VIRTUAL REALITY FOR THE HUMAN-CENTRED DESIGN OF ASSISTIVE DEVICES

Ivan de Souza Rehder¹, Emília Villani¹, Edmar Thomaz da Silva¹

¹Instituto Tecnológico de Aeronáutica, 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 TSens, are used. Among them, are an electrocardiogram sensor (ECG), to gather heart-rate and heartrate 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: virtual reality, human factors, human machine interface.

# Introduction

According to the World Health Organisation (WHO), on the **“**World report on vision” from 2019, there are at least 2.2 billion people with some visual impairment degre**e**. 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,amongothers **(**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 fulfil 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, Miehling 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 assistivedevices 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, 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 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 usertailored 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,butseveraldifferenttechniqueshavebeenproposed 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,orcollaborativedesign,referstoadesignprocess 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?

The concepts of assistive devices presented as part ofthis work are used only as examples for investigating theresearch questions presented. The challenges related to theirfull 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 analyses the results in order to investigate the research questions and Section 4 discusses those results.

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

# 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.

# 

Figure 1: Method's diagram

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

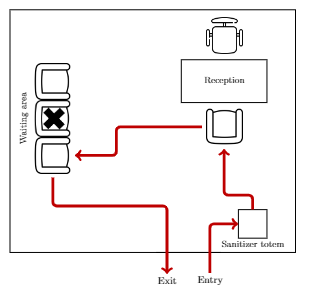


Figure 2: Scheduled task of the experiment and their order

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 3 and Figure 4 show the virtual environment created in Unity3D and the corresponding real environment assembled at the CCM entrance hall.

|  |  |
| --- | --- |
| Figure 3: Real environment photo | Figure 4: Virtual environment screenshot |

## Human factors techniques

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.

1. Situation awareness;

A modified SAGAT (Situation Awareness Global Assessment Technique) questionnaire is used to evaluate the BVI situation awareness.

As original idea from \cite{endsley1988design}, the proposed version is based on 3 levels of situation awareness:

* 1. Perception:

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

* 1. Comprehension:

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

* 1. 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

1. 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.

1. 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.

1. 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.

1. 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.

# Results

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

* 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: %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: %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):

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):

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

### 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 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".

|  |  |
| --- | --- |
| Figure 5: Boxplot of the mental demand of the blind participants grouped by the methods. | Figure 6: Boxplot of the mental demand of the blind participants grouped by the rounds. |

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.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: ANOVA p-value for mental demand – blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.170 | | Rounds | 0.075 | | Interaction | 0.993 | |

#### Analysis of the NASA-TLX score

Figure 7 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. Boxplot of the NASA-TLX score of the blind participants grouped by the rounds. presents a boxplot with the NASA-TLX global score grouped by the rounds, showing that the two groups are still different.

|  |  |
| --- | --- |
| Figure 7: Boxplot of the NASA-TLX score of the blind participants grouped by the methods. | Figure 8: Boxplot of the NASA-TLX score of the blind participants grouped by the rounds. |

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.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: ANOVA p-value for NASA-TLX score – blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.029\*\* | | Rounds | 0.022\*\* | | Interaction | 0.814 | |

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.

### 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 9 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 10 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 9: Boxplot of the SAGAT score of the blind participants grouped by the methods. | Figure 10: Boxplot of the SAGAT score of the blind participants grouped by the rounds. |

Table 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.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: ANOVA p-value for SAGAT score – blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.277 | | Rounds | 0.002\*\* | | Interaction | 0.834 | |

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

|  |
| --- |
| Figure 11: Boxplot of the questionaire score of the blind participants grouped by the methods. |

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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 4: ANOVA p-value for questionnaire score – blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.001\*\* | |

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.

### 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)

Figure 12 and Figure 13 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 12: Boxplot of the BPM of the blind participants grouped by the methods. | Figure 13: 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 significant influence of either the guidance methods or the rounds in the participants' heart rate.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 5: ANOVA p-value for BPM – blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.100 | | Rounds | 0.371 | | Interaction | 0.894 | |

#### Analysis of the heartbeat variance (SDNN)

Figure 14 and Figure 15 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 14: Boxplot of the SDNN of the blind participants grouped by the methods. | Figure 15: Boxplot of the SDNN of the blind participants grouped by the rounds. |

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.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 6: ANOVA p-value for average SDNN– blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.486 | | Rounds | 0.223 | | Interaction | 0.473 | |

### 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.

Figure 16 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 17 presents the GSR grouped by the rounds. In this case, there is no apparent difference between the rounds.

|  |  |
| --- | --- |
| Figure 16: Boxplot of the GSR of the blind participants grouped by the methods. | Figure 17: Boxplot of the GSR of the blind participants grouped by the rounds. |

Table 7 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.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 7: ANOVA p-value for average GSR– blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.051 | | Rounds | 0.722 | | Interaction | 0.996 | |

## 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.

#### Analysis of the mental demand scale

Figure 18 and Figure 19 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.

|  |  |
| --- | --- |
| Figure 18: Boxplot of the mental demand of the participants grouped by the methods. | Figure 19: Boxplot of the mental demand of the participants grouped by the rounds. |

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 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.

Table 8: Anova p-value for the mental demand average on each methods.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (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 | |

#### Analysis of the NASA-TLX score

Figure 20 And Figure 21: Boxplot of the mental demand of the participants grouped by the rounds.Figure 21 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.

|  |  |
| --- | --- |
| Figure 20: Boxplot of the mental demand of the participants grouped by the methods. | Figure 21: Boxplot of the mental demand of the participants grouped by the rounds. |

The p-values for both groups are presented in Table 9It 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.

Table 9: Anova p-value for the NASA-TLX score 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 | |

### Adapted SAGAT

Figure 22 and Figure 23 bring the boxplots. According to Figure 22 both groups presented a higher situation awareness with ‘mixture’ and ‘haptic’. On the other hand, Figure 23 confirms that the difference between the rounds is more significant for blind participants.

|  |  |
| --- | --- |
| Figure 22: Boxplot of the SAGAT score of the participants grouped by the methods. | Figure 23: Boxplot of the SAGAT score of the participants grouped by the rounds. |

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.

Table 10: Anova p-value for the SAGAT average on each methods.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (a): Blind participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.227 | | Rounds | 0.002\*\* | | Interaction | 0.834 | | (b): Sight Participants   |  |  | | --- | --- | | Source | p-value | | Methods | 0.086 | | Rounds | 0.034\*\* | | Interaction | 0.688 | |

# Acknowledgements

This work was supported by CAPES.

# References

1. ALEXANDER, T. et al. Enhancing Human Effectiveness through Embedded Virtual Simulation. I/ITSEC Interservice/Industry Training, Simulation, and Education Conference, 2012.
2. ARTHUR, J. J. et al. A review of head-worn display research at NASA Langley Research Center. SPIE Defense + Security. Baltimore: SPIE, 2015.
3. ARTHUR, J. J. et al. A review of head-worn display for the Next Generation Air Transportation System. SPIE Optical Engineering, Baltimore, v. 56, 2017.
4. AUER, S. et al. Comparison Between Virtual Reality and Physical Flight Simulators for Cockpit Familiarization. MundC: Mensch und Computer, 2021.
5. BCAA. Aviation Safety Information Leaflet Formation Flying. Belgian Civil Aviation Authority, 2016.
6. BELL, H. H.; WAAG, W. L. Evaluating the Effectiveness of Flight Simulators for Training Combat Skills: A Review. THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, v. 8, n. 3, p. 223 - 242, 1998. https://doi.org/10.1207/s15327108ijap0803\_4.
7. BLUME, B. D. et al. Transfer of Training: A Meta-Analytic Review. Journal of Management, 2010.
8. BOUCSEIN, W. Electrodermal Activity. 2ª. ed. New York: Springer, v. 1, 2012.
9. BRAITHWAITE, J. J. et al. A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. University of Birmingham. Birmingham, p. 43. 2015.
10. BÜRKI-COHEN, J.; GO, T. H. The Effect of Simulator Motion Cues on Initial Training of Airline Pilots. AIAA Modeling and Simulation Technologies Conference and Exhibit. San Francisco: AIAA. 2005. p. 12.
11. BÜRKI-COHEN, J.; SPARKO, A. L.; BELLMAN, M. Flight Simulator Motion Literature Pertinent to Airline-Pilot Recurrent Training and Evaluation. AIAA Modeling and Simulation Technologies Conference. Portland: AIAA. 2011. p. 17.
12. DICKSTEIN, C. Air Force instructor, student pilot killed in T-38 crash at Vance Air Force Base. Stars and Stripes, 2019.
13. FAA. "Flight Simulation Training Device Initial and Continuing Qualification and Use," 14 CFR Part 60, NSP Consolidate Version 2016.
14. FOOTE, B.; MELZER, J. A history of helmet mounted displays. SPIE Defense + Security. Baltimore: SPIE. 2015.
15. GAWRON, V. J. Human Performance, Workload, and Situational Awareness Measures Handbook, 2008.
16. HART, S. G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, 2006. p. 904-908.
17. HART, S. G.; STAVELAND, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human Mental Workload, 1988.
18. HAYS, R. T. et al. Flight Simulator Training Effectiveness: A Meta-Analysis. Military Psychology, v. 4, p. 63-74, 1992.
19. HIDALGO-MUÑOZ, A. R. et al. Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation, 2018.
20. JOHANNES, B. et al. Psychophysiological Assessment in Pilots Performing Challenging Simulated and Real Flight Maneuvers. Aerospace Medicine and Human Performance, v. 88, n. 9, p. 834 - 840, 2017.
21. KRIJN, R.; WEDZINGA, G. Development and in-flight demonstration of 'E-CATS', an experimental embedded training system for fighter aircraft. National Aerospace Laboratory NLR. p. 25. 2004. (NLR-TP-2004-393).
22. LABEDAN, P. et al. Virtual Reality for Pilot Training: Study of Cardiac Activity. n Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2021.
23. LEE, A. T. Flight Simulation: Virtual Environments in Aviation. Nova York: Taylor & Francis Group, v. 1, 2016.
24. LEINO, T. Neuroendocrine Responses to Psychological Workload of Military Flying. Oulo: Oulu University Library, 1999.
25. LEMMERS, A. J. J. An Embedded Training Multi-Ship Demonstrator. National Aerospace Laboratory NLR. p. 17. 2008. (NLR-TP-2008-005).
26. MACFARLANE, P. W. et al. Comprehensive Electrocardiology, 2011.
27. MAGEE, L.; SOTTILARE, R.; ROESSINGH, J. J. Human Interaction in Embedded Virtual Simulations. ITEC. Cologne: ITEC. 2011.
28. MANSIKKA, et al. Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. Applied Ergonomics, 2016.
29. MARINESCU, A. et al. Physiological Parameter Response to Variation of Mental Workload. Human Factors, p. 31 - 56, 2018.
30. MCCLERNON, C. K. Stress Training Efficacy in an Aviation Context. International Symposium on Aviation Psychology, 2009.
31. MCCLERNON, C. K. et al. Stress Training Improves Performance During a Stressful Flight. Human Factors, v. 53, n. 3, p. 207-218, 2011.
32. MCSPADDEN, R. What Went Wrong? Formation Flight Turns Tragic. AOPA, 2019.
33. ORITZ, G. A. Transfer of Learning Effectiveness: PC-Based Flight Simulation. Journal of Aviation/Aerospace Education & Research, v. 3, n. 2, p. 29-33, 1993.
34. REID, G. B. Training Transfer of a Formation Flight Trainer. Human Factors, 1975. 470-476.
35. ROBINSON, A. D.; BICANIC, N. Methods and Systems to Allow Real Pilots in Real Aircraft Using Augmented and Virtual Reality to Meet in a Virtual Piece of Airspace. US 2019/0333396, 2019.
36. ROESSINGH, J. J. M. Transfer of Manual Flying Skills From PC-Based Simulation to Actual Flight—Comparison of In-Flight Measured Data and Instructor Ratings. THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, v. 15 (1), p. 67 - 90, 2005.
37. ROESSINGH, J. J. M.; VAN SIJLL, M. C.; JOHNSON, S. P. Embedded Training - An explorative study providing requirements for the display of virtual targets on a Helmet Mounted Display in simulated air-to-air engagements within visual range, Amsterdam, p. 54, 2003. ISSN NLR-TP-2003-262.
38. ROESSINGH, J. J. M.; VERHAAF, G. G. Training Effectiveness of Embedded Training in a (Multi-) Fighter Environment A Discussion Paper. National Aerospace Laboratory NLR. p. 8. 2009.
39. SHAFFER, F.; GINSBERG, J. P. An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health, vol. 28, 2017.
40. TASK GROUP HFM-165. Improving Human Effectiveness Through Embedded Virtual Simulation. OTAN, 2014.
41. VALVERDE, H. H. A Review of Flight Simulator Transfer of Training Studies. Human Factors, Ohio, v. 15, n. 6, p. 512 - 522, 1973. ISSN https://doi.org/10.1177/001872087301500603.
42. VELTMAN, J. A. Comparative Study of Psychophysiological Reactions During Simulator and Real Flight. The International Journal of Aviation Psychology, v. 12, p. 33 - 48, 2002.
43. WEDZINGA, G. E-CATS: First time demonstration of embedded training in a combat aircraft. Aerospace Science and Technology, v. 10, p. 73-84, 2005.
44. WOOD, M. E. et al. Design of the Simplified Formation Trainer. Defense Technical Information Center. Ohio, p. 36. 1972. (AD0754973).