Computers & Graphics 117 (2023) 1–12

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on RAGI 2023

Exploring the user experience of hands-free VR interaction methods during a Fitts' task

Pedro Monteiro *, Hugo Coelho, Guilherme Gonçalves, Miguel Melo, Maximino Bessa

University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
INESC TEC, 4200-465 Porto, Portugal

A R T I C L E I N F O

Keywords:
Hands-free
HCI
Immersive Virtual Reality
Interaction
Usability

A B S T R A C T

Despite advancements in interaction with immersive Virtual Reality (VR) systems, using hand gestures for all interactions still imposes some challenges, especially in interactions with graphical user interfaces that are usually performed with point-and-click interfaces. Therefore, exploring the use of alternative hands-free methods for selection is essential to overcome usability problems and provide natural interaction for users. The results and insights gained from this exploration can lead to enhanced user experiences in VR applications. This study aims to contribute to the literature with the evaluation of the usability of the most commonly used hands-free methods for selection and system control tasks in immersive VR and their impact on standard and validated experience and usability metrics, namely the sense of presence, cybersickness, system usability, workload, and user satisfaction. A Fitts' selection task was performed using a within-subjects design by nine participants experienced in VR. The methods evaluated were the handheld controllers, the head gaze, eye gaze, and voice commands for pointing at the targets, and dwell time and voice commands to confirm the selections. Results show that the methods provide similar levels of sense of presence and low cybersickness while showing low workload values and high user satisfaction, matching the experience of traditional handheld controllers for non-multimodal approaches. The assisted eye gaze with dwell was the preferred hands-free method and the one with the highest values of usability. Still, developers should minimize the number of gaze movements to reduce fatigue. The evaluation also showed that using a multimodal approach for selections, especially using the voice, decreases user satisfaction and increases users' frustration.

1. Introduction

Immersive virtual reality (VR) is a type of VR experience that uses an immersive display such as a head-mounted display (HMD) and allows users to navigate and interact with a 3D virtual environment in real-time [1]. This form of VR provides a more realistic experience than non-immersive VR, which is usually delivered via a computer screen and requires interaction through traditional input devices [2].

In immersive VR, handheld controllers such as those included in standard commercial VR products are used to enable interactions with virtual environments and objects [2] by providing a proxy to users' hands inside those environments, enhancing the sense of presence: the feeling of actually being present in the virtual world despite being aware of its artificial nature [3].

The combination of an HMD and handheld controllers represents the most common VR setup to achieve a high level of immersion and interactivity [4]. Consequently, it has found extensive applications in entertainment, education, industry, research, and even clinical fields [5].

Continuous advances in tracking systems have allowed the reliable use of hand gestures for VR interactions [6]. Hand gestures provide a natural and intuitive way for users to interact with virtual environments and other users, allowing them to manipulate objects, perform actions that resemble real-world interactions, and increase the sense of presence and immersion [7].

Despite the increasing use of hand gestures, some challenges with their use can lead to a negative user experience in VR applications. For instance, the tracking systems can have issues that result in a lack of precision of the hand tracking, erroneous segmentation of the hands, and ultimately, a wrong interpretation of the gestures [8]. Using hand gestures can also hinder using hands for other tasks that may be performed simultaneously, such as directly interacting with objects. Overloading the use of the hands can increase cognitive load and physical fatigue [8], especially when performing gestures for tasks or on interfaces that are usually interacted through point-and-click methods (i.e., graphical user interfaces (GUI) with menus). Thus, because sometimes hand gestures can have suboptimal usability, it is valuable

* Corresponding author at: University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal.
E-mail addresses: monteiro.p@outlook.pt (P. Monteiro), hcoelho@utad.pt (H. Coelho), guilhermeg@utad.pt (G. Gonçalves), mcmelo@inesctec.pt (M. Melo), maxbessa@utad.pt (M. Bessa).

https://doi.org/10.1016/j.cag.2023.10.005
Received 18 July 2023; Received in revised form 9 October 2023; Accepted 10 October 2023
Available online 13 October 2023

to explore alternative hands-free methods (methods that do not use the hands) for those selection tasks [9].

An increasing number of studies are exploring using hands-free interaction methods for tasks in immersive VR [10]. These alternative methods allow users to perform selections and system control tasks while using their hands to grab and manipulate objects in virtual environments. Such methods include voice commands, eye and head gaze, brain–computer interfaces, electromyography, and body gestures. However, the literature lacks a comprehensive evaluation of hands-free methods, with different methods being shown to be the best for the same task in different studies [10].

This study aims to contribute to this field by evaluating the usability of the most widely used and readily available hands-free interfaces when performing selections in a Fitts' task [11,12] using standard and validated metrics. As Fitts tasks are commonly used to evaluate performance objectively, this study extends their use to evaluate the user experience.

Advanced users have qualities that make them ideal first users of new technologies (they are more knowledgeable and, as such, are more willing to explore and adapt to new techniques and are better at communicating their experience, including in VR applications [13]) and, as such, they will be the primary target audience of this study.

Building on earlier investigations of the performance of these methods [14], this study intends to extend the findings to answer the following research questions:

RQ1 – VR Experience Are hands-free methods capable of delivering a similar or better experience regarding common VR metrics (sense of presence and cybersickness) than handheld controllers?

RQ2 – Usability Do hands-free methods have similar or better system usability and workload than handheld controllers?

RQ3 – User Satisfaction How is the general satisfaction with the interactions affected by the interaction method?

Exploring alternative hands-free techniques and answering these research questions is crucial for improving the design and usability of VR systems. Researchers can gain new insights and possibilities for enhancing user experiences by investigating approaches to interact with VR systems using hands-free methods, leading to more intuitive and immersive interactions.

To achieve the defined goal and answer the research questions, a previously developed evaluation testbed [14] was used in which interaction methods and evaluation modules were added and configured regarding the testing requirements.

2. Related work

2.1. VR experience

Immersive VR is an interactive technology that creates a computer-generated environment, allowing users to interact with and explore that environment [15]. These experiences are primarily visual but can also include additional sensory information like sound and tactile feedback. The most common immersive VR setups (e.g., Meta Quest and HTV Vive) encompass a head-mounted display (HMD) and handheld controllers, which can track the position and rotation of the user's head and hands in real-time and update their virtual representations accordingly.

VR is mainly characterized by the immersion and sense of presence it provides to users, which are factors that directly impact the user experience. Immersion refers to the objective features of the simulated environment, including the senses stimulated, the capacity for interaction, and the realism of the simulation [3,16–18]. The sense of presence refers to the sense of being in the virtual environment generated by natural or mediated means. It encompasses the feeling of "being there" and the ability to interact with the virtual environment [3,16–19].

However, using technology to mediate the delivery of stimuli can negatively impact the experience, especially if there is a mismatch between the stimuli delivered and the stimuli felt by the users [20]. This results in a phenomenon called cybersickness [21], which consists of adverse symptoms similar to motion sickness and may include nausea, discomfort, or fatigue.

2.2. Hands-free selection

The interactive nature of VR allows the creation of highly engaging experiences. In this regard, the most common interaction tasks that users can perform [1,22] include the ability to navigate the environments, manipulate objects, and control the simulations mainly through graphical user interfaces (GUI) that require selecting their elements. As VR experiences become more social, communication tasks are also of great importance since, in a social environment, users are expected to perceive others' emotions and social cues.

Selection tasks, in particular, are a two-step process in which the first part is responsible for picking a target (for instance, by using a pointing device) and the second part for confirming the target selection [23]. Traditionally, this is accomplished with tracked handheld controllers that can point a cursor and one of the buttons to confirm the selection.

Hands-free methods can equally allow to perform selection tasks. Methods that provide directional data, such as eye and head gaze, can be used to point a cursor at targets [24–26]. Also, the feet can act as a novel cursor for selections [27]. However, these hands-free pointing methods do not provide a straightforward confirmation method, and moving the eyes and the head are actions that naturally occur while exploring the virtual environments, requiring the distinction of intent (interact vs. explore) known as the "Midas Touch" problem [28]. Additionally, when using a pointing method, the confirmatory method (multimodal or not) must not influence the pointing and lead to wrong confirmations, known as Heisenberg Errors [29]. As such, alternatives must be used to confirm the selections.

A commonly accepted method of confirming selections is by using a fixation time (dwell time) of 300 ms up to 1000 ms, where users are required to fixate their gaze on the target for a small amount of time [30,31]. It is also possible to implement algorithms that detect when a user intentionally gazes at elements with the intent of selecting those elements [26]. Alternatively, a multimodal approach with confirmatory voice commands [32] or, albeit less commonly, electromyography [33], face expressions [34], and head motions [35] can also be used to complete the selections. Eye input data also allows confirmations by detecting blinks [36].

Despite pointing being the usual metaphor for target acquisition, some methods allow one to perform selections without a cursor. For example, a voice command can include both the target and the intent action [37]. More novel and less explored methods, such as brain–computer interfaces (BCI), can also be used to perform selections directly [38].

2.3. User experience of hands-free selection

Given the importance of selection tasks, they are one of the most evaluated interaction tasks in VR using hands-free methods [10]. Despite this, the evaluations do not usually provide a comparison with other interaction methods and use evaluation metrics specific to the interaction method, not allowing a direct comparison with other methods [10].

Some factors can be evaluated to investigate the user experience of hands-free methods and are related to evaluating the VR experience and the usability of the methods.

The VR experience is commonly evaluated by measuring the users' sense of presence after exposure to the VR environments (e.g., [37,39,40]). For instance, speech methods can provide a better sense of presence than controllers [37] but a worse sense of agency [40]

Additionally, cybersickness and fatigue are also measured to understand if the methods have a negative impact on the experience (e.g., [39,41,42]). Hand gestures proved to be more fatiguing than head gaze for selection [8]. Similarly, using controllers results in a higher simulator sickness than using voice commands [42].

The usability of the methods can be directly evaluated with system usability questionnaires and user satisfaction questionnaires [8,43]. The use of head gaze was compared with speech in a search task, and usability was better with speech, while novelty was higher with a head gaze [39]. Voice commands were also more satisfying [37] and provided a higher engagement [44] than handheld controllers for system control tasks. Recently, the use of eye blinks [45] has been shown as advantageous, giving better performance and usability than other confirmatory methods [36,46,47], given its social acceptability and being the recommended option when an eye tracker is available [36].

Several studies have shown that user satisfaction is higher when using head and eye gaze techniques than when using handheld controllers for selection [41,48,49]. Moreover, head gaze has been shown to have a higher user satisfaction than eye gaze [43] and than other non-hands-free methods [8]. Contrarily, a study showed that touching capacitive buttons on an HMD can be more satisfying than head gaze with dwell [50].

Finally, cognitive load is also measured as an important factor in usability since a higher cognitive load can result in a worse experience [33,51]. In this regard, using eye gaze to point and myography to confirm is shown to have a lower cognitive load than eye or head gaze with dwell times but higher than handheld controllers [33]. Blinking for confirmation has a lower workload than voice commands when performing hands-free text selection.

3. Methodology

3.1. Sample

The experiment was carried out by nine participants aged between 23 and 34 years ($M = 26.6$, $SD = 3.36$). Participants were male volunteers recruited in the laboratory where the experiments were performed and consisted of highly technically educated personnel who are proficient users of VR technologies. Despite this, they had no prior contact with the specific implementation of the interaction methods evaluated in this study.

3.2. Apparatus

The apparatus consisted of a combination of an immersive VR platform providing the necessary visual-auditory stimuli and input devices, a computer responsible for the rendering of the VR simulation, and the test application.

3.2.1. Hardware platform

The HTC Vive Pro Eye immersive VR system was used. The system consisted of the following components: a high-resolution (1440 × 1600 pixels per eye) HMD with a 110° wide field of view (FOV) and a maximum refresh rate of 90 Hz; two tracked handheld controllers with buttons and a trigger; an integrated Tobii® eye tracker with a maximum 120 Hz sampling rate, a 110° tracking range, and up to 0.5° accuracy within a 20° FOV; and an embedded close-range microphone. The system uses SteamVR 2.0 tracking, allowing accurate capturing of the position and rotation of the participants' heads and hands. The non-spatial audio stimulus was delivered using the experiment room sound system.

The VR system was connected to a computer equipped with an Intel® Core™ i7-8700K CPU, an NVIDIA GeForce RTX 3090 GPU, 32 GB of RAM, and an SSD. This combination of hardware ensured that the VR system met the target frame rate for visual delivery and input data processing.

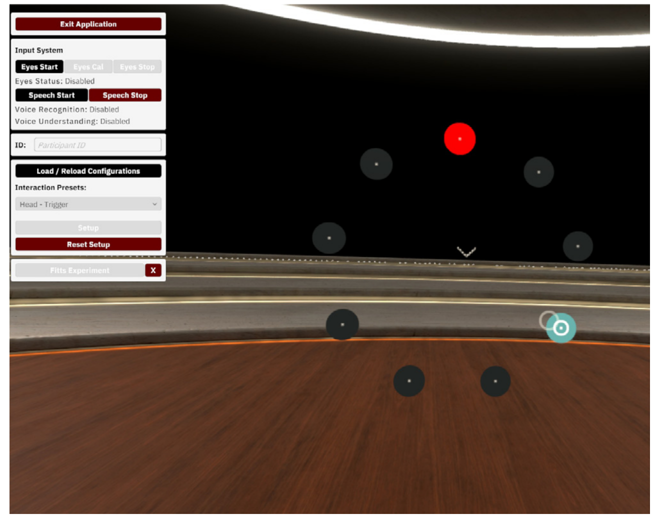

Fig. 1. Screenshot of the application using head gaze for pointing. The red target is the one to be selected. The blue target is being hovered by the cursor.

3.2.2. Software environment

The evaluations were supported by an application (Fig. 1) developed using the Unity 2022.2 game engine. The OpenXR SDK was used to interface with VR devices where possible, allowing the application to be agnostic to the VR setup and using the proprietary device SDKs when required. For voice commands, Azure Cognitive Services[1] and Wit.ai[2] were used for speech-to-text and intent recognition. The Tobii G2OM[3] was used to improve eye gaze with machine learning models and detect which target is being looked at.

The application provides a configurable implementation of interaction interfaces and methods based on modular components: an interaction origin (e.g., hands, eyes), a controller responsible for mapping input data to interaction events, an interactor responsible for responding to interaction events and triggers state changes in interactable objects, a visual representation of the interactor, and a reticle to show the target of interaction better. Interactor visuals, reticles, and targets responded to interaction states (i.e., rest, hover, and select) to provide visual-auditory feedback and affordances to users.

The implemented selection task was a Fitts' task and followed the respective guidelines for this type of task [11,12,52], as well as guidelines for VR interactions [22,53,54]. For sizes, distance-independent millimeters (dmm) were used. The spherical selection targets were placed in a circle pattern, and there were four target diameters (32 dmm, 64 dmm, 96 dmm, and 128 dmm) with five amplitude angles of the circles (30°, 60°, 90°, 120°, and 150°) resulting in 20 possible combinations. Each circle had nine equidistant targets, which were always placed at 1 m from the viewing point. This ensured relevant difficulty indexes for the Fitts' task. To complete the task, the first target of a ring needed to be selected, resulting in ten effective selections for each of the 20 combinations. At any moment, a single target was marked as the correct target for selection.

[1] https://github.com/Azure-Samples/cognitive-services-speech-sdk
[2] https://github.com/wit-ai/wit-unity
[3] https://developer.tobii.com/xr/solutions/tobii-g2om/

3.3. Independent variable

The interaction method is the independent variable in this study. Based on the most commonly used hands-free interaction methods and focusing on the ones with public implementations ready to use by developers, a preliminary test was conducted to anecdotally identify the best combinations of interaction methods while following the standard and best implementation practices. This resulted in the following interaction methods:

C – Controllers: Uses tracked handheld controllers to point at the targets and the trigger button to confirm the selection. The cursor (a single point with 1 dmm) had no smoothing applied to it, and there was a line to show the direction of the pointer.

HV – Head + Voice: Uses head gaze without smoothing applied and with a 6° vertical down offset to point a cursor with a diameter of 16 dmm. To confirm the selection, an affirmative voice command had to be used (e.g., "OK", "Yes", "Confirm") while pointing at a target.

HD – Head + Dwell: Uses the same pointing method and cursor as HV, but confirmations require the fixation on a target for 500 ms.

EV – Eyes + Voice: Uses the smoothed eye gaze to point a cursor with 16 dmm diameter. To confirm the selection, an affirmative voice command had to be used (e.g., "OK", "Yes", "Confirm") while pointing (looking) at the target.

ED – Eyes + Dwell: Uses the same pointing and cursor as EV, but confirmations require the fixation on a target for 500 ms.

AEV – Assisted Eyes + Voice: Uses the eye gaze processed from Tobii G2OM to obtain the target being looked at. Includes a 16 dmm diameter cursor that follows the smoothed eye gaze for visual purposes only. The confirmations use voice commands as the EV method.

AED – Assisted Eyes + Dwell: Uses the same pointing method as AEV with the processed eye gaze and a smoothed cursor. The confirmations are made as in the ED method by a fixation on a target.

V – Voice (Direct): Uses a single voice command to select targets. Each target had a number displayed, and the voice command was the number of the target.

VC – Voice (with Confirmation): Method similar to V; however, the voice command would only result in the target being hovered. A second confirmatory voice command (e.g., as in HV) was required to complete the selection.

3.4. Dependent variables and instruments

This sub-section provides an overview of the instruments used to gather user feedback about the interaction methods. Most of these questionnaires were translated into Portuguese from the original English versions.

3.4.1. Presence

The iGroup Presence Questionnaire (IPQ) [55] was used to assess the users' sense of presence, more specifically, its Portuguese version (IPQp) [56].

This questionnaire, which is applied at the end of the experiment, has 14 sentences where users rank their sense of presence on a 5-point Likert scale from 1 (Totally Disagree) to 5 (Totally Agree).

The results of IPQp are subdivided into four subscales: Spatial Presence, Involvement, Experience Realism, and Overall Presence. The subscales have final scores that range from 1 (worst sense of presence) to 5 (best sense of presence).

3.4.2. Cybersickness

To evaluate the users' cybersickness while interacting with a particular method, the Simulator Sickness Questionnaire (SSQ) was used [57]. The questionnaire was translated from the validated English version to Portuguese.

This questionnaire comprises 16 questions where users assess their symptoms on a four-point scale ranging from 0 (None) to 3 (Severe). Since there is an erroneous assumption that participants have zero symptoms before the experiment [58], the questionnaire is applied before and after the exposure to the interaction method.

Four subscales subdivide the scores that result from the questionnaire, and each has different score ranges: Nausea (values range from 0 to 200.34), Oculomotor (values range from 0 to 159.18), Disorientation (values range from 0 to 292.32), and Overall Cybersickness (values range from 0 to 235.62). The lower the value, the better.

Additionally, total head rotation was also measured as a factor to evaluate better the fatigue of using the interaction methods.

3.4.3. System Usability

The System Usability Questionnaire (SUS) [59] was used to assess the user experience. This questionnaire has ten affirmations that describe multiple aspects of system usability. Users classify each of the affirmations using a 5-point Likert scale, ranging from 1 (completely disagree) to 5 (completely agree). The final SUS Score is considered a raw value that, despite being from 0 to 100, does not represent a percentage, cannot be directly interpreted, and must be used in the context of a comparison.

A commonly used way to interpret these scores is using the scores and grades found in [60]. By comparing the scores of multiple studies and calculating a percentile rank, the authors segregated the scores and attributed a letter grade from A (Excellent) to F (Awful). Additionally, the author also notes that grade A (corresponding to the top 10% of the results) is where users start to recommend the system to others.

Following [61], a 7-point adjective rating scale was added to the SUS questionnaire that ranges from "Worst Imaginable" (1) to "Best Imaginable" (7), where users can directly classify their usability with the interaction method. This was added to interpret the SUS Scores better.

3.4.4. Workload

The workload of interaction methods is subjectively evaluated using the NASA Task Load Index (NASA-TLX) [62]. Since different factors can influence the workload, this questionnaire uses a weighted average of ratings on six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The scores of these subscales with values from 0 up to 500 (depending on the weight of the subscale) calibrate the final Workload score that ranges from 0 (no workload) to 100 (extreme workload).

Additionally, the Integrated Workload Scale (IWS) [63] was used as a more direct, subjective measure of workload. This questionnaire has a single question where the users rank on a scale of 9 items to determine which of the items best describes the perceived workload of the task, where 1 is the lowest workload, and 9 is the highest. This differs from NASA-TLX, which has a broader definition of Workload, whereas the IWS focuses more on task difficulty and demand.

3.4.5. User satisfaction

User satisfaction was also measured using a custom questionnaire derived from questions of [64,65]. This questionnaire comprises eight affirmations related directly to perceived user satisfaction, technology acceptance, and intention to use the system again. Users answer on a 7-point Likert scale from 1 (Totally Disagree) to 7 (Totally Agree). The mean of all questions results in a final user satisfaction score from 1 (Worst) to 7 (Best). Additionally, this questionnaire has an open-ended question where users are asked to enumerate the perceived strengths and weaknesses of the system.

3.5. Design and procedure

The nine interaction methods were evaluated using a within-subjects design, consisting of nine sessions where a single method was evaluated. The order of the methods being evaluated was randomized per session and participant to minimize order effects. Only one interaction method was evaluated in a single session, resulting in participants attending the experiment site for nine consecutive days. Because of this, to guarantee the anonymization of the collected data, a letter was assigned to each participant and a numeric identifier to each session. No collected data allowed a direct connection to a participant.

The experiments were carried out in an experimental room isolated from potential external distraction factors where only the researcher experimenting and the participant were present. To participate in the experiment, volunteers were required to fill out a consent form with information regarding the experiment goals and participant rights, including the right to withdraw without penalization. After accepting this, they filled out a brief sociodemographic questionnaire for sample characterization in the first session.

At the start of each session, participants filled out the SSQ to gather the baseline cybersickness before being equipped with the VR setup. Afterward, they were equipped with the HMD and one handheld controller (if in the Controller condition). Once in the VR environment, an eye-tracker calibration was performed to guarantee its correct adjustment to each participant. Because of limitations on the eye-tracker, participants could not wear corrective glasses as they interfered with the correct tracking, and only participants who stated that this limitation would not penalize their experience and well-being were accepted. This calibration was performed in all conditions as it also ensured the correct placement of the HMD in the participants' faces for optimal viewing conditions. A second calibration ensured the placement of the virtual targets to the participants' heights.

The calibrations were followed by a familiarization phase where a tutorial was performed using the evaluated interaction method, consisting of a reduced version of the main experiment task. This allowed participants to understand the method and task and perceive how the interface responded to user interactions. Participants could request help and additional information during this phase and were instructed to complete the selection task as fast as possible. The participants were not informed of the difference between methods with and without assisted eye gaze. The main experiment task was performed after the tutorial was finished and with the participant's approval. For a single evaluation session, each participant made a total of 240 correct selections (40 in the tutorial and 200 during the main task). In total, after all of the nine sessions, each participant made a total of 2160 correct selections.

At the end of a single session, after completing the task and the removal of the VR equipment, participants filled out the remaining questionnaires, and a brief interview was conducted to gather feedback from the participants about the interaction method used. The whole session lasted for approximately 30 min. A final interview was conducted after all of the sessions to ask each participant their view and opinions on the experiment and interaction methods. The entire study was conducted over approximately one month, accounting for the participant sessions, final interview, and data processing.

3.6. Data analysis

The data was collected in CSV format and processed and analyzed using the RStudio[4] software with packages that support the required statistical tests and data visualization. The significance level was set at 95% (alpha level of 0.05) to determine statistical significance for all tests.

No outliers were removed because of the subjective nature of questionnaire answers and sample size. A Shapiro–Wilk test was performed for each group and variable, confirming that the data did not follow a normal distribution ($p < 0.05$). As such, the groups were compared using the non-parametric Friedman test as an alternative to the repeated-measures ANOVA, given its tolerance to outliers and data that does not follow a normal distribution. The Friedman test uses the ranks of the values of each group for comparisons. When statistically significant differences are observed, a post-hoc analysis was performed using the Conover test with a Bonferroni correction to understand which groups are statistically different. Only statistically significant differences will be reported for this post-hoc analysis.

Additionally, as described in [60], the SUS scores were also transformed into percentile ranks and graded in a process similar to a norm-referenced test, allowing the correct interpretation and comparison of those scores.

To increase confidence in the results, Kendall's coefficient of concordance (W) is also reported as a measure of effect size and is classified following Cohen's interpretation guidelines [66] in small ($<$ 0.3), medium (0.3 – 0.5) and large ($>$ 0.5) effect.

[4] https://posit.co/products/open-source/rstudio/

Table 1
Descriptive statistics for the IPQp data.

Subscale	Method	M	SD	Mdn	IQR	Min	Max
Spatial presence	C	3.519	0.444	3.667	0.667	2.833	4.000
	HV	3.407	0.678	3.167	0.667	2.667	4.500
	HD	3.593	0.401	3.667	0.333	2.667	4.000
	EV	3.611	0.534	3.833	1.000	2.833	4.333
	ED	3.444	0.486	3.667	0.833	2.833	4.000
	AEV	3.574	0.553	3.667	0.833	2.667	4.167
	AED	3.685	0.530	3.833	0.500	2.667	4.500
	V	3.704	0.576	3.833	0.667	2.833	4.667
	VC	3.463	0.803	3.833	1.000	1.833	4.333
Involvement	C	3.167	0.673	3.000	1.000	2.500	4.250
	HV	2.944	1.014	3.000	0.750	1.250	4.500
	HD	3.361	0.945	3.250	1.750	2.250	4.750
	EV	3.167	0.875	3.000	1.000	2.250	4.750
	ED	3.389	0.830	3.500	1.250	2.250	4.500
	AEV	2.972	0.897	2.750	1.250	2.000	4.500
	AED	3.500	0.927	3.500	1.500	2.250	4.750
	V	3.250	0.893	3.250	1.250	2.250	4.750
	VC	2.861	0.397	3.000	0.500	2.250	3.500
Experienced realism	C	2.222	0.972	2.250	1.750	1.000	3.500
	HV	2.306	0.899	2.250	1.500	1.250	3.750
	HD	2.194	0.958	1.750	1.750	1.000	3.500
	EV	2.028	1.034	1.750	1.750	1.000	3.500
	ED	2.250	0.866	2.000	1.000	1.000	3.500
	AEV	1.972	0.870	1.750	1.250	1.000	3.250
	AED	2.194	1.021	2.500	2.000	1.000	3.500
	V	2.222	0.964	2.250	2.000	1.000	3.500
	VC	2.139	0.885	2.000	1.500	1.000	3.250
Presence	C	2.969	0.431	3.028	0.500	2.250	3.667
	HV	2.886	0.662	3.000	0.778	1.972	4.250
	HD	3.049	0.416	2.972	0.583	2.583	3.861
	EV	2.935	0.611	2.917	0.472	2.083	4.111
	ED	3.028	0.421	3.167	0.528	2.194	3.528
	AEV	2.840	0.498	3.056	0.667	2.139	3.417
	AED	3.127	0.556	3.222	0.639	2.139	4.083
	V	3.059	0.598	3.111	0.500	2.194	4.222
	VC	2.821	0.458	2.944	0.750	2.139	3.306

4. Results

The results were subdivided into subsections according to the dependent variables.

4.1. Presence

The descriptive values of IPQp subscales are in Table 1 and the overall values in Fig. 2. No statistically significant differences were found between the groups regarding Spatial Presence ($\chi^2(8) = 6.703,\ p = 0.569,\ W = 0.09$), Involvement ($\chi^2(8) = 15.134,\ p = 0.570,\ W = 0.21$), Experienced Realism ($\chi^2(8) = 7.983,\ p = 0.435,\ W = 0.11$), and overall Presence ($\chi^2(8) = 13.200,\ p = 0.105,\ W = 0.18$).

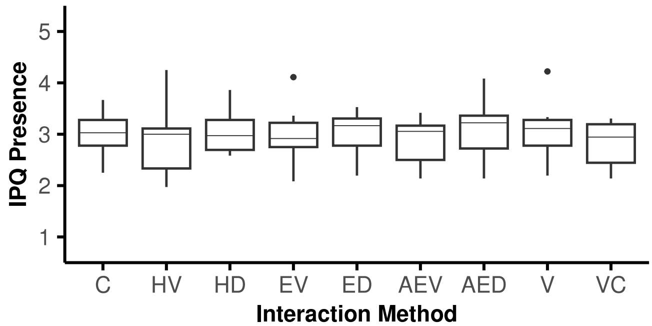

Fig. 2. Overall presence values for each of the interaction methods.

Table 2
Descriptive statistics for the SSQ data.

Subscale	Method	M	SD	Mdn	IQR	Min	Max
Nausea	C	0.000	0.000	0.000	0.000	0.000	0.000
	HV	2.120	4.207	0.000	0.000	0.000	9.540
	HD	0.000	0.000	0.000	0.000	0.000	0.000
	EV	4.240	5.028	0.000	9.540	0.000	9.540
	ED	1.060	3.180	0.000	0.000	0.000	9.540
	AEV	1.060	3.180	0.000	0.000	0.000	9.540
	AED	1.060	3.180	0.000	0.000	0.000	9.540
	V	1.060	3.180	0.000	0.000	0.000	9.540
	VC	6.360	6.746	9.540	9.540	0.000	19.080
Oculomotor discomfort	C	−0.842	4.555	0.000	0.000	−7.580	7.580
	HV	3.369	5.507	0.000	7.580	0.000	15.160
	HD	4.211	7.685	0.000	7.580	0.000	22.740
	EV	5.896	6.317	7.580	7.580	0.000	15.160
	ED	4.211	10.107	0.000	0.000	0.000	30.320
	AEV	6.738	7.034	7.580	15.160	0.000	15.160
	AED	2.527	3.790	0.000	7.580	0.000	7.580
	V	1.684	3.342	0.000	0.000	0.000	7.580
	VC	6.738	12.821	0.000	7.580	0.000	37.900
Disorientation	C	1.547	4.640	0.000	0.000	0.000	13.920
	HV	1.547	4.640	0.000	0.000	0.000	13.920
	HD	4.640	9.843	0.000	0.000	0.000	27.840
	EV	1.547	4.640	0.000	0.000	0.000	13.920
	ED	0.000	0.000	0.000	0.000	0.000	0.000
	AEV	3.093	6.138	0.000	0.000	0.000	13.920
	AED	3.093	9.280	0.000	0.000	0.000	27.840
	V	1.547	4.640	0.000	0.000	0.000	13.920
	VC	4.640	6.960	0.000	13.920	0.000	13.920
Cybersickness	C	0.000	1.870	0.000	0.000	−3.740	3.740
	HV	2.909	4.087	0.000	3.740	0.000	11.220
	HD	3.324	6.043	0.000	3.740	0.000	18.700
	EV	4.987	4.181	3.740	3.740	0.000	11.220
	ED	2.493	6.202	0.000	0.000	0.000	18.700
	AEV	4.571	4.868	3.740	7.480	0.000	11.220
	AED	2.493	3.740	0.000	3.740	0.000	11.220
	V	1.662	1.971	0.000	3.740	0.000	3.740
	VC	7.064	9.433	3.740	7.480	0.000	26.180

4.2. Cybersickness

The descriptive values of SSQ subscales can be found in Table 2 and the overall values in Fig. 3. Statistically significant differences appear to be found between the values of Nausea within the groups, $\chi^2(8) = 19.788$, $p = 0.011$, $W = 0.27$. However, after a post-hoc analysis, we verified this is a type I error as no post-hoc comparison was statistically significant ($p >= 0.05$). No statistically significant differences were found between the values of Oculomotor Discomfort ($\chi^2(8) = 11.446$, $p = 0.178$, $W = 0.16$), the values of Disorientation ($\chi^2(8) = 6.181$, $p = 0.627$, $W = 0.09$), and values of overall Cybersickness ($\chi^2(8) = 15.349$, $p = 0.053$, $W = 0.21$).

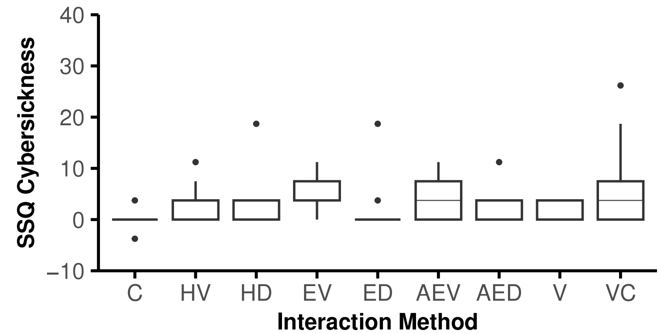

Fig. 3. Overall cybersickness values for each of the interaction methods.

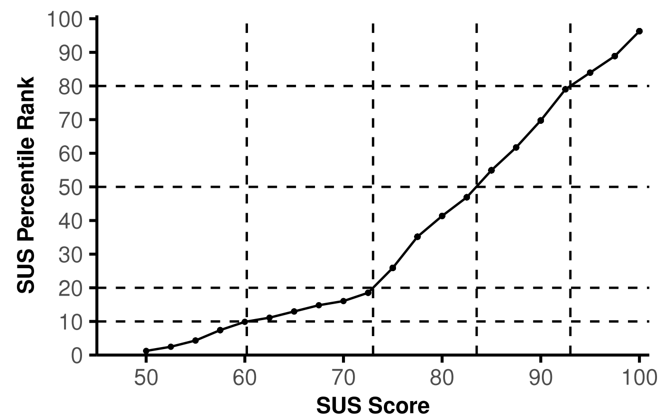

Fig. 4. Distribution of percentile ranks by SUS Score. Ranks were calculated using only this experiment data.

Table 3
SUS Score grading based on percentile rank values of this study.

Grade	Percentile rank	SUS score
A - Excellent	≥ 80	≥ 93
B - Good	< 80	< 93
C - OK	< 50	< 83.5
D - Poor	< 20	< 73
F - Awful	< 10	< 60.2

4.3. System usability

The analysis of SUS data requires processing the SUS Scores since the raw scores are not meaningful. The ranks and respective grades calculated in this study are only meaningful in the context of this study. Table 3 shows the mapping of letter grade to percentile rank and SUS score limits. Moreover, Fig. 4 shows this association.

Regarding SUS, the methods were compared using the grades found by [60], the percentile ranks calculated for this study, and the grades associated with those percentile ranks. Table 4 shows the descriptive values for this SUS data.

When comparing the SUS Grades, no statistically significant differences were found in the post-hoc tests despite the significant result of the Friedman test ($\chi^2(8) = 24.039$, $p = 0.002$, $W = 0.33$) and despite the significant difference found when comparing the raw scores ($\chi^2(8) = 36.913$, p< 0.001, $W = 0.51$).

Statistically significant differences were found when comparing the percentile ranks of the SUS scores, $\chi^2(8) = 36.913$, p< 0.001, $W = 0.51$. The post-hoc analysis showed that the method with the highest percentile ranks AED (Grade A, $Mdn = 89.951$) had significantly higher values than those with the lower ranks, namely, HV (Grade C, $Mdn =$

Table 4
Descriptive statistics for the SUS Scores.

	Method	M	SD	Mdn	IQR	Min	Max	Grade[a]	Grade[b]
SUS score raw	C	86.111	9.110	90.000	5.000	70.000	97.500	A	A
	HV	76.111	17.235	80.000	20.000	50.000	100.000	B	A
	HD	86.389	11.118	90.000	15.000	65.000	100.000	A	A
	EV	77.500	13.807	77.500	17.500	57.500	95.000	B	B
	ED	83.611	13.057	82.500	10.000	57.500	100.000	A	A
	AEV	76.111	9.772	77.500	7.500	55.000	85.000	B	A
	AED	93.056	8.176	95.000	10.000	75.000	100.000	A	A
	V	84.167	12.930	85.000	17.500	60.000	100.000	A	A
	VC	75.000	10.232	75.000	7.500	55.000	85.000	B	B
SUS percentile rank	C	58.711	25.366	69.753	14.815	16.049	88.889	B	B
	HV	41.015	33.486	41.358	41.975	1.235	96.296	C	B
	HD	60.974	27.709	69.753	43.827	12.963	96.296	B	B
	EV	40.947	29.875	35.185	51.235	7.407	83.951	C	C
	ED	53.978	29.985	46.914	28.395	7.407	96.296	C	C
	AEV	34.431	17.991	35.185	20.988	4.321	54.938	C	C
	AED	77.572	22.919	83.951	26.543	25.926	96.296	A	A
	V	55.556	30.768	54.938	48.765	9.877	96.296	B	C
	VC	32.373	18.283	25.926	20.988	4.321	54.938	C	C

[a] Grades associated with the Median value. For raw scores, the values from [60] were used.
[b] The most frequent grade in the group.

41.358, $p = 0.042$), EV (Grade C, $Mdn = 35.185$, $p = 0.016$), AEV (Grade C, $Mdn = 35.185$, $p = 0.005$), and VC (Grade C, $Mdn = 25.926$, $p = 0.001$).

4.4. Workload

The descriptive values for all the NASA-TLX subscales and the IWS can be found in Table 5. Fig. 5 shows the overall Workload values for each method.

Statistically significant differences were found between the values of Mental Demand within the groups ($\chi^2(8) = 21.778$, $p = 0.005$, $W = 0.30$) and in the values of Physical Demand ($\chi^2(8) = 20.687$, $p = 0.008$, $W = 0.29$). As with Nausea comparisons, the post-hoc analysis revealed that no significant differences exist in each pairwise comparison in these two subscales, indicating a type I error.

Regarding the remaining subscales of the NASA-TLX, no statistically significant differences were found in Temporal Demand ($\chi^2(8) = 5.614$, $p = 0.690$, $W = 0.09$), Performance ($\chi^2(8) = 9.020$, $p = 0.341$, $W = 0.13$), Effort ($\chi^2(8) = 8.603$, $p = 0.377$, $W = 0.12$), Frustration ($\chi^2(8) = 15.021$, $p = 0.059$, $W = 0.21$), and Workload ($\chi^2(8) = 5.034$, $p = 0.754$, $W = 0.07$).

The differences between the IWS values in the groups were also not statistically significant despite the test results ($\chi^2(8) = 19.198$, $p = 0.014$, $W = 0.27$) since the posthoc analysis showed no significant differences in any of the pairwise comparisons (type I error).

4.5. User satisfaction

Table 6 shows the descriptive statistics for the usability adjective and user satisfaction data. Fig. 6 shows the overall user satisfaction. Results from comparing the adjective used for the usability of the methods show that no statistically significant differences exist despite the Friedman test showing significance ($\chi^2(8) = 26.443$, $p = 0.001$, $W = 0.37$) since the pairwise comparisons were not significant (type I error). User satisfaction was significantly different between the methods ($\chi^2(8) = 19.969$, $p = 0.010$, $W = 0.28$), more specifically AED ($Mdn = 6.750$) had much higher user satisfaction values than HV ($Mdn = 5.750$, $p = 0.035$) and AEV ($Mdn = 5.625$, $p = 0.027$).

4.6. Total head rotation

Table 7 shows the descriptive statistics for the total head rotation of the users during the experiment. Fig. 7 shows this data graphically.

Statistically significant differences were found between the interaction methods, $\chi^2(8) = 42.696$, p< 0.001, $W = 0.59$. For the pairwise

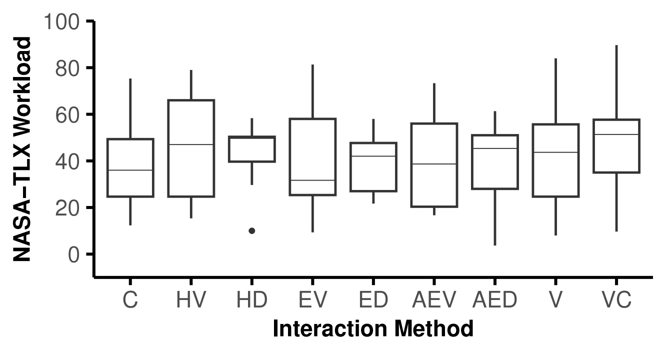

Fig. 5. NASA-TLX Workload values for each of the interaction methods.

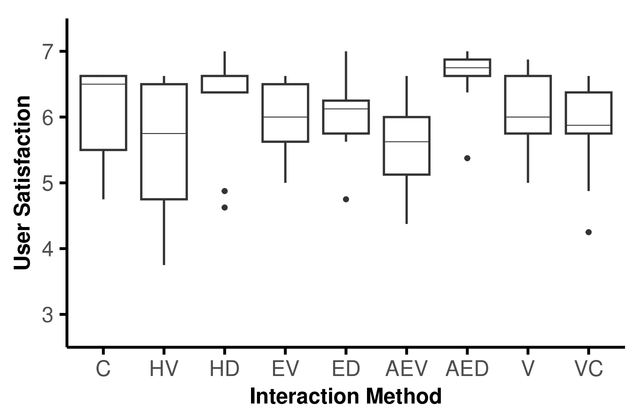

Fig. 6. User satisfaction values for each of the interaction methods.

comparisons, the following differences in total head rotation angle were significant:

- HV ($Mdn = 16697.235$) was significantly higher than for ED ($Mdn = 11262.635$, $p = 0.015$), AEV($Mdn = 11839.909$, $p = 0.04$), AED ($Mdn = 9901.292$, $p = 0.003$), V ($Mdn = 9360.191$, $p = 0.002$), and VC ($Mdn = 10171.455$ $p = 0.020$);
- HD ($Mdn = 16135.494$) was significantly higher than for AED ($Mdn = 9901.292$, $p = 0.015$), and V ($Mdn = 9360.191$, $p = 0.012$).

Table 5
Descriptive statistics for the Workload data from NASA-TLX and IWS.

Metric	Method	M	SD	Mdn	IQR	Min	Max
Mental demand	C	33.889	49.547	10	35	0	135
	HV	42.778	70.539	10	50	0	220
	HD	20.556	25.304	15	20	0	80
	EV	67.222	131.318	0	10	0	325
	ED	28.889	49.103	10	40	0	150
	AEV	47.778	96.537	10	30	0	300
	AED	45.556	83.045	10	40	0	260
	V	70.556	81.603	40	110	0	225
	VC	116.667	151.781	60	65	5	475
Physical demand	C	137.222	135.603	100	155	20	450
	HV	146.667	121.527	100	185	0	340
	HD	175.556	127.216	100	210	60	375
	EV	124.444	113.011	90	120	20	380
	ED	102.222	110.880	45	165	0	320
	AEV	107.778	156.127	30	110	0	475
	AED	61.667	83.179	20	80	0	260
	V	74.444	131.943	10	40	0	360
	VC	71.111	103.706	20	75	0	300
Temporal demand	C	79.444	86.907	55	100	0	250
	HV	147.222	110.259	150	145	20	320
	HD	120.000	130.767	50	100	10	400
	EV	98.333	97.660	60	165	0	240
	ED	105.000	101.581	40	145	0	250
	AEV	142.778	117.210	140	105	10	400
	AED	99.444	116.228	30	185	0	325
	V	128.889	70.966	150	50	10	240
	VC	179.444	108.179	200	150	15	350
Performance	C	161.111	147.813	110	175	0	400
	HV	101.667	149.917	45	50	0	475
	HD	137.778	138.949	90	60	15	475
	EV	95.556	79.587	45	110	10	210
	ED	92.222	62.755	60	90	25	200
	AEV	85.556	70.774	60	115	15	220
	AED	181.667	180.520	120	280	0	475
	V	140.000	121.037	100	170	15	360
	VC	82.222	90.661	50	60	0	260
Effort	C	127.222	98.714	90	60	40	340
	HV	193.333	113.771	210	210	45	340
	HD	171.111	134.508	260	250	20	320
	EV	172.778	141.490	160	200	30	450
	ED	152.778	103.987	120	75	30	375
	AEV	145.000	136.999	70	245	25	380
	AED	122.778	124.978	90	175	15	400
	V	132.222	145.648	30	220	5	360
	VC	141.667	132.169	105	225	0	350
Frustration	C	56.111	121.135	10	30	0	375
	HV	76.667	145.516	0	90	0	450
	HD	8.889	19.808	0	5	0	60
	EV	68.889	77.370	40	95	0	210
	ED	100.000	111.608	30	150	0	300
	AEV	96.667	119.791	60	160	0	340
	AED	40.000	98.362	0	30	0	300
	V	76.111	121.957	30	80	0	375
	VC	131.667	166.827	45	255	0	400
Workload	C	39.667	20.549	36.000	24.667	12.333	75.333
	HV	47.222	24.332	47.000	41.333	15.333	79.000
	HD	42.259	14.782	49.667	10.667	10.000	58.333
	EV	41.815	23.828	31.667	32.667	9.333	81.333
	ED	38.741	13.358	42.000	20.667	21.667	58.000
	AEV	41.704	22.514	38.667	35.667	16.667	73.333
	AED	36.741	20.621	45.333	23.000	3.667	61.333
	V	41.481	23.573	43.667	31.000	8.000	84.000
	VC	48.185	23.566	51.333	22.667	9.667	89.667
IWS	C	3.333	1.581	3	2	1	6
	HV	3.778	1.202	4	1	2	6
	HD	3.000	1.118	3	2	2	5
	EV	3.556	0.882	4	1	2	5
	ED	2.778	0.833	3	1	2	4
	AEV	3.556	0.882	4	1	2	5
	AED	3.000	1.500	3	2	1	6
	V	2.556	0.726	2	1	2	4
	VC	3.556	1.590	3	2	2	6

(continued on next page)

Table 5 (continued).

Table 6
Descriptive statistics for the Usability adjective and User satisfaction.

Metric	Method	M	SD	Mdn	IQR	Min	Max
Usability adjective	C	5.667	0.866	6	1	4	7
	HV	4.778	1.302	5	2	2	6
	HD	5.333	0.707	5	1	4	6
	EV	4.889	0.601	5	0	4	6
	ED	5.778	0.667	6	1	5	7
	AEV	5.111	0.601	5	0	4	6
	AED	6.000	0.866	6	2	5	7
	V	5.556	0.726	6	1	4	6
	VC	4.667	0.707	5	1	4	6
User satisfaction	C	6.069	0.798	6.500	1.125	4.750	6.625
	HV	5.528	1.075	5.750	1.750	3.750	6.625
	HD	6.208	0.850	6.625	0.250	4.625	7.000
	EV	5.889	0.611	6.000	0.875	5.000	6.625
	ED	6.056	0.641	6.125	0.500	4.750	7.000
	AEV	5.556	0.796	5.625	0.875	4.375	6.625
	AED	6.625	0.508	6.750	0.250	5.375	7.000
	V	5.988	0.669	6.000	0.875	5.000	6.875
	VC	5.833	0.803	5.875	0.625	4.250	6.625

Table 7
Descriptive statistics for the total head rotation (in degrees).

Method	M	SD	Mdn	IQR	Min	Max
C	12 814.827	2040.557	12 772.771	2285.175	10 756.402	17 131.266
HV	16 587.936	761.555	16 697.235	1266.498	15 377.538	17 535.901
HD	16 205.181	880.358	16 135.494	845.145	14 840.465	17 460.700
EV	11 787.961	1511.278	11 629.127	1886.314	10 056.262	14 550.154
ED	10 843.334	1625.707	11 262.635	2167.499	8720.409	13 652.418
AEV	11 203.376	1975.644	11 839.909	3204.800	8338.694	13 812.196
AED	10 317.254	1900.635	9901.292	2290.654	8154.302	13 929.502
V	8746.558	3967.294	9360.191	6752.019	2286.855	13 057.414
VC	11 424.381	8654.366	10 171.455	3225.105	2727.580	32 262.461

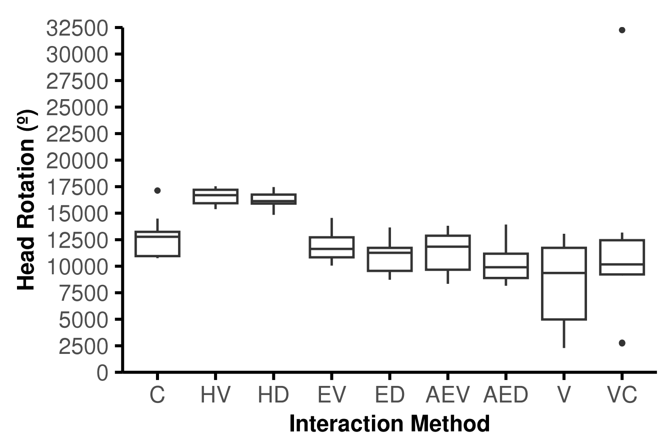

Fig. 7. Total head rotation (in degrees) for each interaction method.

Generally, the head rotation is significantly higher in the methods that use the head for pointing.

4.7. User feedback

Participants reported that the methods were “easy to use, learn, and intuitive” and appreciated the fact that the methods did not require the use of the hands, providing more “freedom” and “better immersion”.

Despite the use of voice commands being well received, it was highly mentioned that the “recognition was slow” or the “commands not recognized”, leading to a “feeling of frustration”, especially in the VC method which required a second voice command. Additionally, methods that required a voice confirmation were classified as “frustrating” since they required the cursor to stay fixated on the targets during voice recognition.

Some participants also reported that the methods with “a larger cursor allowed snapping the cursor to targets and made the selections more accurate and fast”. At the same time, controllers were “too dependent on stable hand movements”. Hands-free methods resulted in “neck discomfort for targets outside the field of view”. In contrast, the handheld controllers resulted in “fatigue in hand and trigger finger” for the used selection task. While the methods with assisted eye gaze were perceived as “more precise” than their non-assisted counterparts, participants pointed out the “cursor jitter and loss of tracking when looking through the edges of the lens”. Methods considered the fastest (C and AED) were also perceived as “highly demanding regarding temporal effort”, making the task “too competitive”.

Regarding the feedback provided on the application, it was noted that the targets should provide a “distinct affordance for when the correct target is being hovered, instead of the standard hovering color”. Some participants also referred that the “targets were too close” to the source of interaction. Additionally, for methods with eye gaze, participants reported that the visuals of the cursor could be omitted, resorting to the affordances provided by the targets.

5. Discussion

5.1. RQ1 - VR experience

When analyzing the methods regarding their impact on the VR experience, we found that the users’ sense of presence and cybersickness was not significantly different depending on the method. Overall presence values were low [67], as the low values of the experience realism subscale impacted them. We believe this is due to the nature of the task since the environment was visually simple and did not require more spatialization than what was in front of the user. The task does not have a real-world equivalent (resulting in lower experience realism values) and was also very specialized without the need to explore the environment. The use of eye gaze alone (ED/AED) provided slightly better presence than the head gaze when used with dwell (HD) and handheld controllers (C). Moreover, using the voice as a confirmatory method (HV/EV/AEV/VC) showed a slightly worse presence than not using it (HD/ED/AED/V).

Values of cybersickness were residual throughout the experiments, suggesting that none of the interaction methods negatively impacted the VR experience. However, there was a slight increase in ocular discomfort in the conditions that used rapid eye movements (ED/AED).

5.2. RQ2 - system usability

Regarding the system usability of the studied interaction methods, we found that the methods were not significantly different if the classification of [60] is followed. This might happen since the pool of studies used to obtain the grades does not include the type of experience found in this study. When computing the percentile ranks of this study, we observed that the assisted eye gaze with dwell (AED) registered significantly higher system usability than the methods that use the voice as a confirmatory method (HV/EV/AEV/VC). This contradicts [39] where in a task with context (unlike this study), direct speech (similar to V) had higher usability than head gaze. This suggests that results differ in a more immersive task than an elementary and direct task like the Fitts’. We also found that using two voice commands for each selection (VC) was worse than all the other methods, which could be attributed to the longer times required for voice recognition to complete [14].

The subjective data results of the NASA-TLX and IWS questionnaires have shown that the hands-free methods do not result in a significantly higher workload than the handheld controllers. Despite this, as found in [33], the head gaze and eye gaze with dwell time had higher workloads than the eye gaze with a confirmatory method (the voice in this study) and the controllers. This contradicts [41], where eye gaze alone had a lower cognitive load. Contrary to what we expected, the multimodal methods had a slightly lower workload than their non-multimodal counterparts. Workload values from both questionnaires were low across all methods. Additionally, the use of eye data could be improved by using eye blinks since it can have better results than the dwell method [47].

Overall, we observed that the hands-free methods, especially the non-multimodal ones, produced similar or better usability than the handheld controllers in most metrics. It was also observable that users preferred assisted eye gaze with dwell.

5.3. RQ3 - user satisfaction

For user satisfaction, all the methods registered high values namely, the satisfaction with the usability was rated by users with “good” for multimodal methods and “excellent” for the non-multimodal ones. Moreover, the overall user satisfaction with the experiment followed the same trend, with the eye gaze assisted with machine learning and dwell (AED) registering a significantly higher user satisfaction than the same method and head gaze combined with the voice (AEV/HV). This was also verified with the anecdotal data, where that method was perceived as faster, more responsive to input, and more performant. Despite the significantly higher head rotations in the methods that used head gaze for pointing (as in [51]), user satisfaction and usability were not significantly lower in those methods.

Participants’ frustration with methods that use the voice could be attributed to the longer times required to process the voice input [14], suggesting faster recognition and the ability to lock the cursor on the target automatically are possible solutions to increase the usability of such methods. When inquired about the usage of the methods in public spaces, participants mentioned that they would prefer to use the voice for experiences where they are isolated or in a comfortable environment (e.g., at home), which is in line with studies that found the voice to be less socially acceptable for users [36,68]. Additional care must be taken when placing interactable elements since elements outside viewing angles seem to generate discomfort when interacted with for long periods. Selections can also be improved (including for handheld controllers) by snapping the cursor to the elements.

6. Limitations and future work

We have identified some limitations and opportunities for future work regarding the evaluation of user experience while using hands-free selection methods.

This study was limited by the constraints of the available VR system, leading to the focus on methods whose implementation is public and readily available to be used by developers. For instance, eye blinks were excluded because the hardware did not provide a consistent user experience, leading to high frustration. Moreover, the system used in this study could not work when users had corrective glasses and showed limitations in tracking accuracy on the edges of the HMD lenses. Furthermore, the speed of the voice recognition system directly impacted the performance and, consequently, the user experience.

We must note that the findings are limited to a sample of advanced users, and future studies are required to increase the confidence and generalization of the results to general VR users. Future work should be conducted to evaluate more novel hands-free interaction methods (e.g., brain–computer interfaces [38] and motion-based techniques [69]), as well as the use of machine learning to improve the existing methods.

Further evaluations should be conducted to better understand how different methods' intrinsic characteristics affect their user experience and usability. Additionally, the context of interactions should also be studied to verify the applicability of the results for selections that occur during a typical VR experience.

7. Practical implications

The study presents takeaways for application developers that can guide the development of VR applications. Hand-free methods are not detrimental to the VR experience and can be used to provide engaging and enjoyable experiences.

While assisted eye gaze is the preferred hands-free pointing method, developers should take into account the fatigue it causes, meaning that if their application has too many selections to be performed, it is pertinent to provide an alternative method, such as voice commands or organize the selection targets so they require fewer gaze movements and consequently less discomfort to users. Implementing techniques that improve target acquisition (e.g., cursor snapping) is also advantageous to improve the user experience of the selection methods.

Application developers must consider that using an inherently slower method, such as voice, will usually result in worse values across the metrics, so they should explore and prefer faster recognition systems. Moreover, they should consider the context of the application and where it will be used; as such, they should provide methods that are well integrated with the application and make sense for its purpose, but also consider that if the application is used in a public setting, it might be wise not to offer voice recognition as the only input method.

8. Conclusion

This study evaluated the most commonly used hands-free interfaces for immersive VR selection and system control tasks and their impact on user experience and usability metrics. The nine evaluated methods include the traditional handheld controllers as a baseline, three hands-free methods that can perform the selections directly with dwell, two hands-free methods where only the voice is used, and three hands-free methods with a multimodal approach where the voice is used as a confirmatory method.

A Fitts' selection task was conducted, and we found that the different methods did not significantly differ in terms of the user's sense of presence and cybersickness. Also, we found that workload was low across all methods while user satisfaction was high. Moreover, we observed that using assisted eye gaze with dwell time was the hands-free method with the highest usability and was preferred by users.

The evaluation also showed that using a multimodal approach for selections, especially using the voice, decreases user satisfaction and increases users' frustration. As such, the use of faster voice recognition methods is of most importance to enable satisfying interactions.

In conclusion, this study provides valuable insights into the design of applications that use hands-free methods for selection in immersive VR experiences. It contributes to the literature by presenting a comprehensive evaluation of the usability of the methods and provides an initial classification of SUS grades for VR studies. Finally, it can be used as a guide for implementing hands-free selection in VR scenarios.

CRediT authorship contribution statement

Pedro Monteiro: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Hugo Coelho:** Software, Investigation, Resources, Data curation. **Guilherme Gonçalves:** Software, Validation, Formal analysis, Resources, Writing – review & editing. **Miguel Melo:** Methodology, Supervision, Writing – review & editing, Project administration. **Maximino Bessa:** Conceptualization, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is financed in part by National and European Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, Portugal, within the project SFRH/BD/147813/2019.

References

[1] Sherman WR, Craig AB. Understanding virtual reality: Interface, application, and design. Morgan Kaufmann; 2002.
[2] Parong J, Mayer RE. Learning science in immersive virtual reality. J Educ Psychol 2018;110(6):785–97. http://dx.doi.org/10.1037/edu0000241.
[3] Gonçalves G, Coelho H, Monteiro P, Melo M, Bessa M. Systematic review of comparative studies of the impact of realism in immersive virtual experiences. ACM Comput Surv 2022;55(6):1–36. http://dx.doi.org/10.1145/3533377.
[4] Jacobsen LF, Krogsgaard-Jensen NM, Peschel AO. Shopping in reality or virtuality? A validation study of consumers' price memory in a virtual vs. Physical supermarket. Foods 2022;11(14):2111. http://dx.doi.org/10.3390/foods11142111.
[5] Melo M, Goncalves G, Monteiro P, Coelho H, Vasconcelos-Raposo J, Bessa M. Do multisensory stimuli benefit the virtual reality experience? A systematic review. IEEE Trans Vis Comput Graph 2020;28(01):1428–42. http://dx.doi.org/10.1109/tvcg.2020.3010088.
[6] Frey G, Jurkschat A, Korkut S, Lutz J, Dornberger R. Intuitive hand gestures for the interaction with information visualizations in virtual reality. In: Augmented reality and virtual reality. Springer International Publishing; 2019, p. 261–73. http://dx.doi.org/10.1007/978-3-030-06246-0_19.
[7] Lee YS, Sohn B-S. Immersive gesture interfaces for navigation of 3D maps in HMD-based mobile virtual environments. Mob Inf Syst 2018;2018:1–11. http://dx.doi.org/10.1155/2018/2585797.
[8] Yan Y, Shi Y, Yu C, Shi Y. HeadCross: Exploring head-based crossing selection on head-mounted displays. Proc ACM Interact Mob Wearab Ubiquit Technol 2020;4(1):1–22. http://dx.doi.org/10.1145/3380983.
[9] Zheng JM, Chan KW, Gibson I. A VR based 3D graphics user interface for CAD modeling system. In: Volume 1: 25th design automation conference. American Society of Mechanical Engineers; 1999, p. 1029–36. http://dx.doi.org/10.1115/detc99/dac-8627.
[10] Monteiro P, Goncalves G, Coelho H, Melo M, Bessa M. Hands-free interaction in immersive virtual reality: A systematic review. IEEE Trans Vis Comput Graph 2021;27(5):2702–13. http://dx.doi.org/10.1109/tvcg.2021.3067687.
[11] ISO 9241-210:2019. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. Vol. 2019. Geneva, CH: International Organization for Standardization; 2019.
[12] Soukoreff RW, MacKenzie IS. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. Int J Hum-Comput Stud 2004;61(6):751–89. http://dx.doi.org/10.1016/j.ijhcs.2004.09.001.
[13] Lee Y-J, Park M, Silva D, Joo J. Power users' branded virtual reality experience: A preliminary study. J Commun Technol 2022;5(3):53–77. http://dx.doi.org/10.51548/joctec-2022-013.
[14] Monteiro P. Evaluation of hands-free VR interaction methods during a Fitts' task: Efficiency and effectiveness. IEEE Access 2023. http://dx.doi.org/10.1109/ACCESS.2023.3293057.
[15] Fuchs P. Virtual reality concepts and technologies. CRC Press; 2011, p. 409.
[16] Slater M, Wilbur S. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence: Teleoper Virt Environ 1997;6(6):603–16. http://dx.doi.org/10.1162/pres.1997.6.6.603.
[17] Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc B 2009;364(1535):3549–57. http://dx.doi.org/10.1098/rstb.2009.0138.
[18] Skarbez R, Frederick P. Brooks J, Whitton MC. A survey of presence and related concepts. ACM Comput Surv 2017;50(6):1–39. http://dx.doi.org/10.1145/3134301.

[19] Schuemie MJ, van der Straaten P, Krijn M, van der Mast CAPG. Research on presence in virtual reality: A survey. CyberPsychol Behav 2001;4(2):183–201. http://dx.doi.org/10.1089/109493101300117884.

[20] Chang E, Kim HT, Yoo B. Virtual reality sickness: A review of causes and measurements. Int J Hum Comput Interact 2020;36(17):1658–82. http://dx.doi.org/10.1080/10447318.2020.1778351.

[21] LaViola JJ. A discussion of cybersickness in virtual environments. ACM SIGCHI Bull 2000;32(1):47–56. http://dx.doi.org/10.1145/333329.333344.

[22] Jerald J. The VR book: Human-Centered design for virtual reality. Association for Computing Machinery and Morgan & Claypool; 2015.

[23] LaViola JJ, Kruijff E, McMahan RP, Bowman DA, Poupyrev I. 3D user interfaces: Theory and practice. In: Addison-Wesley usability and HCI series, Addison-Wesley; 2017.

[24] Majaranta P, Bulling A. Eye tracking and eye-based human–computer interaction. In: Human–computer interaction series. London: Springer; 2014, p. 39–65. http://dx.doi.org/10.1007/978-1-4471-6392-3_3.

[25] Mardanbegi D, Pfeiffer T. EyeMRTK. In: Proceedings of the 11th ACM symposium on eye tracking research & applications. ACM; 2019, p. 1–5. http://dx.doi.org/10.1145/3317956.3318155.

[26] Yi X, Lu Y, Cai Z, Wu Z, Wang Y, Shi Y. GazeDock: Gaze-only menu selection in virtual reality using auto-triggering peripheral menu. In: 2022 IEEE conference on virtual reality and 3D user interfaces. IEEE; 2022, p. 1–11. http://dx.doi.org/10.1109/vr51125.2022.00105.

[27] Gelsomini M, Leonardi G, Garzotto F. Embodied learning in immersive smart spaces. In: Proceedings of the 2020 CHI conference on human factors in computing systems. ACM; 2020, p. 1–14. http://dx.doi.org/10.1145/3313831.3376667.

[28] Jacob RJK. The use of eye movements in human-computer interaction techniques: What you look at is what you get. ACM Trans Inf Syst (TOIS) 1991;9(2):152–69. http://dx.doi.org/10.1145/123078.128728.

[29] Wolf D, Gugenheimer J, Combosch M, Rukzio E. Understanding the heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices. In: Proceedings of the 2020 CHI conference on human factors in computing systems. ACM; 2020, p. 1–10. http://dx.doi.org/10.1145/3313831.3376876.

[30] Mutasim AK, Batmaz AU, Stuerzlinger W. Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality. In: ACM symposium on eye tracking research and applications. ACM; 2021, p. 1–7. http://dx.doi.org/10.1145/3448018.3457998.

[31] Majaranta P, Ahola U-K, Špakov O. Fast gaze typing with an adjustable dwell time. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM; 2009, p. 357–60. http://dx.doi.org/10.1145/1518701.1518758.

[32] Cao L, Zhang H, Peng C, Hansberger JT. Real-time multimodal interaction in virtual reality - a case study with a large virtual interface. Multimedia Tools Appl 2023. http://dx.doi.org/10.1007/s11042-023-14381-6.

[33] Pai YS, Dingler T, Kunze K. Assessing hands-free interactions for VR using eye gaze and electromyography. Virt Real 2018;23(2):119–31. http://dx.doi.org/10.1007/s10055-018-0371-2.

[34] Wang K-J, Liu Q, Zhao Y, Zheng CY, Vhasure S, Liu Q, et al. Intelligent wearable virtual reality (VR) gaming controller for people with motor disabilities. In: 2018 IEEE international conference on artificial intelligence and virtual reality. IEEE; 2018, p. 161–4. http://dx.doi.org/10.1109/aivr.2018.00034.

[35] Yu D, Liang H-N, Lu X, Zhang T, Xu W. DepthMove: Leveraging head motions in the depth dimension to interact with virtual reality head-worn displays. In: 2019 IEEE international symposium on mixed and augmented reality. IEEE; 2019, p. 103–14. http://dx.doi.org/10.1109/ismar.2019.00-20.

[36] Meng X, Xu W, Liang H-N. An exploration of hands-free text selection for virtual reality head-mounted displays. In: 2022 IEEE international symposium on mixed and augmented reality. IEEE; 2022, p. 74–81. http://dx.doi.org/10.1109/ismar55827.2022.00021.

[37] Hepperle D, Weiß Y, Siess A, Wölfel M. 2D, 3D or speech? A case study on which user interface is preferable for what kind of object interaction in immersive virtual reality. Comput Graph 2019;82:321–31. http://dx.doi.org/10.1016/j.cag.2019.06.003.

[38] McClinton W, Caprio D, Laesker D, Pinto B, Garcia S, Andujar M. P300-based 3D brain painting in virtual reality. In: Extended abstracts of the 2019 CHI conference on human factors in computing systems. ACM; 2019, p. 1–6. http://dx.doi.org/10.1145/3290607.3312968.

[39] Speicher M, Cucerca S, Krüger A. VRShop: A mobile interactive virtual reality shopping environment combining the benefits of on- and offline shopping. Proc ACM Interact Mob Wearab Ubiquit Technol 2017;1(3):1–31. http://dx.doi.org/10.1145/3130967.

[40] Winkler P, Stiens P, Rauh N, Franke T, Krems J. How latency, action modality and display modality influence the sense of agency: a virtual reality study. Virt Real 2019. http://dx.doi.org/10.1007/s10055-019-00403-y.

[41] Luro FL, Sundstedt V. A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In: Proceedings of the 11th ACM symposium on eye tracking research & applications. New York, NY, USA: ACM; 2019, p. 1–9. http://dx.doi.org/10.1145/3317956.3318153.

[42] Mirzaei MA, Chardonnet J-R, Mérienne F, Genty A. Navigation and interaction in a real-scale digital mock-up using natural language and user gesture. In: Proceedings of the 2014 virtual reality international conference. New York, NY, USA: ACM Press; 2014, p. 1–4. http://dx.doi.org/10.1145/2617841.2620716.

[43] Qian YY, Teather RJ. The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In: Proceedings of the 5th symposium on spatial user interaction. New York, NY, USA: ACM Press; 2017, p. 91–8. http://dx.doi.org/10.1145/3131277.3132182.

[44] Osking H, Doucette JA. Enhancing emotional effectiveness of virtual-reality experiences with voice control interfaces. In: Beck D, Peña Rios A, Ogle T, Economou D, Mentzelopoulos M, Morgado L, Eckhardt C, Pirker J, Koitz-Hristov R, Richter J, Gütl C, Gardner M, editors. Immersive learning research network. Cham: Springer International Publishing; 2019, p. 199–209. http://dx.doi.org/10.1007/978-3-030-23089-0_15.

[45] Wei Y, Shi R, Yu D, Wang Y, Li Y, Yu L, et al. Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions. In: Proceedings of the 2023 CHI conference on human factors in computing systems. ACM; 2023, p. 1–14. http://dx.doi.org/10.1145/3544548.3581042.

[46] Lu X, Yu D, Liang H-N, Xu W, Chen Y, Li X, et al. Exploration of hands-free text entry techniques for virtual reality. In: 2020 IEEE international symposium on mixed and augmented reality. IEEE; 2020, p. 344–9. http://dx.doi.org/10.1109/ismar50242.2020.00061.

[47] Lu X, Yu D, Liang H-N, Goncalves J. Itext: Hands-free text entry on an imaginary keyboard for augmented reality systems. In: The 34th annual ACM symposium on user interface software and technology. ACM; 2021, p. 815–25. http://dx.doi.org/10.1145/3472749.3474788.

[48] Giannopoulos I, Komninos A, Garofalakis J. Natural interaction with large map interfaces in VR. In: Proceedings of the 21st pan-hellenic conference on informatics. PCI 2017, New York, NY, USA: ACM Press; 2017, p. 1–6. http://dx.doi.org/10.1145/3139367.3139424.

[49] Yan Y, Yu C, Yi X, Shi Y. HeadGesture: Hands-free input approach leveraging head movements for HMD devices. Proc ACM Interact Mob Wearab Ubiquit Technol 2018;2(4):1–23. http://dx.doi.org/10.1145/3287076.

[50] Ganapathi P, Sorathia K. Investigating controller less input methods for smartphone based virtual reality platforms. In: Proceedings of the 20th international conference on human-computer interaction with mobile devices and services adjunct. New York, NY, USA: ACM Press; 2018, p. 166–73. http://dx.doi.org/10.1145/3236112.3236136.

[51] Blattgerste J, Renner P, Pfeiffer T. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In: Proceedings of the workshop on communication by gaze interaction. New York, NY, USA: ACM Press; 2018, p. 1–9. http://dx.doi.org/10.1145/3206343.3206349.

[52] McKenzie C, Glazier A. Designing screen interfaces for VR (Google I/O '17). 2017, Google Developers. URL https://www.youtube.com/watch?v=ES9jArHRFHQ.

[53] Strasburger H, Rentschler I, Juttner M. Peripheral vision and pattern recognition: A review. J Vis 2011;11(5). http://dx.doi.org/10.1167/11.5.13.

[54] Purwar S. Designing user experience for virtual reality (VR) applications. 2019, URL https://uxplanet.org/designing-user-experience-for-virtual-reality-vr-applications-fc8e4faadd96.

[55] Schubert T, Friedmann F, Regenbrecht H. The experience of presence: Factor analytic insights. Presence: Teleoper Virt Environ 2001;10(3):266–81. http://dx.doi.org/10.1162/105474601300343603.

[56] Vasconcelos-Raposo J, Bessa M, Melo M, Barbosa L, Rodrigues R, Teixeira CM, et al. Adaptation and validation of the igroup presence questionnaire (IPQ) in a portuguese sample. Presence: Teleoper Virt Environ 2016;25(3):191–203. http://dx.doi.org/10.1162/pres_a_00261.

[57] Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993;3(3):203–20. http://dx.doi.org/10.1207/s15327108ijap0303_3.

[58] Brown P, Spronck P, Powell W. The simulator sickness questionnaire, and the erroneous zero baseline assumption. Front Virt Real 2022;3. http://dx.doi.org/10.3389/frvir.2022.945800.

[59] Brooke J. SUS: A quick and dirty usability scale. In: Usability evaluation in industry. CRC Press; 1995, p. 189–94.

[60] Sauro J. Measuring usability with the system usability scale (SUS). 2011, MeasuringU. URL https://measuringu.com/sus/.

[61] Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usab Stud 2009;4(3):114–23.

[62] Hart SG, Staveland LE. Development of NASA-TLX (task load index): Results of empirical and theoretical research. 1988.

[63] Pickup L, Wilson JR, Norris BJ, Mitchell L, Morrisroe G. The integrated workload scale (IWS): A new self-report tool to assess railway signaller workload. Applied Ergon 2005;36(6):681–93. http://dx.doi.org/10.1016/j.apergo.2005.05.004.

[64] Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: Four longitudinal field studies. Manage Sci 2000;46(2):186–204. http://dx.doi.org/10.1287/mnsc.46.2.186.11926.

[65] Disztinger P, Schlögl S, Groth A. Technology acceptance of virtual reality for travel planning. In: Information and communication technologies in tourism 2017. Springer International Publishing; 2017, p. 255–68. http://dx.doi.org/10.1007/978-3-319-51168-9_19.

[66] Cohen J. Statistical power analysis for the behavioral sciences. second ed.. Routledge; 1988, http://dx.doi.org/10.4324/9780203771587.

[67] Melo M, Gonçalves G, Vasconcelos-Raposo J, Bessa M. How much presence is enough? Qualitative scales for interpreting the igroup presence questionnaire score. IEEE Access 2023;11:24675–85. http://dx.doi.org/10.1109/access.2023.3254892.

[68] Alallah F, Neshati A, Sakamoto Y, Hasan K, Lank E, Bunt A, et al. Performer vs. observer. In: Proceedings of the 24th ACM symposium on virtual reality software and technology. ACM; 2018, p. 1–9. http://dx.doi.org/10.1145/3281505.3281541.

[69] Xu W, Liang H-N, Zhao Y, Yu D, Monteiro D. DMove. In: Proceedings of the 2019 CHI conference on human factors in computing systems. ACM; 2019, p. 1–14. http://dx.doi.org/10.1145/3290605.3300674.