IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 June 2023, accepted 4 July 2023, date of publication 7 July 2023, date of current version 18 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3293057

RESEARCH ARTICLE

Evaluation of Hands-Free VR Interaction Methods During a Fitts' Task: Efficiency and Effectiveness

PEDRO MONTEIRO, GUILHERME GONÇALVES, BRUNO PEIXOTO, MIGUEL MELO, AND MAXIMINO BESSA

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

Corresponding author: Pedro Monteiro (monteiro.p@outlook.pt)

This work was supported in part by the National and European Funds through the Portuguese Funding Agency, FCT—Fundaço para a Ciãncia e a Tecnologia, under Project SFRH/BD/147813/2019 and Project UIDB/50014/2020.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Currently, it is standard to use tracked handheld controllers for interaction in immersive virtual reality (VR). However, since VR interactions are becoming more natural with hand tracking, it is important to provide hands-free alternatives for selection and system control tasks. As such, this study aims to provide an exploratory evaluation of the effectiveness and efficiency of commonly used hands-free interfaces in selection and system control tasks. Nine interaction methods were evaluated while performing a Fitts' law task with nine advanced users of VR in a within-subject experiment. We evaluated handheld controllers as a baseline, against head gaze, eye gaze, and voice commands for pointing at the targets, and dwell time and voice commands to confirm selections. We found that using eye gaze with a 500 ms dwell time proved to be the hand-free method with the highest performance, matching the handheld controllers and being preferred by users. The evaluation also showed that using a multimodal approach to selection, especially using the voice, decreases performance, but increases effectiveness. Moreover, we verified that Fitts' law can be applied to hands-free methods, but its usage is limited when the methods have very short travel times. We then suggest selections per minute as a more robust comparative performance metric. Further studies should expand the audience and interaction tasks and focus on the confirmatory method of selection.

INDEX TERMS Hands-free, HCI, immersive virtual reality, interaction, usability.

I. INTRODUCTION

Virtual reality (VR) has become an increasingly popular platform for a wide range of applications, from entertainment and gaming to education and training [1]. Immersive VR, in particular, has become a recent hot topic with the emergence of the metaverse, allowing users to immerse themselves in a virtual environment and interact in real time [2]. This usually comprises a headset with handheld controllers that track the position and rotation of the users' heads and hands, thereby allowing for a sense of presence in the virtual environment.

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

Handheld controllers are the most used method of interaction with virtual environments, as they are bundled with most VR systems and use an interaction metaphor that is intuitive to users. For interactions with graphical user interfaces (GUIs), these are used as a point-and-click metaphor where users can make the controller a pointer and use a button to act on the element being pointed at [3].

In a future where immersive VR experiences push the limits of fidelity and realism [4] and with constant technological advances made to provide better, more natural, and intuitive experiences to users, interactions will tend to mirror their real-world counterparts, especially interactions with objects and other users. This results in the study of

alternative methods to overcome the problem of not having a controller for interaction, such as the use of hand gestures [5] or other non-hand interaction methods [6], [7]. We consider hand interactions for system control tasks to be disruptive to other interactions that naturally use the hands; for instance, if a user needs to perform a control task while grabbing an object, it first needs to release the object to be able to use the hand for that task. For this reason, and since literature is not consensual regarding the classification of hand gestures as a hands-free method (e.g., [6], [8]), this study does not consider hand gestures as hands-free interactions. Therefore, there is a need to consider and evaluate novel hands-free interaction methods.

As found by [6], there are an increasing number of studies that explore the use of hands-free interaction methods in immersive VR, but the literature lacks a comprehensive evaluation of these methods and consequently a clear comparison of the advantages and disadvantages of the methods compared to each other. The study also found that voice, eyes, and head are the most studied interaction sources for immersive VR in selection and system control tasks.

The Fitts law is widely recognized [9], [10] in the field of human-computer interaction (HCI) as a metric to evaluate the selection performance (and consequently the efficiency) of interaction methods, which can also be applied in immersive VR [11]. Furthermore, effectiveness is usually evaluated by completion or error rates [6].

As such, this study aims to contribute to this field by providing an exploratory evaluation of the effectiveness and efficiency of the most commonly used hands-free interfaces when performing a Fitts' law task. Given the exploratory nature of the study, the evaluation will target advanced users, who are generally testers and adopters of new technologies. We intend to answer the following research questions:

RQ1 How is the effectiveness of users affected by the interaction method?
RQ1.1 Can hands-free methods be more effective than handheld controllers?
RQ1.2 Does a multimodal approach lead to better effectiveness?
RQ2 How is the efficiency of users affected by the interaction method?
RQ2.1 Can hands-free methods match the efficiency of handheld controllers?
RQ2.2 Does a multimodal approach lead to better efficiency?

To accomplish this goal and answer the research questions, an evaluation testbed was developed in which both interaction methods and evaluation modules can be added and configured according to the testing requirements. Ideally, the results of this study contribute to the search for better, more reliable, and more comprehensive methodologies for the evaluation of interaction in immersive VR experiences.

II. RELATED WORK

A. HANDS-FREE SELECTION IN VR

Immersive VR refers to VR experiences that provide users with a high level of immersion using immersive devices. This allows users to feel fully present [12] in a virtual environment while interacting with its contents in real time [13]. Immersive VR systems typically consist of a head-mounted display (HMD) for visual stimulation and headphones for audio. User interactions in virtual environments are diverse [2], [14], with common tasks in immersive VR including system control, selection, navigation, and manipulation.

Although the most used interface is the handheld controllers (which are included in commonly sold VR systems such as the Meta Quest and HTC Vive), advances in HCI make novel and advanced interaction methods possible. To improve user-centered interactions [14] and increase the sense of presence in immersive VR systems, hand recognition is used. This allows an accurate virtual representation of users' hands and allows for more natural interactions through hand gestures that mimic real-world actions [15]. However, the use of hands for object interactions can limit their simultaneous use in other tasks, such as system control tasks, which involve selecting GUI elements.

Selection typically involves a two-step procedure in which, first, there is a target acquisition (also known as aiming or pointing) and second, target confirmation (also known as activation) [16]. While handheld controllers are a prime method to perform this task as they support aiming and have buttons for confirmation, theoretically any method that provides a direction can be used for aiming, and any method that can trigger a single action can be used for confirmation.

For instance, eye and head gaze can serve as effective pointing methods to select GUI elements and objects [7], [17], [18]. However, when the interface does not have a clear confirmation method, confirming the selection can be challenging. To address this, the fixation time on the target (dwelling) can be used to confirm the selection. Because the eyes and the head move naturally, it is important to differentiate between the resting gaze and the gaze intended for interaction to avoid the ''Midas Touch'' problem [19]. To mitigate this issue, different dwell times can be used, and studies have found that fixation times ranging from 300 ms to 1000 ms are suitable depending on the level of expertise of the users, allowing high performance in selection tasks [20], [21].

In the selection process, a multimodal interaction approach is possible because it consists of two phases. For example, alternative methods can be employed instead of dwell time for confirmation. A widely used approach is the use of voice commands, in which users can confirm a target or specify their intent before or after pointing at the target [22]. The use of voice commands has the added benefit of enabling selection or system control without the need for a pointing phase [23], allowing a single command to perform the desired action. Less explored methods include brain-computer interfaces (BCI) [24], muscle activity [25], face expressions [26], and body actions [27], which have also been shown to be an

alternative for hands-free interaction in VR. However, care must be taken to avoid the use of a confirmation method that influences pointing and leads to incorrect confirmations (Heisenberg errors) [28].

The next section provides an overview of the evaluation of these interfaces for the selection and system control tasks.

B. EVALUATION FOR SELECTION AND SYSTEM CONTROL TASKS

The evaluation of interaction methods in VR usually focuses on assessing usability (user satisfaction, efficiency, and effectiveness) and system performance. Furthermore, the mental and physical states of users after using these methods can be evaluated using VR experience metrics. Although HCI studies have already explored hands-free interaction methods, [6] found that the evaluation of these methods lacks a formal methodology and that studies usually do not provide comprehensive comparisons of interaction methods for interaction tasks, instead tailoring the evaluations to the interaction method being used.

In a study by [25], the use of eye gaze with myography as an input method for selection tasks in VR was compared to two types of handheld controllers (stationary and tracked) and head or eye gaze with a 750 ms dwell time. Researchers evaluated the efficiency and cognitive load of these methods and found that the novel eye gaze with myography method was more efficient than the other hands-free methods, except for tracked controllers.

Regarding pointing methods, [29] found that eye gaze was the most efficient method, while [11] and [30] reported that head gaze was more efficient.

In [7] the use of eye gaze was studied to select items from a menu arranged on the periphery of the field of view (FOV). The study found that the technique was able to outperform the dwelling and pursuit techniques and overcome the problems associated with false triggering of the menu and false confirmations.

In a study by [31], the use of head gaze was compared with speech for searching for products in a shopping environment. The study found that the task time and error rate were higher with head-pointing and lower with speech.

A study by [32] compared head-gaze selection techniques with non-hands-free interfaces and found that the researchers' head-gaze technique was efficient, whereas hand gestures were fatiguing.

Several studies have shown that head and eye gaze techniques provide better efficiency than handheld controllers for selection purposes [33], [34], [35]. However, the results vary in terms of accuracy and efficacy. In the context of smartphone VR, [36] found that touching capacitive buttons is more efficient than head-gazing with dwell time.

[37] found that blinking is a viable hands-free text selection solution and, as such, it is recommended as the default option when an eye tracker is available, as it has the best performance and a low error rate. Head gaze with dwell is an acceptable alternative when an eye tracker is unavailable, and voice input should be avoided because of its poor performance.

In these studies, effectiveness was mostly measured using objective metrics such as error rates [7], [29], [31] or task times [23], [33], [37]. On the other hand, effectiveness is measured mainly by evaluating user performance using the methods. Table 1 shows a summary of studies that evaluated selection tasks concerning the efficiency and effectiveness of at least one hands-free method. A robust and reliable performance metric is Fitts' law. Because the use of an objective and well-understood performance metric is pertinent, this metric is used to compare interaction methods [11], [25], [38].

C. FITTS' LAW FOR SELECTION PERFORMANCE EVALUATION IN VR

The Fitts' law [39], [40] is a predictive model that describes the relationship between the size and distance of the targets and the time it takes to move and select the target. It states that the time taken to move to a target is proportional to the distance to the target and inversely proportional to the size of the target. This law is often applied in HCI and user interface design to optimize the placement and sizing of interactive elements. Although it was originally developed for a one-dimensional selection task [41], many successful extensions have been made to adapt it to 2D and 3D tasks, despite its shortcomings [9], [10], [11], [25], [38], [41], [42], [43], [44].

The original law [39] describes the linear relationship between the Index of Difficulty of a target (ID) and the movement time (MT) shown in Equation 1, where a and b are empirical constants determined by linear regression. The ID (Equation 2), which has bits as the unit, considers the amplitude (A) of movement and the width (W) of the target.

$$\mathrm{MT} = a + b\,\mathrm{ID} \tag{1}$$

$$\mathrm{ID} = \log_2\left(\frac{2A}{W}\right) \tag{2}$$

However, as stated in [10] and [41], Equation 3 for ID is commonly accepted for pointing tasks. Furthermore, the performance throughput (TP) can be calculated using Equation 4.

$$\mathrm{ID} = \log_2\left(\frac{A}{W} + 1\right) \tag{3}$$

$$\mathrm{TP} = \frac{\mathrm{ID}}{\mathrm{MT}} \tag{4}$$

It has been shown that Fitts' law can be used for gaze interactions in VR environments while using Equation 3 as is [38] or replacing W by an effective (W_e) target width [25], [42], aligning with the procedures found in ISO 9241 [9]. Other variations include angular interpretations of ID [11] or polar coordinate systems [43]. Despite the multiple existing models, when evaluating a "point and click" task for a GUI, the recommendations found in [9] should be followed [10].

TABLE 1. Summary of studies of efficiency and effectiveness evaluation of selection task in immersive VR.

Study	Metrics	Methods Evaluated
[7]	Task Time; False Positives; False negatives	Eye Gaze
[11]	Fitts' Law; Movement Time; Error Rate	Eye Gaze; Head Gaze
[23]	Task Time; Errors	Voice; Controller
[25]	Fitts' Law	Eye Gaze; Arm Force; Head Gaze; Controllers
[29]	Task Times; Errors	Eye Gaze; Head Gaze
[30]	Pointing Time	Eye Gaze; Head Gaze; Controller
[31]	Task Times; Error Rate	Head Gaze; Voice
[32]	Pointer trajectory; Unintentional Passes; Speed	Head Gaze; Controller; Hand Gestures
[33]	Number of Actions; Task Times	Head Gaze; Controller
[34]	Task Times; Angle Between Rays and Targets; Failed Aims	Eye Gaze; Controller
[35]	Speed; Accuracy	Head Gaze; Hand Gestures
[36]	Task Times; Error Rate	Head Gaze; Button
[37]	Task Time; Error Rate	Eye Gaze; Voice; Neck Gesture
[38]	Fitts' Law; Task Times; Reaction Time; Error Rate	Voice (Non-verbal); Controller; Hand Gestures

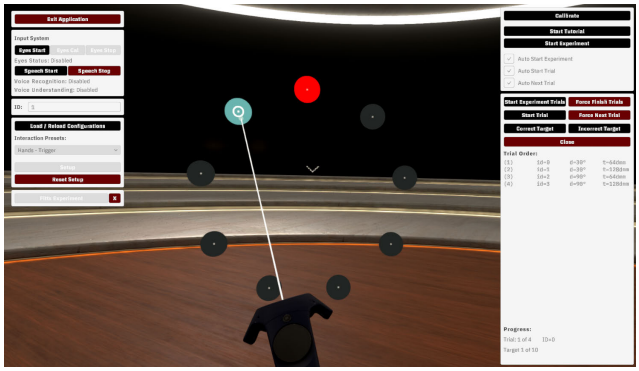


FIGURE 1. Screenshot of the application configured for the handheld controller's interaction and with an ongoing Fitts task.

III. TESTBED APPLICATION

An application was developed to enable this study and its conditions (Fig. 1). The Unity 2022.2 game engine was used as it provided faster integration with the VR platform. Unity's use of C# as its programming language also provided great compatibility with the SDKs of the interaction devices. The OpenXR standard was used to interface with the VR devices where possible, mainly allowing the application to be agnostic to the VR setup.

A. INTERACTION COMPONENTS

The application was developed as a test bed for multiple interaction experiments. Thus, it provides a highly customizable and modular implementation of different interactions. A single interaction method is defined by five components that are aligned with Unity's XR Interaction Plugin but tailored to the application:

- **Origin:** The origin is the point of reference that is used to control the position and rotation of the interaction. Origins coincide with different body parts (e.g., hands, eyes, head, or the whole body);
- **Controller:** Component responsible for mapping the input events and triggering interaction events in the application. Two examples of controllers are the "Left Hand Trigger" which translates a press of the left handheld controller trigger button to a "select" event; or the "Voice Actions" controller which interprets an affirmative word from the user's speech to a "select" command. For voice detection, the Azure Cognitive Services[1] and Wit.ai[2] are supported as both speech-to-text and text to action services;
- **Interactor:** Component that listens to interaction events and triggers state changes in the target objects that are being interacted with. Also used to query which targets are interactable and the state of the interaction (e.g., hovering and selecting). For pointing, target acquisition can be performed by ray casts and sphere casts (i.e., a ray cast with a sphere detection area), which are configurable (e.g., sphere diameter and ray smoothing). An example of a used non-pointing method is the Tobii G2OM[3] which processes eye gaze and through machine learning detects which object is being looked at;
- **Visual:** Component that gives a visual representation of the interactor. For instance, when using ray casts, the ray line is displayed depending on the interaction state;
- **Reticle:** This is a visual component that is displayed on the target of the interaction and is mainly used to display a cursor so users can better perceive where they are pointing. This component can also respond to the different states of interaction.

These components can be configured and configured to create different interaction methods, and their configurations can then be saved as JSON files, allowing further customization without the need to recompile the application.

B. FITTS' TASK

To evaluate the interaction methods, a Fitts' law task was used. This consists of a consecutive selection of a series of spherical targets displayed in a circular pattern [9] with varying sizes and distances between them.

[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/

For this task, a module was added to the application, which allowed the configuration of the trials. A trial is a specific configuration of the following three factors: First, target sizes are defined by their diameter measured in distance-independent millimeters (dmm), which is one millimeter at a distance of one meter from the point of view; Second, because the targets were arranged in a ring (see Fig. 1), the angle amplitude of the ring defines how far from the center point and each other the targets are, and consequently their distance from each other; Third, since the targets are evenly distributed in the circle, the number of targets to select controls the direction angle of each target (azimuth angle), which defines the position of the target on the ring. The first target was always vertically aligned at a 90-degree angle.

Developers can configure how the targets respond to user interaction, e.g. the color of the different target states and reticles. Additionally, the center of the circle was placed 6° below the line of sight parallel to the ground, as this is the angle of the resting eye level. The module also supports a tutorial mode in which a small subset of trials is used.

This module is also responsible for recording all the data required to evaluate the interaction methods; more specifically, the experiment time, the selection time for each target, the number of times a target is hovered (correctly or incorrectly), the number of times a target is incorrectly confirmed, the position of the cursor during the experiment (if a cursor exists), and the number of voice activations, actions, and errors. These data were recorded as a CSV file for each participant trial.

IV. METHODOLOGY

A. SAMPLE

Given the exploratory nature of the study and its target sample of advanced VR users, the experiment was performed by nine male participants aged 23 to 34 ($M = 26.6$, $SD = 3.36$). All participants were volunteers recruited at the laboratory where the experiments were performed and consisted of highly technically educated personnel who are accustomed to VR technologies. However, they had no prior contact with the specific implementation of the technologies used in this study.

B. APPARATUS

In this study, an HTC Vive Pro Eye was used as the immersive VR system. This system uses SteamVR 2.0 tracking and encompasses (1) an HMD responsible for the delivery of the visual stimulus with a 110° FOV, a per-eye resolution of 1440 × 1600 pixels, and 90 Hz target refresh rate; (2) two tracked handheld controllers for hand interaction; (3) an embedded Tobii® eye tracker with a 120 Hz sampling rate, a 110° tracking range, and down to 0.5° accuracy within a 20° FOV; and (4) an embedded close-range microphone for voice interaction. Additionally, the audio stimulus was delivered using the sound system of the experimental room.

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

C. INDEPENDENT VARIABLES

1) INTERACTION METHODS

The interaction method was used as an independent variable. Given the capabilities of the developed application, a preliminary test was conducted to anecdotally identify the interaction methods that made sense for users while following common and best implementation practices. The resulting final interaction methods are:

C – Controllers: Handheld controllers were used for pointing at the targets and the trigger button to confirm the selection. Smoothing was not applied to the cursor, which was a single point (1 dmm).

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

HD – Head + Dwell: The pointing and cursor were the same as in HV, while the confirmation used a 500 ms dwell time.

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

ED – Eyes + Dwell: The pointing and cursor were the same as in EV, while the confirmation used a 500 ms dwell time.

AEV – Assisted Eyes + Voice: Confirmation similar to EV; however, instead of using the smoothed gaze to point, the hovered object was obtained via the Tobii G2OM. Participants could still see the smoothed gaze cursor.

AED – Assisted Eyes + Dwell: Confirmation similar to ED, but as in AEV, instead of using the smoothed gaze to point, the hovered object was obtained via the Tobii G2OM. Participants could still see the smoothed gaze cursor.

V – Voice (Direct): The full selection process was performed with a single voice command, in this case, the number displayed on the targets.

VC – Voice (with confirmation): To perform the selection, the participants required two voice commands. The first is the same as V, resulting in the target being hovered. Subsequently, a second confirming voice command had to be issued.

2) FITTS PARAMETERS AND INDEXES OF DIFFICULTY

The three factors that define the targets were configured according to the general guidelines in [45]:

It is recommended that the minimum hit size of a target be 64 dmm and the comfortable size be 96 dmm, both with 16 dmm padding. Thus, we used four target sizes: 32 dmm (a bellow minimum size, for extra difficulty), 64 dmm, 96 dmm, and 128 dmm (as an over-comfortable size). The sizes had a constant increment value of 32 dmm from the smallest to the largest size.

The amplitude angles of the rings were derived from human motion and viewing zones. The motion range zones of the head (horizontal × vertical) are considered comfortable within the ranges of 60° × 30° up to a maximum motion of 110° × 100°. The eye view zones [46] define central vision within 30°, near peripheral vision within 60°, medium peripheral vision with a 120° range, and 200° for far peripheral vision. As such, we used five ring amplitude angles (Fig. 2b): 30° (central vision), 60° (central vision with comfortable head rotation), 90° (comfortable head rotation with central vision and within the HMD FOV), 120° (high head rotation with peripheral vision), and 150° (highly outside of comfortable ranges with both head and eye rotations combined).

For each trial (combination of target diameter and ring amplitude), the participants completed 10 target selections in a clockwise star pattern, resulting in nine equally spaced targets with the following direction angles (Fig. 2): 10°, 50°, 90°, 130°, 170°, 210°, 250°, 290°, and 330°.

Using Equation 3, we obtained the following IDs: 2.33, 2.68, 3.14, 3.18, 3.51, 3.59, 3.86, 3.98, 4.01, 4.06, 4.1, 4.25, 4.4, 4.53, 4.81, 4.96, 5.01, 5.5, 5.78, 5.94.

Finally, the targets were placed at a constant distance of 1 m from the participant's point of view, since this is a distance outside of the ''no zone'' and within the distance where the foreground content is placed [47]. Independent of the placement and size of the targets, the distance between them was always greater than 16 dmm.

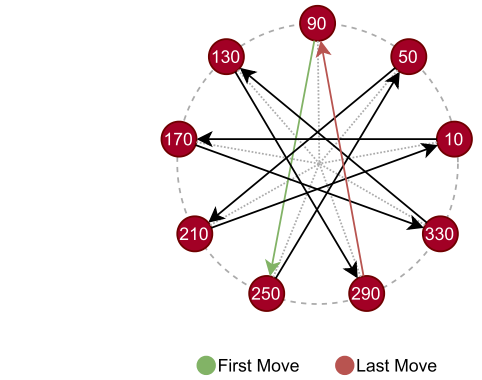
The order of the trials was randomized for each participant and interaction method. In total, to complete the task, each participant had to perform 200 confirmations (4 target sizes × 5 ring amplitude angles × 10 targets) for each of the interaction methods.

D. DEPENDENT VARIABLES

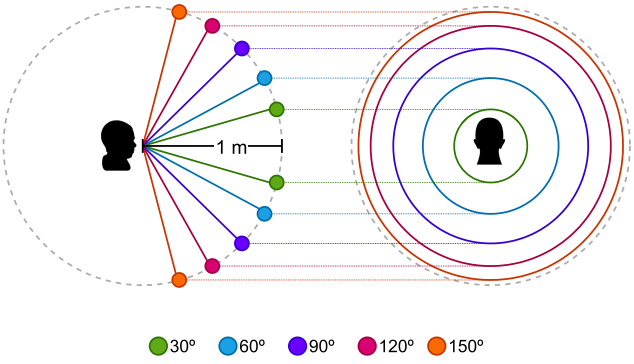
The dependent variables were the objective metrics of effectiveness and efficiency: the number of times a correct target was hovered, the number of times an incorrect target was hovered, the number of times an incorrect target was confirmed, the total experiment time, the number of selections per minute, Fitts' performance throughput, and Fitts' regression model.

E. DESIGN AND PROCEDURE

Each participant was assigned a letter anonymizing the collected data and was required to attend the laboratory for 9 consecutive work days. The experiments were carried out in a laboratory environment in which participants were isolated from external factors. Only the researcher conducting the experiment and the participant were present in the room.



(a) Direction angles with the target selection pattern.



(b) Side and front view of the ring amplitude angles.

FIGURE 2. Factors for the target placement.

To volunteer to participate in the experiment, participants had to first fill out an informed consent form providing information on the experiment's goals and their rights. Participants could withdraw from the experiment without penalization. After accepting to participate, participants filled out a brief sociodemographic questionnaire to characterize the sample. No information that allowed a direct connection between a participant and the data was recorded.

The experiment was performed using a within-subjects design, with only one of the interaction methods evaluated per participant on a single day. For a single evaluation day, a participant used one of the interaction methods and performed the Fitts' task following the procedure described below. The method was randomly assigned on the day and previously used methods were excluded. During the evaluation, each participant made a total of 2160 correct selections (9 methods × 200 selections in the experiment plus 9 methods × 2 target sizes × 2 ring amplitude angles × 10 targets in the tutorial).

Participants were equipped with the HMD and with one handheld controller (if using the C method). In the VR environment, a calibration procedure was performed to ensure the correct calibration of the eye tracker for the participant. This eye tracker calibration was performed regardless of the method being evaluated because it also ensured the correct placement of the HMD on the face for optimal viewing

conditions. We must note that participants who wore glasses were required to not use them since tests with them showed that the eye tracker did not work correctly. These people were only accepted to participate if they stated that not using glasses would not penalize their experience and well-being. A second calibration was performed to obtain the participant's height and calibrate the origin of the targets to their viewing point.

First, participants went through a tutorial phase in which they were able to test and get used to the interaction method and task. The researcher also helped and explained the task and the importance of completing the task as quickly as possible. This phase was also important for participants to understand how the interaction method responded to their input and how the application provided feedback on the interaction. Participants were not informed about the differences between the EV/AEV and ED/AED methods. With the tutorial finished and with the approval of the participant, the main experiment task was performed.

After completing the task, the VR equipment was removed and a brief interview was conducted to gather anecdotal feedback from the participants about the interaction method they used. The entire session was approximately 30 minutes long. A further interview was conducted after the last interaction method to ask participants their views of the overall experiment and interaction methods.

F. DATA ANALYSIS

Due to the sample size and nature of the data, we were only able to remove outliers of target selection time by the index of difficulty for each interaction method. Outliers were removed through Z-score filtering, with a data point being considered an outlier if it is above or below the mean for more than 3 standard deviations. This resulted in the removal of 908 data points. After this, the Shapiro-Wilk test was performed for each condition group and dependent variables and showed that the data did not follow a normal distribution ($p < 0.05$).

As such, the groups were compared using the Friedman test as an alternative to repeated-measures ANOVA, given its tolerance to outliers and data that do not follow a normal distribution. Contrary to repeated-measures ANOVA, the Friedman test uses the ranks of the values of each group for comparison. To better interpret the results in light of the small sample size, Kendall's coefficient of concordance (W) is also reported as a measure of effect size and is classified according to Cohen's interpretation guidelines [48] as small (< 0.3), medium ($0.3 - 0.5$) and large (> 0.5).

Whenever the Friedman test produced statistically significant values, a post-hoc analysis was performed to understand which groups were statistically different using the Conover test with Bonferroni correction given its conservative nature of type I errors. Only statistically significant differences were reported for this post-hoc analysis.

Regarding Fitts' law, Spearman correlations were used to assess the impact of the index of difficulty factors on target selection times, and simple linear regressions were performed following the model with Equation 1.

Statistical procedures were performed using the RStudio[4] software with packages that support the required statistical tests and data visualization. The level of significance was maintained at 95% (alpha level of 0.05) for all statistical tests.

TABLE 2. Descriptive statistics for the number of times a target was interacted with.

Interaction	Method	M	SD	Mdn	IQR	Min	Max
Correct Target Hovers	C	326.444	54.118	312	36	280	454
	HV	250.444	22.995	242	23	220	292
	HD	240.111	21.705	240	17	211	287
	EV	305.333	49.533	307	19	234	415
	ED	402.778	217.220	269	296	227	787
	AEV	302.222	243.071	219	31	203	949
	AED	290.222	123.944	201	156	200	532
	V	200.000	0.000	200	0	200	200
	VC	200.111	0.333	200	0	200	201
Incorrect Target Hovers	C	53.111	46.015	39	13	21	168
	HV	1.444	0.882	1	1	0	3
	HD	1.444	1.333	1	1	0	4
	EV	9.444	5.411	8	10	2	17
	ED	24.444	18.365	29	21	2	56
	AEV	18.222	21.212	12	6	3	72
	AED	27.333	37.822	8	29	1	114
	V	0.444	1.014	0	0	0	3
	VC	0.556	0.882	0	1	0	2
Incorrect Target Confirmations	C	30.444	16.957	26	16	15	62
	HV	0.000	0.000	0	0	0	0
	HD	0.222	0.441	0	0	0	1
	EV	0.333	0.500	0	1	0	1
	ED	1.333	1.936	0	2	0	5
	AEV	1.000	0.866	1	2	0	2
	AED	0.667	0.500	1	1	0	1
	V	0.444	1.014	0	0	0	3
	VC	5.778	7.981	3	3	1	26

V. RESULTS

Given the amount of data, the results from Fitts' law were separated from the others.

A. GENERAL EFFECTIVENESS AND EFFICIENCY

1) TARGET INTERACTIONS

The descriptive statistics of the number of times a target was interacted with are shown in Table 2, in Fig. 3 for hovers, and in Fig. 4 for confirmations.

Regarding hovers on the correct target, statistically significant differences were found when comparing interaction methods, $\chi^2(8) = 47.423$, $p < 0.001$, $W = 0.659$. Post-hoc pairwise comparisons revealed that:

- V ($Mdn = 200$) had significantly lower correct hovers than C ($Mdn = 312$, $p < 0.001$), EV ($Mdn = 307$, $p = 0.003$), and ED ($Mdn = 269$, $p = 0.011$);
- VC ($Mdn = 200$) had significantly lower correct hovers than C ($Mdn = 312$, $p = 0.001$), EV ($Mdn = 307$, $p = 0.004$), and ED ($Mdn = 269$, $p = 0.016$);

Regarding the hovers of incorrect targets, statistically significant differences were also found between the methods: $\chi^2(8) = 57.896$, $p < 0.001$, $W = 0.804$. The post-hoc analysis showed that:

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

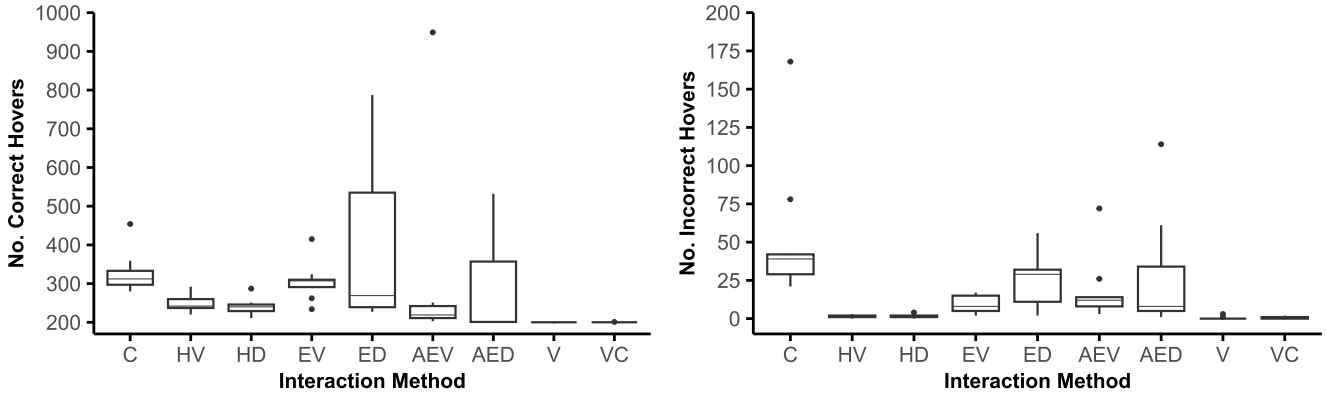


FIGURE 3. Number of times a correct (left) or incorrect (right) target was hovered by interaction method.

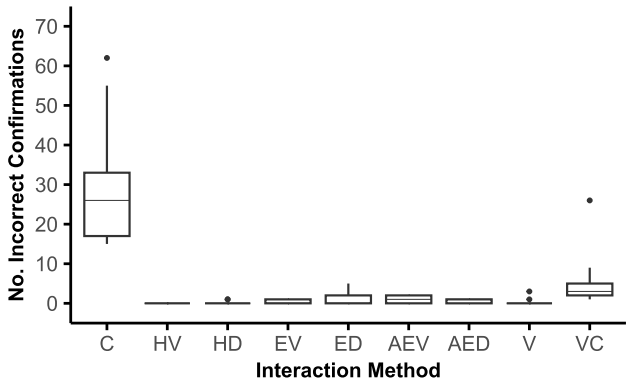


FIGURE 4. Number of times an incorrect target was confirmed by interaction method.

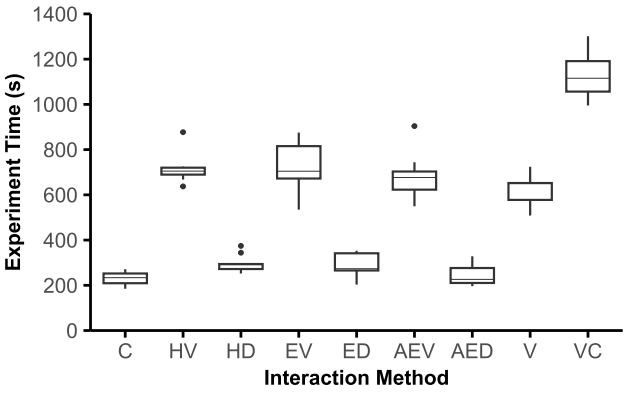


FIGURE 5. Experiment time (in seconds) for each of the interaction methods.

- C ($Mdn = 39$) had significantly higher incorrect hovers than HV ($Mdn = 1$, $p = 0.005$), HD ($Mdn = 1$, $p = 0.007$), V ($Mdn = 0$, $p < 0.001$), and VC ($Mdn = 0$, $p < 0.001$);
- ED ($Mdn = 29$) had significantly higher incorrect hovers than V ($Mdn = 0$, $p = 0.01$) and VC ($Mdn = 0$, $p = 0.015$);
- AEV ($Mdn = 12$) had significantly higher incorrect hovers than V ($Mdn = 0$, $p = 0.03$) and VC ($Mdn = 0$, $p = 0.044$);
- AED ($Mdn = 8$) had significantly higher incorrect hovers than V ($Mdn = 0$, $p = 0.02$) and VC ($Mdn = 0$, $p = 0.03$);

Finally, when comparing the number of times a confirmation was wrongly performed, statistically significant differences were also found ($\chi^2(8) = 46.325$, $p < 0.001$, $W = 0.643$), more specifically:

- C ($Mdn = 26$) had significantly more wrong confirmations than HV ($Mdn = 0$, $p < 0.001$), HD ($Mdn = 0$, $p = 0.002$), EV ($Mdn = 0$, $p = 0.003$), ED ($Mdn = 0$, $p = 0.045$), AED ($Mdn = 1$, $p = 0.045$), and V ($Mdn = 0$, $p = 0.003$).
- VC ($Mdn = 3$) had significantly more incorrect confirmations than HV ($Mdn = 0$, $p = 0.008$).

TABLE 3. Descriptive statistics for experiment time (in seconds).

Method	M	SD	Mdn	IQR	Min	Max
C	230.977	31.444	233.669	43.104	184.744	271.196
HV	715.838	66.696	704.880	30.655	637.312	876.988
HD	296.265	39.327	291.243	22.123	252.238	374.295
EV	729.639	107.789	704.395	143.252	534.626	875.252
ED	287.878	53.183	273.381	75.967	203.170	352.806
AEV	686.176	99.745	676.646	80.224	549.365	904.028
AED	241.986	45.893	226.149	66.062	196.292	328.228
V	624.879	70.840	649.817	74.686	508.479	724.159
VC	1134.391	106.883	1115.661	134.687	995.109	1301.293

2) EXPERIMENT TIME

The experiment time was compared between each interaction method. Table 3 lists the descriptive values for the time (in seconds) of each method. Additionally, Fig. 5 shows this data graphically.

Analysis shows that the experiment time was significantly different between the interaction methods, $\chi^2(8) = 65.067$, $p < 0.001$, $W = 0.904$. Generally, a significant difference was found between the methods with and without voice, namely:

- C ($Mdn = 233.669$) was significantly lower than HV ($Mdn = 704.880$, $p = 0.004$), EV ($Mdn = 704.395$, $p = 0.007$), AEV ($Mdn = 676.646$, $p = 0.043$), and VC ($Mdn = 1115.661$, $p < 0.001$);

- HD ($Mdn = 291.243$) was significantly lower than VC ($Mdn = 1115.661$, $p = 0.004$);
- ED ($Mdn = 273.381$) was significantly lower than VC ($Mdn = 1115.661$, $p < 0.002$);
- AED ($Mdn = 226.149$) was significantly lower than HV ($Mdn = 704.880$, $p = 0.005$), EV ($Mdn = 704.395$, $p = 0.009$), and VC ($Mdn = 1115.661$, $p < 0.001$).

3) TARGET SELECTION TIMES

The time a user spent confirming the correct target (in seconds) using a certain interaction method is shown in Fig. 6. Results show that methods that use voice generally have higher target times than the others. In addition, except for methods that only use the voice (V and VC), the larger the ring angle, the longer the time needed to confirm the targets. Moreover, to a lesser extent, when the target diameter increases, the time decreases.

4) SELECTIONS PER MINUTE AND THROUGHPUT

To evaluate the overall performance of the methods, the number of selections per minute was derived from the target selection times, and the throughput was calculated using Equation 4. These data are shown in Table 4 and Fig. 7.

TABLE 4. Descriptive statistics for the number of selections per minute and respective mean performance throughput for each method.

Method	M	SD	Mdn	IQR	Min	Max	TP
C	66.235	9.808	61.283	13.534	53.856	81.610	4.245
HV	17.949	1.415	17.975	1.271	14.729	19.350	1.226
HD	43.614	4.105	43.297	3.207	36.589	49.788	2.913
EV	18.328	2.283	18.198	3.112	16.129	23.135	1.248
ED	51.031	6.528	49.632	2.380	42.647	64.016	3.353
AEV	19.230	1.674	19.251	1.020	16.591	22.069	1.313
AED	56.886	5.260	56.079	7.263	49.549	64.977	3.794
V	21.879	2.049	21.403	1.815	18.533	25.781	1.522
VC	11.489	0.627	11.358	0.869	10.785	12.657	0.799

Statistically significant differences were found when comparing the number of selections per minute between methods: $\chi^2(8) = 68.415$, $p < 0.001$, $W = 0.95$. Analogously, the performance throughput was also significantly different between the methods: $\chi^2(8) = 68.385$, $p < 0.001$, $W = 0.95$.

The post-hoc analysis for selections per minute showed that:

- C ($Mdn = 61.283$) was more performant than HV ($Mdn = 17.975$, $p < 0.001$), EV ($Mdn = 18.198$, $p = 0.001$), AEV ($Mdn = 19.251$, $p = 0.007$), and VC ($Mdn = 11.358$, $p < 0.001$);
- VC ($Mdn = 11.358$) was less performant than C ($Mdn = 61.283$, $p < 0.001$), HD ($Mdn = 43.297$, $p = 0.012$), ED ($Mdn = 49.632$, $p < 0.001$), and AED ($Mdn = 56.079$, $p < 0.001$);
- AED ($Mdn = 56.079$) was more performant than HV ($Mdn = 17.975$, $p = 0.012$) and EV ($Mdn = 18.198$, $p = 0.02$);
- ED ($Mdn = 49.632$) was more performant than HV ($Mdn = 17.975$, $p = 0.043$).

Similarly, the post-hoc analysis for performance throughput showed that:

- C ($Mdn = 3.959$) was more performant than HV ($Mdn = 1.224$, $p < 0.001$), EV ($Mdn = 1.245$, $p = 0.002$), AEV ($Mdn = 1.318$, $p = 0.005$), and VC ($Mdn = 0.778$, $p < 0.001$);
- VC ($Mdn = 0.778$) was less performant than C ($Mdn = 3.959$, $p < 0.001$), HD ($Mdn = 2.89$, $p = 0.009$), ED ($Mdn = 3.248$, $p = 0.001$), and AED ($Mdn = 3.766$, $p < 0.001$);
- AED ($Mdn = 3.766$) was more performant than HV ($Mdn = 1.224$, $p = 0.007$), EV ($Mdn = 1.245$, $p = 0.015$), and AEV ($Mdn = 1.318$, $p = 0.033$).

B. FITTS' LAW MODEL ADJUSTMENT

To better understand the results of Fitts' model, we first evaluated the main effects of factors of the indexes of difficulty on the target times and their correlation, followed by the results of model adjustment to the data.

1) MAIN EFFECTS

To investigate the influence of the different factors of the Fitts task on the target times, the main effects were calculated. The factors of interest were the angle of the rings, the diameter of the targets, and the sine angles of the direction of the targets. Table 5 shows the results of the tests for each factor by the interaction methods.

The results show that the angle of the rings and the diameter of the targets had a significant effect on the target selection times, which means that the target times were significantly different for all methods (except V and VC) by varying those factors. Furthermore, the target times were positively correlated with the angle of the rings (i.e., longer times to select the targets were verified with larger ring angles) and negatively correlated with the diameters of the targets (i.e., faster selections of the targets occurred with bigger targets). The sine of target direction angles showed small effect sizes in general and weak correlations with target selection times.

Handheld controllers that require stable hands for precision movements (associated with smaller targets) showed a stronger correlation with the diameter of the targets than the remaining methods.

The exception to this was voice-only methods (V and VC) with selection times that appear to vary significantly with the angle of the rings but with much smaller effect sizes and weaker correlations; and without a main effect or correlation with the diameter of the targets. For the sine of the direction angle, these methods showed a significant difference in target times, but weak correlations.

2) LINEAR REGRESSIONS

Simple linear regressions were performed using the model with Equation 1.

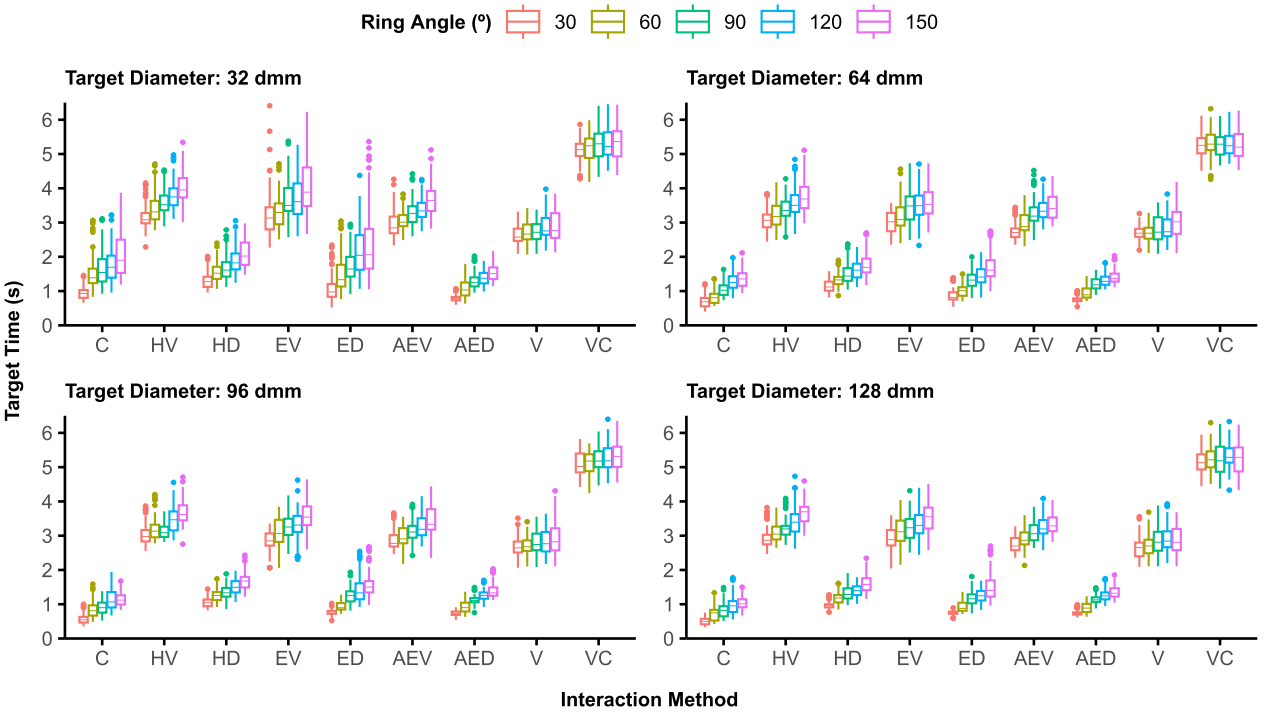


FIGURE 6. Time to confirm a correct target (in seconds) by interaction method and ring angle (°) for each of the target diameters (dmm).

TABLE 5. Results of the Friedman test and Spearman correlation to assess the main effects of ring angle, target diameter, and sine of target direction angle in the target times. (* $p < 0.005$).

			Friedman				Spearman		
Factor	Method	df	χ^2	p		W	ρ	p	
Ring Angle	C	4	35.289	0.000	*	0.980	0.583	0.000	*
	HV	4	34.578	0.000	*	0.960	0.567	0.000	*
	HD	4	36.000	0.000	*	1.000	0.648	0.000	*
	EV	4	30.667	0.000	*	0.852	0.422	0.000	*
	ED	4	36.000	0.000	*	1.000	0.684	0.000	*
	AEV	4	33.867	0.000	*	0.941	0.564	0.000	*
	AED	4	36.000	0.000	*	1.000	0.807	0.000	*
	V	4	17.600	0.001	*	0.489	0.234	0.000	*
	VC	4	11.822	0.019	*	0.328	0.121	0.000	*
Target Diameter	C	3	25.933	0.000	*	0.960	-0.542	0.000	*
	HV	3	19.533	0.000	*	0.723	-0.251	0.000	*
	HD	3	27.000	0.000	*	1.000	-0.388	0.000	*
	EV	3	22.600	0.000	*	0.837	-0.218	0.000	*
	ED	3	24.733	0.000	*	0.916	-0.368	0.000	*
	AEV	3	23.800	0.000	*	0.881	-0.144	0.000	*
	AED	3	21.133	0.000	*	0.783	-0.162	0.000	*
	V	3	4.867	0.182		0.180	0.019	0.469	
	VC	3	2.200	0.532		0.081	-0.013	0.606	
Sine of Target Direction Angle	C	8	2.578	0.958		0.036	-0.014	0.591	
	HV	8	16.089	0.041	*	0.223	-0.061	0.017	*
	HD	8	30.074	0.000	*	0.418	-0.065	0.010	*
	EV	8	12.207	0.142		0.170	-0.034	0.184	
	ED	8	14.933	0.060		0.207	-0.033	0.201	
	AEV	8	16.059	0.042	*	0.223	-0.008	0.772	
	AED	8	17.570	0.025	*	0.244	-0.042	0.103	
	V	8	54.133	0.000	*	0.752	-0.177	0.000	*
	VC	8	35.348	0.000	*	0.491	-0.036	0.162	

Results show (Table 6) that the linear regressions were statistically significant in all the interaction methods, as well as the respective values of slope (b) and intercept (a). Fig. 8 shows a visual representation of the linear regressions.

TABLE 6. Results of the simple linear regression for the Fitts' Law model for each of the interaction methods. (** $p < 0.001$).

Method	*RSE*	R^2*Adj.*	*a* [CI 95%]	*b* [CI 95%]	*F*
C	0.34	0.57	-0.612 [-0.687, -0.536] **	0.408 [0.391, 0.426] **	$F(1, 1540) = 2066.633$ **
HV	0.40	0.29	2.282 [2.191, 2.373] **	0.268 [0.247, 0.289] **	$F(1, 1518) = 607.406$ **
HD	0.27	0.49	0.287 [0.226, 0.348] **	0.281 [0.267, 0.295] **	$F(1, 1578) = 1505.864$ **
EV	0.67	0.18	1.983 [1.83, 2.135] **	0.336 [0.3, 0.371] **	$F(1, 1514) = 339.863$ **
ED	0.44	0.45	-0.429 [-0.53, -0.327] **	0.425 [0.401, 0.448] **	$F(1, 1529) = 1251.439$ **
AEV	0.38	0.22	2.262 [2.174, 2.349] **	0.216 [0.196, 0.237] **	$F(1, 1491) = 427.701$ **
AED	0.23	0.44	0.241 [0.188, 0.293] **	0.213 [0.201, 0.225] **	$F(1, 1519) = 1176.335$ **
V	0.37	0.02	2.547 [2.461, 2.632] **	0.058 [0.038, 0.078] **	$F(1, 1484) = 32.904$ **
VC	0.41	0.01	5.047 [4.953, 5.14] **	0.047 [0.025, 0.069] **	$F(1, 1481) = 17.907$ **

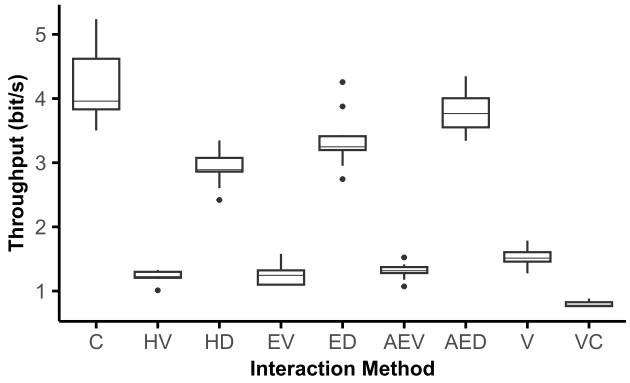


FIGURE 7. Performance throughput (in bits/s) for each of the interaction methods.

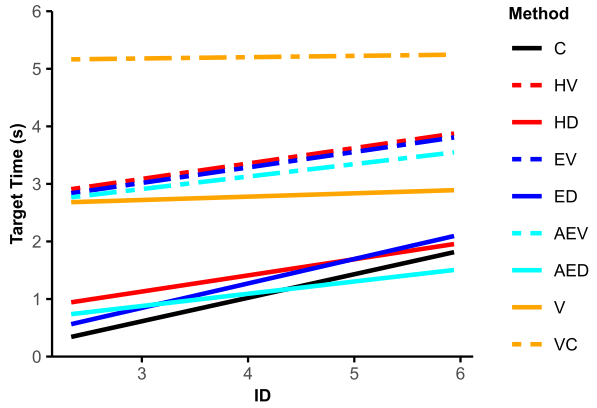


FIGURE 8. Plot of the simple linear regression for the Fitts' Law model for each of the methods.

These results also reflect those of the main effects. Methods with stronger correlations in target diameters and ring angles (reflected in ID) show bigger slopes. Additionally, the methods with lower slope values indicate that they are not as influenced by ID (and respective factors) as the others with higher slopes.

VI. DISCUSSION

In interpreting the results of this study, it is important to consider the context of the targeted audience consisting of advanced VR users and limit the generalization of the findings to that user group.

Focusing on objective effectiveness metrics (Table 2), voice-only methods by having a more direct pointing phase registered, as expected, a lower number of hovers on the correct and incorrect targets than the other methods. We hypothesize that the controllers and eye gaze registered higher hovers for being less stable for pointing, as we also verified that during the experiment users were frustrated with the precision of these methods. Additionally, when using controllers or dwell, users were more engaged and tried to be faster, resulting in a higher number of hovers and higher performance.

Regarding the number of errors caused by confirmation of the wrong target, handheld controllers performed significantly worse than the remaining methods resulting from Heisenberg errors [28] where the movement of pressing the trigger changed the pointing sufficiently for the selection to be performed outside the targets.

In general, we observed that hands-free methods were able to produce similar or better effectiveness than handheld controllers in most metrics and that assisted eye gaze with dwell was preferred by users (RQ1.1). Furthermore, our results also showed that the multimodal methods were similar to the other hands-free methods in terms of effectiveness (RQ1.2).

Despite the lower effectiveness of the handheld controllers, considering the experiment time (Table 3), they were still the most efficient methods, with hands-free methods matching the efficiency of the controllers (RQ2.1). As expected, methods that used the voice resulted in longer experiment times (roughly twice as long) than their non-multimodal counterparts. Using the voice with one command (V) was slightly more efficient than using the voice as a confirmation method. This suggests that the voice detection system had a significant impact on the interactions, making them slower and resulting in a higher frustration level as anecdotally found, especially as users had to keep pointing at the target during the recognition. Faster recognition and the ability to lock the pointing on the target while recognition is ongoing are possible solutions to increase the efficiency and performance of methods that use the voice.

Performance data (Table 4) followed a pattern similar to that of efficiency. We found that handheld controllers had the highest performance, followed by the methods that used dwell, and finally by the methods that used voice (RQ2.2).

Again, using two voice commands (VC) was the slowest and least performant method for this task. Given the additional time required for voice recognition, these results are expected. Despite the hands-free methods with dwell being slightly less performant than the handheld controllers, the difference was not significant. We believe that these methods, especially assisted eye gaze with dwell, can match the selection performance of controllers across the tested target sizes and placements (RQ2.1). We also verified that using an improved detection method for eye gaze is beneficial for the performance of the interaction, with users reporting that pointing was more stable with the assisted gaze.

As found in [7] the selection times are in the 500 ms to 1000 ms range for dwelling in targets that are within the eye's FOV (Fig. 6). Similarly to [7] and [25] we would expect that using a confirmation method for eye gaze that is near instant (e.g., button or movement) would make it the most performant method of all, bringing the selection time to under 500 ms and outperforming the controllers. Our results did not show a clear difference between the efficiency of handheld controllers and near-instant hands-free methods such as the differences found in [33], [34], and [35].

Regarding Fitts' law factors (Table 5), we verified that the angle of the rings and the diameter of the targets had a significant main effect on the target times for all the methods, except for the voice-only methods, whose target times were not affected by the target diameter. As expected, the rings were positively correlated with target times (i.e., a larger angle had more distance between the targets and, consequently, higher target times). On the contrary, the diameter of the target was negatively correlated with the target times (i.e., the smaller the target, the harder and longer it took to select it). The sine of the target direction angle had an effect on the head-gaze methods, with a negative correlation, meaning that the top hemisphere placement had lower times than the lower hemisphere and that the placement of interactable elements should account for this.

Although the simple linear regressions are significant (Table 6), we found that the adjustment of the model to the data was poor in the methods with fast movement times. As Fitts' law models the relationship between selection time and the difficulty of the targets (which depends on distance), when the movement time is very low, the variation in selection time does not significantly increase or decrease with the difficulty. Therefore, care must be taken when using Fitts' law as an objective measure of performance for hands-free methods.

Looking at the b coefficients (slope) for the ID between the methods, we can observe that the performance of the voice-only methods is practically not influenced by the difficulty of the targets. Furthermore, almost all hands-free methods were less dependent on target difficulty than handheld controllers. The results of using an angular model were not reported because we did not verify a better fit using it.

The Fitts' models (Fig. 8) suggest that there are three performance tiers. The first with the controllers and dwell methods, the second with multimodal approaches and one voice command, and the third with two voice commands. The performance of assisted eye gaze is higher than purely using eye gaze and than the other direct methods the more difficult the targets are to select, while controllers are still the most performant method when it comes to less difficult targets. All the multimodal approaches were less performant than only using the one voice command, however not enough to prove it as a better method since this analysis is naive and requires further exploring with user satisfaction and system usability metrics.

A more robust performance metric for this type of task is selections per minute (Table 4), which can be easily derived from target selection times. In traditional 2D interfaces and interfaces where a cursor requires high pointing precision, Fitts' Law can be a good predictor for selection times, whereas the use of hands-free interfaces allows for more direct interaction paradigms for selection (i.e., as long as the targets are within the users' FOV, they can be easily pointed at). As such, given our results, we believe that the evaluation of confirmatory methods is of greater importance for evaluating the performance of hands-free methods. For instance, we verified that using voice commands for confirmation decreased the performance of hands-free methods.

VII. CONCLUSION

This study evaluated the effectiveness and efficiency of the most commonly used hands-free interfaces for selection and system control tasks in immersive VR. The nine methods evaluated include (1) traditional handheld controllers as a baseline, (2) three hands-free methods that can perform the selections directly with dwell, (3) two hands-free methods where only the voice is used, and (4) three hands-free methods with a multimodal approach where the voice is used as a confirmatory method.

A Fitts task was performed and we found that using eye gaze with a 500 ms dwell time proved to be the hands-free method with the highest performance, matching the handheld controllers and being preferred by users. However, we must note that eye gaze is limited by the tracking system. For example, the system used in this study was unable to work when users had corrective glasses. The evaluation also showed that using a multimodal approach to selection, especially using the voice, decreases performance, but increases effectiveness. Moreover, using the voice can be challenging in public spaces (due to privacy and other concerns) Therefore, we believe that conducting studies on alternative confirmation methods for selection is important. Studying the usability of interaction interfaces goes beyond the study of performance, and requires accounting for the system usability, user satisfaction, and user experience. These additional metrics will be explored in a comprehensive follow-up study.

Additionally, the study demonstrates that Fitts' law can be applied to hands-free VR methods, albeit with its limitations, as some hands-free methods have negligible movement times.

We suggest selections per minute as an alternative objective metric for performance, as it more directly reflects the task.

Future work is needed to verify the applicability of techniques used in augmented reality and mixed reality systems (e.g., [49]) and the results of this study in such systems given their recent convergence with typical VR systems. In future studies, the use of more novel interaction methods, such as motion-based techniques [50], electromyography [25] and brain-computer interfaces [24] should also be explored. Furthermore, the results of this study are limited to the context of the target sample (advanced users) and sample size and, as such, cannot be generalized to larger audiences. It shall be used as a starting point for future studies to explore the impact of these methods with other audiences and on different VR scenarios and tasks.

In conclusion, this study provides valuable information for the design of better and more reliable methodologies to evaluate interactions in immersive VR experiences. It also contributes to the literature by evaluating hands-free interaction methods in immersive VR.

REFERENCES

[1] M. Melo, G. Gonçalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa, "Do multisensory stimuli benefit the virtual reality experience? A systematic review," *IEEE Trans. Vis. Comput. Graphics*, vol. 28, no. 2, pp. 1428–1442, Feb. 2022.

[2] W. R. Sherman and A. B. Craig, *Understanding Virtual Reality: Interface, Application, and Design*. San Mateo, CA, USA: Morgan Kaufmann, 2002.

[3] P. Monteiro, H. Coelho, G. Gonçalves, M. Melo, and M. Bessa, "Comparison of radial and panel menus in virtual reality," *IEEE Access*, vol. 7, pp. 116370–116379, 2019.

[4] G. Gonçalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa, "Systematic review of comparative studies of the impact of realism in immersive virtual experiences," *ACM Comput. Surv.*, vol. 55, no. 6, pp. 1–36, Dec. 2022.

[5] G. Frey, A. Jurkschat, S. Korkut, J. Lutz, and R. Dornberger, "Intuitive hand gestures for the interaction with information visualizations in virtual reality," in *Augmented Reality and Virtual Reality*. Berlin, Germany: Springer, 2019, pp. 261–273.

[6] P. Monteiro, G. Gonçalves, H. Coelho, M. Melo, and M. Bessa, "Hands-free interaction in immersive virtual reality: A systematic review," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 5, pp. 2702–2713, May 2021.

[7] X. Yi, Y. Lu, Z. Cai, Z. Wu, Y. Wang, and Y. Shi, "GazeDock: Gaze-only menu selection in virtual reality using auto-triggering peripheral menu," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2022, pp. 832–842.

[8] J. Kang and J. Lim, "Storytelling-based hand gesture interaction in a virtual reality environment," in *Advances in Human Factors in Wearable Technologies and Game Design*. Berlin, Germany: Springer, Jun. 2017, pp. 169–176.

[9] *Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems*, Standard ISO 9241-210:2019, Geneva, CH, USA, Jul. 2019.

[10] R. W. Soukoreff and I. S. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI," *Int. J. Hum.-Comput. Stud.*, vol. 61, no. 6, pp. 751–789, Dec. 2004.

[11] Y. Y. Qian and R. J. Teather, "The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality," in *Proc. 5th Symp. Spatial User Interact.*, New York, NY, USA, Oct. 2017, pp. 91–98, doi: 10.1145/3131277.3132182.

[12] M. J. Schuemie, P. van der Straaten, M. Krijn, and C. A. van der Mast, "Research on presence in virtual reality: A survey," *Cyberpsychology Behav.*, vol. 4, no. 2, pp. 183–201, Apr. 2001.

[13] P. Fuchs, *Virtual Reality Concepts and Technologies*. Boca Raton, FL, USA: CRC Press, 2011.

[14] J. Jerald, *The VR Book: Human-Centered Design for Virtual Reality*. San Rafael, CA, USA: Association for Computing Machinery and Morgan & Claypool, 2015.

[15] M. Höll, M. Oberweger, C. Arth, and V. Lepetit, "Efficient physics-based implementation for realistic hand-object interaction in virtual reality," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2018, pp. 175–182.

[16] J. J. LaViola, E. Kruijff, R. P. McMahan, D. A. Bowman, and I. Poupyrev, *3D User Interfaces: Theory and Practice* (Addison-Wesley Usability and HCI Series). Reading, MA, USA: Addison-Wesley, 2017.

[17] P. Majaranta and A. Bulling, "Eye tracking and eye-based human-computer interaction," in *Advances in Physiological Computing* (Human-Computer Interaction Series). London, U.K.: Springer, 2014, pp. 39–65.

[18] D. Mardanbegi and T. Pfeiffer, "EyeMRTk: A toolkit for developing eye gaze interactive applications in virtual and augmented reality," in *Proc. 11th ACM Symp. Eye Tracking Res. Appl.*, Jun. 2019, pp. 1–5.

[19] R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: What you look at is what you get," *ACM Trans. Inf. Syst.*, vol. 9, pp. 152–169, Apr. 1991.

[20] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger, "Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality," in *Proc. ACM Symp. Eye Tracking Res. Appl.*, May 2021, pp. 1–7.

[21] P. Majaranta, U.-K. Ahola, and O. Spakov, "Fast gaze typing with an adjustable dwell time," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2009, pp. 357–360.

[22] L. Cao, H. Zhang, C. Peng, and J. T. Hansberger, "Real-time multimodal interaction in virtual reality—A case study with a large virtual interface," *Multimedia Tools Appl.*, vol. 82, no. 16, pp. 25427–25448, Feb. 2023.

[23] D. Hepperle, Y. Weiß, A. Siess, and M. Wölfel, "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.*, vol. 82, pp. 321–331, Aug. 2019.

[24] W. McClinton, D. Caprio, D. Laesker, B. Pinto, S. Garcia, and M. Andujar, "P300-based 3D brain painting in virtual reality," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–6.

[25] Y. S. Pai, T. Dingler, and K. Kunze, "Assessing hands-free interactions for VR using eye gaze and electromyography," *Virtual Reality*, vol. 23, no. 2, pp. 119–131, Nov. 2018.

[26] K. Wang, Q. Liu, Y. Zhao, C. Y. Zheng, S. Vhasure, Q. Liu, P. Thakur, M. Sun, and Z. Mao, "Intelligent wearable virtual reality (VR) gaming controller for people with motor disabilities," in *Proc. IEEE Int. Conf. Artif. Intell. Virtual Reality (AIVR)*, Dec. 2018, pp. 161–164.

[27] M. Gelsomini, G. Leonardi, and F. Garzotto, "Embodied learning in immersive smart spaces," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–14.

[28] D. Wolf, J. Gugenheimer, M. Combosch, and E. Rukzio, "Understanding the Heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–10.

[29] J. Blattgerste, P. Renner, and T. Pfeiffer, "Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views," in *Proc. Workshop Commun. Gaze Interact.*, Jun. 2018, pp. 1–9, doi: 10.1145/3206343.3206349.

[30] K. A. M. Heydn, M. P. Dietrich, M. Barkowsky, G. Winterfeldt, S. von Mammen, and A. Nuchter, "The golden bullet: A comparative study for target acquisition, pointing and shooting," in *Proc. 11th Int. Conf. Virtual Worlds Games Serious Appl. (VS-Games)*, Sep. 2019, pp. 1–8.

[31] M. Speicher, S. Cucerca, and A. Kruger, "VRShop: A mobile interactive virtual reality shopping environment combining the benefits of on- and offline shopping," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–31, Sep. 2017, doi: 10.1145/3130967.

[32] Y. Yan, Y. Shi, C. Yu, and Y. Shi, "HeadCross: Exploring head-based crossing selection on head-mounted displays," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–22, Mar. 2020, doi: 10.1145/3380983.

[33] I. Giannopoulos, A. Komninos, and J. Garofalakis, "Natural interaction with large map interfaces in VR," in *Proc. 21st Pan-Hellenic Conf. Inform.*, Sep. 2017, pp. 1–6, doi: 10.1145/3139367.3139424.

[34] F. L. Luro and V. Sundstedt, "A comparative study of eye tracking and hand controller for aiming tasks in virtual reality," in *Proc. 11th ACM Symp. Eye Tracking Res. Appl.*, New York, NY, USA, Jun. 2019, pp. 1–9, doi: 10.1145/3317956.3318153.

[35] Y. Yan, C. Yu, X. Yi, and Y. Shi, "HeadGesture: Hands-free input approach leveraging head movements for HMD devices," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 1–23, Dec. 2018, doi: 10.1145/3287076.
[36] P. Ganapathi and K. Sorathia, "Investigating controller less input methods for smartphone based virtual reality platforms," in *Proc. 20th Int. Conf. Human-Comput. Interact. Mobile Devices Services Adjunct*, Sep. 2018, pp. 166–173, doi: 10.1145/3236112.3236136.
[37] X. Meng, W. Xu, and H. Liang, "An exploration of hands-free text selection for virtual reality head-mounted displays," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2022, pp. 74–81.
[38] D. Zielasko, N. Neha, B. Weyers, and T. W. Kuhlen, "A reliable non-verbal vocal input metaphor for clicking," in *Proc. IEEE Symp. 3D User Interfaces*, Mar. 2017, pp. 40–49.
[39] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Experim. Psychol.*, vol. 47, no. 6, pp. 381–391, 1954.
[40] P. M. Fitts and J. R. Peterson, "Information capacity of discrete motor responses," *J. Exp. Psychol.*, vol. 67, no. 2, pp. 103–112, 1964.
[41] E. Triantafyllidis and Z. Li, "The challenges in modeling human performance in 3D space with Fitts' law," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–9.
[42] X. Zhang and I. S. MacKenzie, "Evaluating eye tracking with ISO 9241—Part 9," in *Human-Computer Interaction, HCI Intelligent Multimodal Interaction Environments*. Berlin, Germany: Springer, 2007, pp. 779–788.
[43] L. D. Clark, A. B. Bhagat, and S. L. Riggs, "Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology," *Int. J. Hum.-Comput. Stud.*, vol. 139, Jul. 2020, Art. no. 102413.
[44] Y. Cha and R. Myung, "Extended Fitts' law for 3D pointing tasks using 3D target arrangements," *Int. J. Ind. Ergonom.*, vol. 43, no. 4, pp. 350–355, Jul. 2013.
[45] C. McKenzie and A. Glazier. (May 2017). *Designing Screen Interfaces for VR (Google I/O '17)*. Google Developers. [Online]. Available: https://www.youtube.com/watch?v=ES9jArHRFHQ
[46] H. Strasburger, I. Rentschler, and M. Juttner, "Peripheral vision and pattern recognition: A review," *J. Vis.*, vol. 11, no. 5, p. 13, Dec. 2011.
[47] S. Purwar. (2019). *Designing User Experience for Virtual Reality (VR) Applications*. [Online]. Available: https://uxplanet.org/designing-user-experience-for-virtual-reality-vr-applications-fc8e4faadd96
[48] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Evanston, IL, USA: Routledge, 1988.
[49] Y. Wei, R. Shi, D. Yu, Y. Wang, Y. Li, L. Yu, and H.-N. Liang, "Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2023, pp. 1–14.
[50] W. Xu, H.-N. Liang, Y. Zhao, D. Yu, and D. Monteiro, "DMove: Directional motion-based interaction for augmented reality head-mounted displays," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–14.

PEDRO MONTEIRO received the M.Sc. degree in computer science from the University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, where he is currently pursuing the Ph.D. degree in informatics. Since 2016, he has been a Research Fellow with INESC TEC, Porto, Portugal. His research interests include virtual reality, virtual reality interaction, and virtual reality user interfaces. He was awarded the JLE Award (Grupo Portugus de Computao Grfica, Porto), in 2021.

GUILHERME GONÇALVES received the M.Sc. degree in multimedia from the University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, where he is currently pursuing the Ph.D. degree in informatics. Since 2018, he has been a Research Fellow with INESC TEC, Porto, Portugal. His research interest includes multisensory virtual reality.

BRUNO PEIXOTO received the B.S. degree in communication and multimedia and the M.Sc. in multimedia from the University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, where he is currently pursuing the Ph.D. degree in informatics. Since 2019, he has been a Research Fellow with INESC TEC, Porto, Portugal. His research interest includes virtual reality for education.

MIGUEL MELO is currently an Assistant Professor with the Department of Engineering, University of Trás-os-Montes e Alto Douro, Portugal, and a Senior Researcher with INESC TEC. He is also the Research and Development Manager of the MASSIVE Virtual Reality Laboratory and a member of the Eurographics Executive Committee. His research interests include computer graphics, HDR, and multisensory virtual reality.

MAXIMINO BESSA is currently an Assistant Professor (Habilitation) with the Department of Engineering, University of Trás-os-Montes e Alto Douro, Portugal, has been a Senior Researcher with INESC TEC, since 2009, and the Director of the Multisensory Virtual Reality Laboratory MASSIVE. He has been a member of the Eurographics Association, since 2003, and the Vice-President of the Portuguese Computer Graphics Chapter (2016–2018).