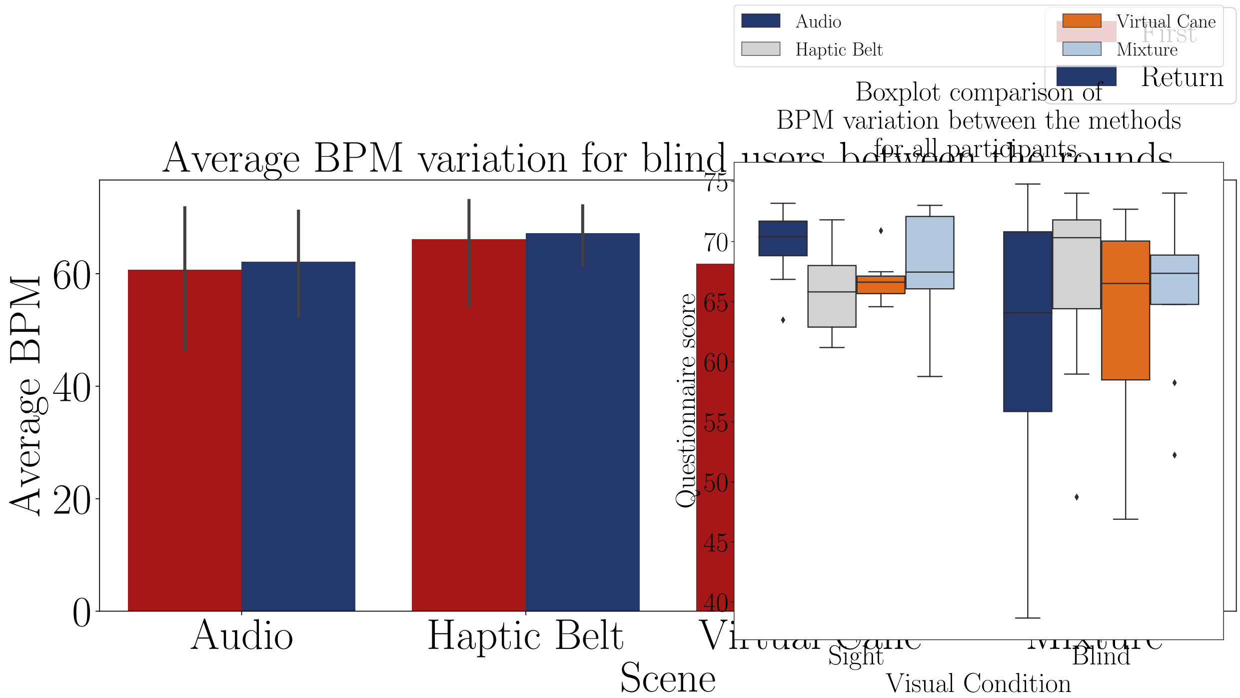
Tabela 31: Anova p-value for the BPM on each method.

(a) Blind participants

Source	P-Value
Methods	0.100
Rounds	0.371
Interaction	0.894

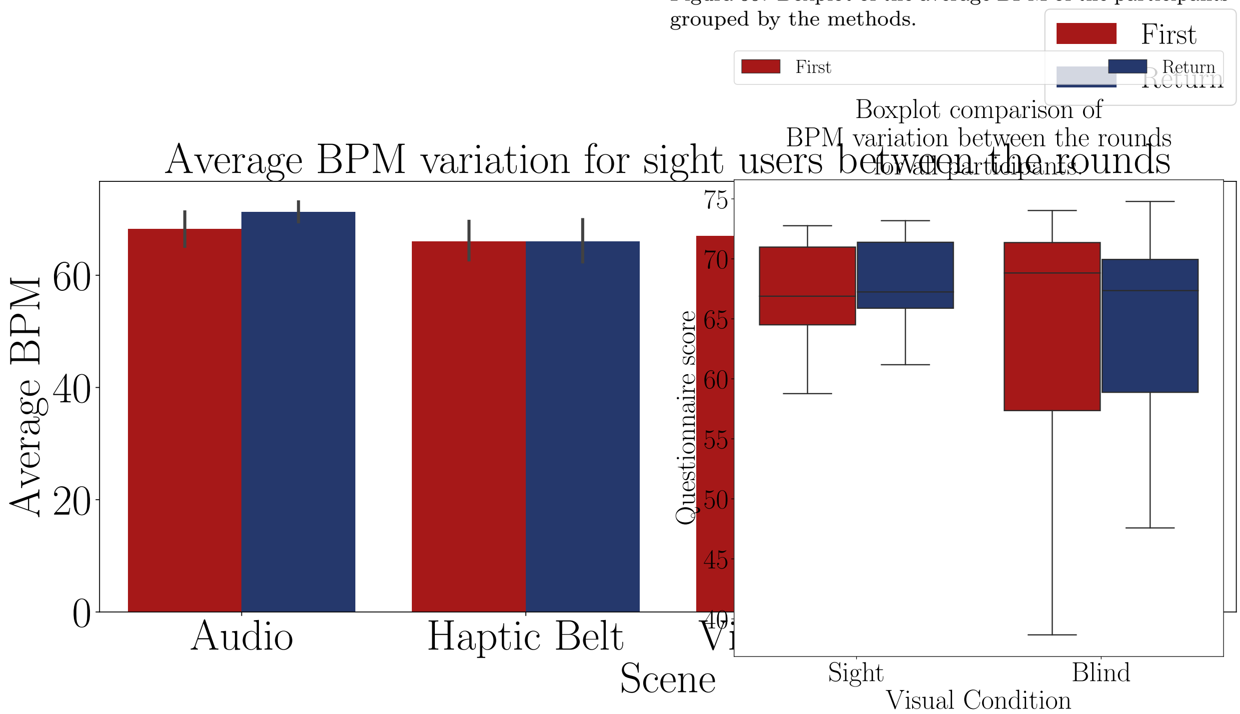
(b) Sight participants

Source	P-Value
Methods	0.166
Rounds	0.308
Interaction	0.631



(a) Blind participants.

Figura 59: Boxplot of the average BPM of the participants grouped by the methods.



(b) Sight participants.

Figura 60: Boxplot of the average BPM of the participants grouped by the rounds.

Figura 58: Barplot of the average BPM of the on each method and each round.

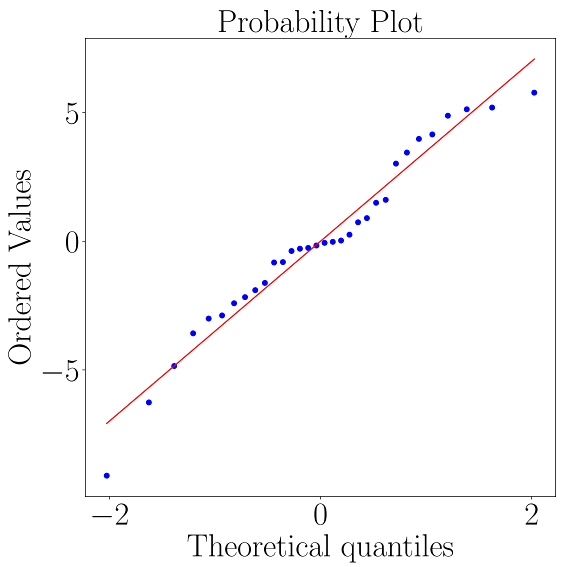


Figura 61: QQ plot of the BPM of the sight participants on each method.

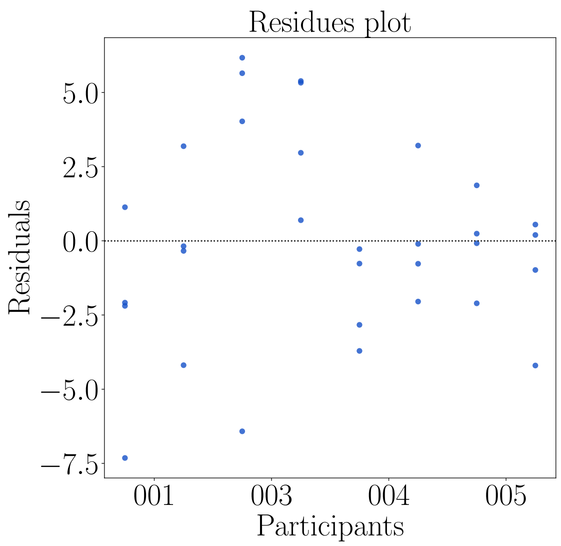


Figura 62: Residual plot of the BPM score the sight participants on each method.

Analysis of the heartbeat variance (SDNN)

Table 32 presents the SDNN for both sighted and blind participants. Different to the heart rate frequency, if the variation between the First and the Return round is negative, it means that the user had an increase on his/her mental workload and vice-versa. The mean values are presented in the barplots of Figure 63.

Tabela 32: Average SDNN by the participants during the each round and method [ms].

Part.	Visual Condition	Round	Audio	Haptic Belt	Virtual Cane	Mix
001C	Blind	First	107.061	124.737	163.968	129
		Return	130.885	131.590	157.589	124
002C	Blind	First	98.863	81.140	33.977	79
		Return	49.627	42.815	114.057	107
003C	Blind	First	38.325	35.101	42.392	43
		Return	41.196	44.256	42.602	46
004C	Blind	First	86.827	62.560	85.900	70
		Return	74.895	70.017	66.089	104
001	Sight	First	82.185	134.530	134.773	225
		Return	69.479	318.747	116.003	136
003	Sight	First	79.600	51.782	68.676	60
		Return	45.709	40.927	66.323	47
004	Sight	First	121.130	154.718	128.477	125
		Return	100.366	122.563	140.115	119
005	Sight	First	87.686	120.522	88.591	102
		Return	93.207	122.839	141.305	96

No clear pattern is evident from this figure. For some methods, the return round resulted in a decrease in the SDNN, while for others, it increased.

Figures 64 and 65 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.

Figures 66 and 67 bring the QQ Plot and residual distribution. Figure 66 hints that the data from sighted users contain two outliers. Table 33 shows the ANOVA test p-values. For both groups, none of the factors have a significant influence on the SDNN value.

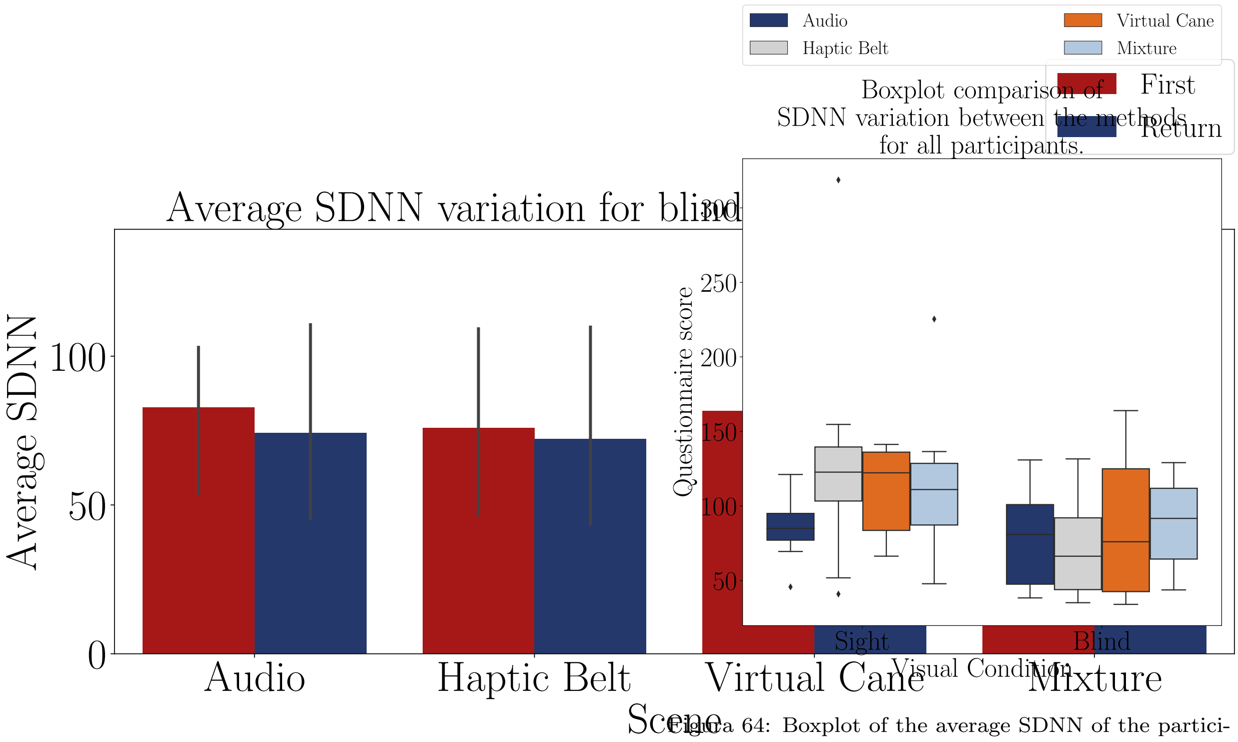


Figura 64: Boxplot of the average SDNN of the participants grouped by the methods.

(a) Blind participants.

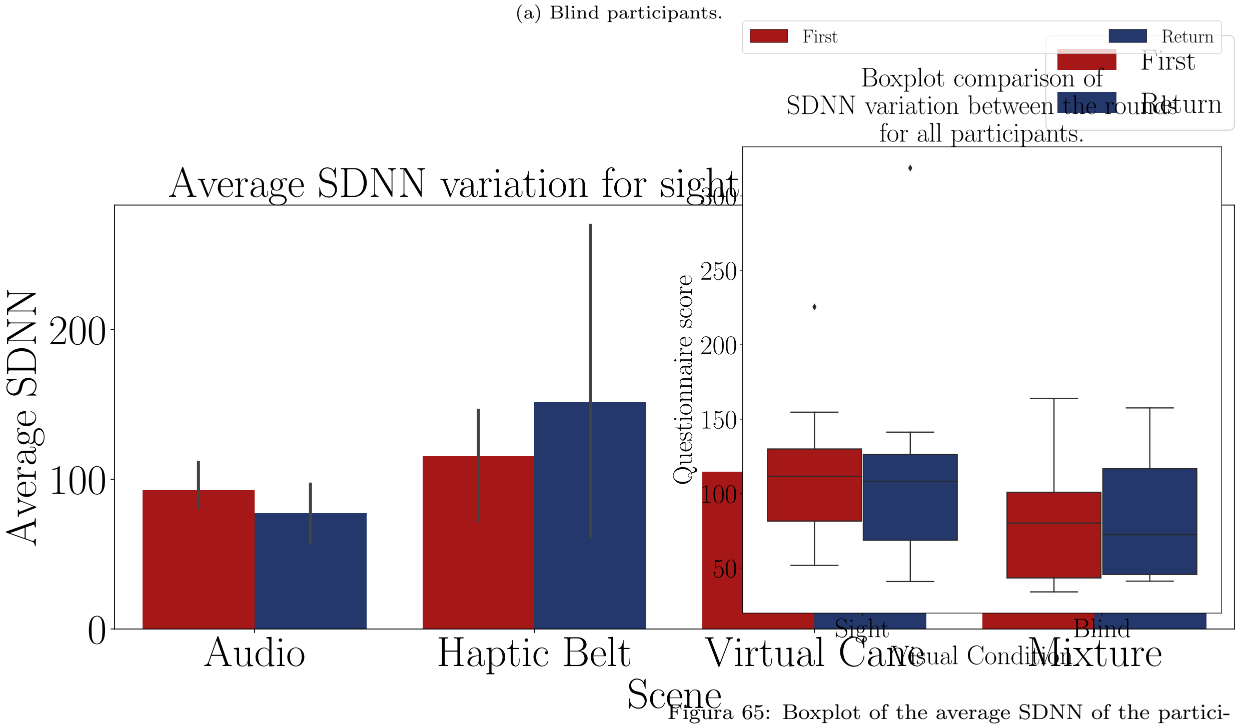


Figura 65: Boxplot of the average SDNN of the participants grouped by the rounds.

(b) Sight participants.

Figura 63: Barplot of the average SDNN of the on each method and round.

Tabela 33: Anova p-value for the average SDNN on each method.'

(a) Blind participants

Source	P-Value
Methods	0.486
Rounds	0.223
Interaction	0.473

(b) Sight participants

Source	P-Value
Methods	0.189
Rounds	0.969
Interaction	0.455

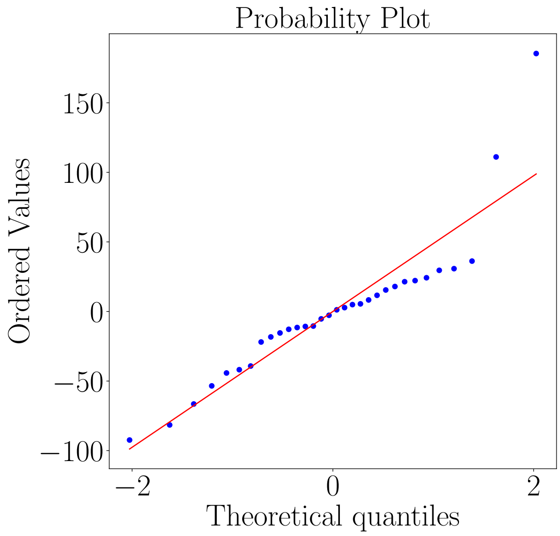


Figura 66: QQ plot of the average SDNN of the sight participants on each method.

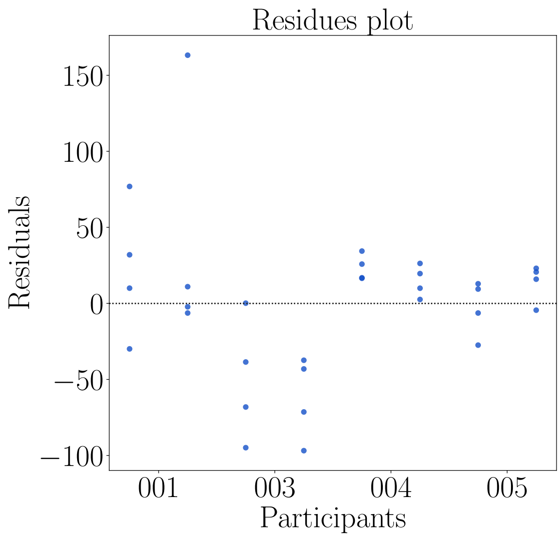


Figura 67: Residual plot of the average SDNN score the sight participants on each method.

5.4.1. Galvanic skin response and temperature data;

Table 34 presents the average skin conductance for both groups, while the percentual variation related to the baseline is presented in Table 19.

Tabela 34: Average GSR felled by the participants [μS].

Participant	Visual Condition	Round	Baseline	Audio	Haptic Belt	Virtual Cane	Mixture
001C	Blind	First	0.37	1.03	3.14	3.79	3.90
		Return		1.58	2.81	4.04	4.57
003C	Blind	First	0.30	0.56	0.62	0.85	1.09
		Return		0.63	0.65	0.92	1.06
004C	Blind	First	1.24	3.07	3.49	2.28	2.23
		Return		2.95	3.20	2.21	2.24
001	Sight	First	4.27	15.19	15.67	15.19	14.13
		Return		14.95	15.09	15.72	21.52
004	Sight	First	2.60	11.18	12.60	12.92	10.34
		Return		11.97	12.25	13.47	10.16
005	Sight	First	0.47	1.58	1.44	1.37	1.33
		Return		1.53	1.47	1.49	1.33

Tabela 35: Average GSR variation in relation to the baseline in each round [μS].

Participant	Visual Condition	Round	Audio	Haptic Belt	Virtual Cane	Mixture
001C	Blind	First	176.54%	746.10%	920.72%	951.71%
		Return	327.42%	656.99%	988.93%	1132.39%
003C	Blind	First	84.23%	104.19%	182.35%	258.80%
		Return	109.23%	112.95%	202.35%	249.72%
004C	Blind	First	148.53%	182.84%	84.33%	80.69%
		Return	138.64%	159.00%	78.73%	81.61%
001	Sight	First	255.76%	266.93%	255.69%	231.52%
		Return	250.18%	253.32%	268.25%	403.90%
004	Sight	First	329.08%	383.54%	395.83%	297.05%
		Return	359.53%	370.35%	417.17%	289.96%
005	Sight	First	239.16%	207.74%	193.85%	184.71%
		Return	227.06%	214.91%	219.59%	185.86%

The barplots of the two groups are presented in Figure 68. 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 64 and 65.

Figures 71 and 72 bring the QQ Plot and residual distribution. The results from ANOVA are presented in Table 36. 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 sighted participants, where the p-value of the method factor is the highest and well above the 0.05 threshold.

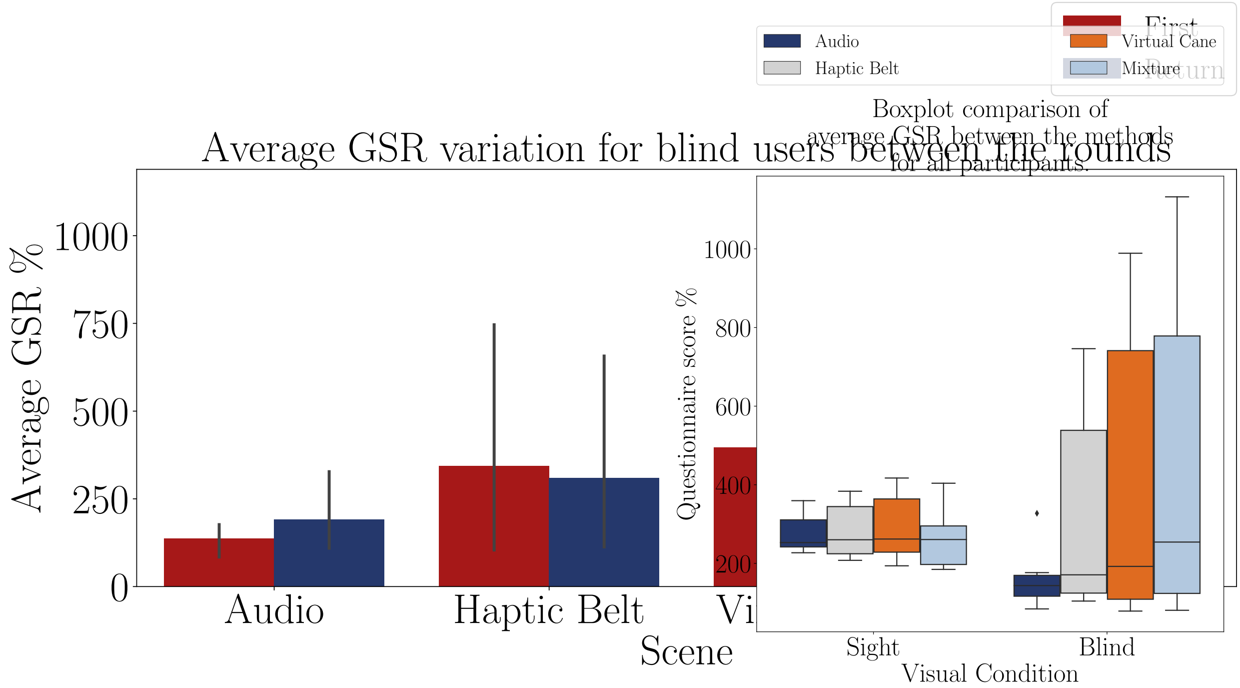
Tabela 36: Anova p-value for the skin conductance average on each method

(a) Blind participants

Source	P-Value
Methods	0.051
Rounds	0.722
Interaction	0.996

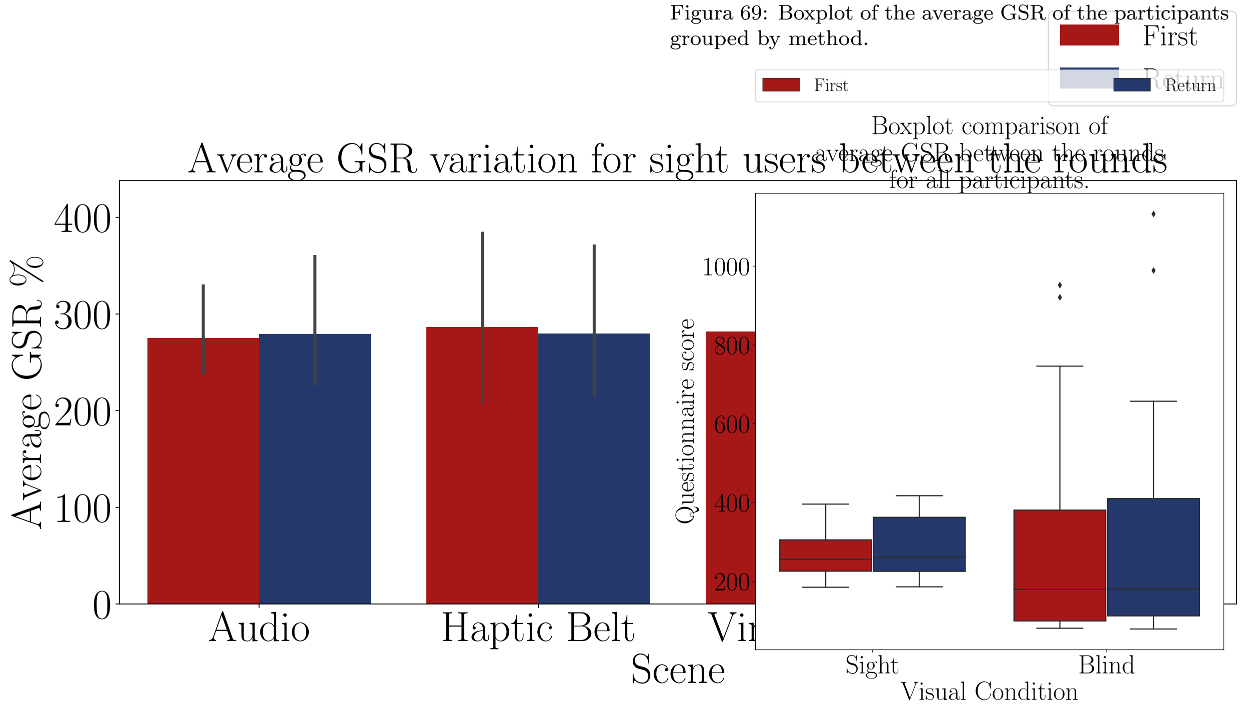
(b) Sight participants

Source	P-Value
Methods	0.802
Rounds	0.354
Interaction	0.686



(a) Blind participants.

Figura 69: Boxplot of the average GSR of the participants grouped by method.



(b) Sight participants.

Figura 70: Boxplot of the average GSR of the participants grouped by round.

Figura 68: Barplot of the average GSR on each method and round.

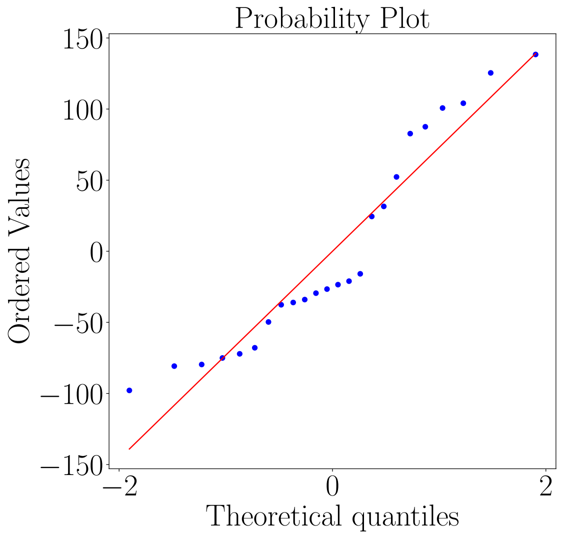


Figura 71: QQ plot of the average skin conductance of the sight participants on each method.

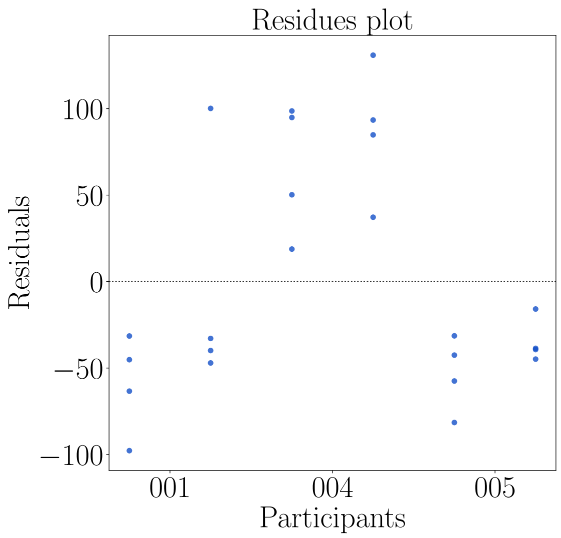


Figura 72: Residual plot of the average skin conductance score the sight participants on each method.

5.5. Final Remarks

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.

Referências

World report on vision. Technical report, 2019.

Rupert Bourne, Jaimie D Steinmetz, Seth Flaxman, Paul Svitil Briant, Hugh R Taylor, Serge Resnikoff, Robert James Casson, Amir Abdoli, Eman Abu-Gharbieh, Ashkan Afshin, et al. 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(2):e130–e143, 2021.

Nicholas A Bradley and Mark D Dunlop. 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, 2002.

Nicholas A Bradley and Mark D Dunlop. An experimental investigation into wayfinding directions for visually impaired people. Personal and Ubiquitous Computing, 9(6): 395–403, 2005.

Mao-Lin Chiu. An organizational view of design communication in design collaboration. Design studies, 23(2):187–210, 2002.

Mica R Endsley. Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society annual meeting, volume 32, pages 97–101. Sage Publications Sage CA: Los Angeles, CA, May 1988.

Mica R Endsley. Measurement of situation awareness in dynamic systems. Human factors, 37(1):65–84, 1995.

Mica R Endsley. Automation and situation awareness. In Automation and human performance: Theory and applications, pages 163–181. CRC Press, 2018.

Wendy A Farrell. Learning becomes doing: Applying augmented and virtual reality to improve performance. Performance Improvement, 57(4):19–28, 2018.

Claudio Kirner, Tereza Gonçalves Kirner, Roberto Sussumu Wataya, and José Armando Valente. Using augmented reality to support the understanding of three-dimensional concepts by blind people. 2011.

Cecil A Lozano, Kurt A Kaczmarek, and Marco Santello. Electrotactile stimulation on the tongue: Intensity perception, discrimination, and cross-modality estimation. Somatosensory & motor research, 26(2-3):50–63, 2009.

Ivana Moerland-Masic, Fabian Reimer, Thomas M Bock, Frank Meller, and Björn Nagel. Application of vr technology in the aircraft cabin design process. CEAS Aeronautical Journal, pages 1–10, 2021.

Mark S Sanders and Ernest James McCormick. Human factors in engineering and design. Industrial Robot: An International Journal, 1998.

Alexa F Siu, Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. 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, pages 1–13, 2020.

Neville Anthony Stanton, Alan Hedge, Karel Brookhuis, Eduardo Salas, and Hal W Hendrick. Handbook of human factors and ergonomics methods. CRC press, 2004.

Alexander Wolf, Nicole Binder, Jörg Miehling, and Sandro

Wartzack. 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, volume 1, pages 4029–4038. Cambridge University Press, 2019.