

Hands-free interaction in immersive virtual reality: A systematic review

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

Abstract—Hands are the most important tool to interact with virtual environments, and they should be available to perform the most critical tasks. For example, a surgeon in VR should keep his/her hands on the instruments and be able to do secondary tasks without performing a disruptive event to the operative task. In this common scenario, one can observe that hands are not available for interaction. The goal of this systematic review is to survey the literature and identify which hands-free interfaces are used, the performed interaction tasks, what metrics are used for interface evaluation, and the results of such evaluations. From 80 studies that met the eligibility criteria, the voice is the most studied interface, followed by the eye and head gaze. Some novel interfaces were brain interfaces and face expressions. System control and selection represent most of the interaction tasks studied and most studies evaluate interfaces for usability. Despite the best interface depending on the task and study, the voice was found to be versatile and showed good results amongst the studies. More research is recommended to improve the practical use of the interfaces and to evaluate the interfaces more formally.

Index Terms—Systematic Review, Virtual Reality, Human-Computer Interaction, Hands-free.

1 INTRODUCTION

Nowadays, with the modernization of industries using new technologies, virtual reality (VR) became a key tool for the future of training and certification of teams in an expeditious way [6]. It can promote success and safety in training tasks [7] in several areas such as medicine and education. Thus, VR is expected to boost higher levels of productivity through technological modernization and innovation, making it a key piece for the future of education and the training of skilled professionals.

Most of the interactions with VR applications are currently performed by using the remote controllers which are sold out-of-the-box with the most sold VR consumer systems [101]. These are made possible because of tracking systems [18] that allow the use of tangible interfaces (like remote controllers [79] of the VR setup apparatus) in 3D spaces. This means that the common interactions with a VR system become hindered and limited in scenarios where the users' hands are used to directly interact with the virtual environment (VE). For instance, in a high-fidelity training scenario, when trainees are learning to use a tool that they hold in their hands, the common system interaction becomes hindered when there is a need to invoke an agent for help. Also, with collaborative and cooperative VR experiences, when users interact naturally with each other using their hands, an alternative hands-free mechanism is needed to allow for system control. Another example is in tourism applications where the users might be driving a buggy to explore a desert or fly gliding and receive feedback through tangible interfaces, requiring hands-free techniques to navigate through the application menus.

Previous literature reviews and systematic surveys have discussed interaction in immersive VR. For instance, [20] reviewed the literature regarding the available 3D UI menus for object and option selection, providing a taxonomy for this type of menus and classifying them regarding their intention, appearance and structure, placement, invocation and availability, interaction, and combinability. Regarding navigation, [118] presents a comprehensive review of the methods that allow users to navigate through VEs grouping them by the type of action

required to perform the navigation (e.g., walking, steering, selecting). At [74], an extensive review of the transformation of 3D virtual objects is provided, specifying the techniques used from desktop to VR to transform the objects regarding changing their location, rotation, and scale. Similarly, [40] presents a survey of 3D interaction techniques for object manipulation, viewpoint manipulation and, system control. Regarding the system control tasks, the author denotes that the techniques for this are less explored, as usually, traditional 2D techniques (e.g., 2D menus) are incorporated in 3D VE, possibly making tasks more difficult to perform. This happens because using an HMD prevents users from using traditional input devices (e.g., mouse and keyboard), a problem that can be solved by using virtual lasers to point at the 2D interfaces that are floating in the 3D space.

Although the current literature discusses interaction in VR scenarios, the focus of the studies is on generic interaction interfaces for those environments with a big focus on object transformation and navigation. As such, literature lacks a comprehensive literature review that surveys interaction within the immersive VE without using the hands. In fact, during a preliminary search, most studies considered two groups of interaction tasks - system control and navigation. Since in most cases navigation is already a hands-free task by nature, this systematic review focuses on all other interaction tasks except navigation (Table 1). Usually, hand gestures are considered a hands-free interaction method. However, since in our scenario they can be a disruptive event to the hand interaction (e.g., a surgeon will not lose the focus or put down its instruments to perform a hand gesture), hand gestures are not considered hands-free interaction in this systematic review.

The four main goals of this systematic review are: (1) identify which hands-free interfaces are used to interact with the VE. Secondly, we aim to (2) identify which interactions tasks in immersive VR are usually performed using hands-free methods; Lastly, our goal is also to (3) identify which metrics are currently used to evaluate the hands-free interactions and interfaces, and (4) understand how they perform in the context of those metrics. To answer these questions, we have performed an extensive literature review.

With such systematic review, this work will provide an overview of the current state-of-the-art, namely the different hands-free interfaces available across the different interaction tasks and how they perform. Moreover, the results can help aid in the design and creation of applications that use hands-free interfaces, by showing the interaction possibilities and implications of the interfaces. The analysis will also provide better insights into existing research opportunities on this topic.

- *Pedro Monteiro, Guilherme Gonçalves, Hugo Coelho, and Miguel Melo where with INESC TEC, 4200-465 Porto, Portugal. E-mail: monteiro.p@outlook.pt, guilhermeg@utad.pt, hugo.r.mendes@inesctec.pt, and mcmelo@inesctec.pt.*
- *Maximino Bessa was with INESC TEC, 4200-465 Porto, Portugal and with Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal. E-mail: maxbessa@utad.pt*

1.1 Background

VR encompasses three components [29]: a user, a VE, and an interface that is used to interact with the VE and to immerse the user. VR can immerse users in a VE [93] through their senses with the various interfaces between the user and the 3D simulations [29]. Following the definition provided in [89], a VR system is immersive when users are completely or almost completely integrated into a VE through devices such as head-mounted displays (HMD) and CAVE Automatic Virtual Environments (CAVE). As such, interaction methods should evolve to support high levels of presence [97] (for a more comprehensive overview of research on presence in VR please refer to [91]) and authentic virtual experiences [61].

Human-Computer Interaction (HCI) is already widely studied in conventional interaction contexts [51]. However, with the addition of another spatial dimension (2D to 3D), and consequently more degrees of freedom, it is required that the most common tasks and methods of interaction are rethought and adapted to VR and its technologies [9]. Examples of such tasks include finding the best way to perform basic interaction tasks (selection, insertion, manipulation) or how to perform gesture recognition in 3D environments when the user has an HMD.

In [52], it is argued that VR interaction should be human-centered, and a set of guidelines are provided for the implementation of such interactions, especially how they should be intuitive, follow the principles of interaction, and the importance of the human hands for interaction. The focus on body-centered interaction is also defended in [96], with authors mentioning the body as the main point of interaction with the VE.

To design good interactions, a set of principles around the discoverability of the interactions should be followed [80], i.e., how to design interactions that correctly communicate to users their intent. The five principles are affordances (how something can be interacted with), signifiers (indicators of the interaction), constraints (restrictions on the interaction), feedback (results of the interaction), and mappings (relationships between the interactions). The interactions with a 3D user interface (UI) can also be classified by their intent in several interaction tasks [52,65,93]. Table 1 provides a summary of these interaction tasks. Different techniques and devices can be used to perform each of these types of tasks (e.g., selection can be performed by touching an object or by pointing at it). Additionally, multiple devices or techniques can be combined resulting in multimodal interaction, for instance, selecting an object can be done by pointing at the object, and the selection can be confirmed by voice command.

An increase in functionality is associated with an increased cognitive load [14], which can occur when the buttons of tangible interfaces (e.g., VR controllers) are used to perform multiple tasks. Thus, this approach or the use of commonly used tracked controllers is not ideal

for interaction with the VE in such conditions and hands-free techniques for system manipulation and control can be used. Therefore, it is crucial to study the most appropriate interaction metaphors and their applicability to VR.

To overcome this limitation, alternative natural user interfaces (NUI) can be used to allow interactions while maintaining the hands of the users free. The concept of NUI comes from the way people interact with each other and with the world [5], either by gestures, expressions, movements, or communication, physically manipulating the objects around them. NUI in VR are usually associated with naturally grasping and grabbing objects with the hands instead of using handheld devices, and thus more related to the selection and manipulation tasks. However, VR input can be done with more than the users' hands which can be particularly useful for system control and symbolic input tasks.

2 METHODS

The methodology followed in this survey was based on PRISMA [77]. This methodology provides guidelines for the development of systematic reviews and meta-analyses to ensure a transparent and complete reporting of the surveyed topics.

2.1 Eligibility Criteria

Studies that explore the use of hands-free interaction interfaces and methods in an immersive VR context were considered for the systematic review. The inclusion and exclusion criteria for this review are presented in Table 2.

There is a possible limitation with the search query not returning all the relevant studies. To overcome and mitigate this limitation a second inclusion criterion was added, allowing for relevant studies to be added through a manual search of the most relevant journal, conferences, and most cited works.

2.2 Search Strategy

The literature was identified by conducting an extensive search on three electronic science databases: Web of Science (Clarivate Analytics), Scopus (Elsevier), and ACM Digital Library (ACM). The first search was performed on 14 May 2020. No other databases were used since according to our searches, the three used databases already index the studies from other publishers such as Springer, PubMed, and IEEE.

The search query consisted of four base components: "VR" and its equivalent terms, "immersive" and its equivalent terms, "HCI" and its equivalent terms, and lastly a set of terms that denote "hands-free interaction" as well as some interfaces that we know *a priori* that can enable hands-free interaction. As such, the query used on the Scopus database was: *TITLE-ABS-KEY ("virtual reality" OR "vr" OR "virtual environment") AND TITLE-ABS-KEY ("immersive" OR "cave"*

Table 1. Types of interaction tasks for 3D user interfaces. Adapted from [52,65,93].

Task	Description
Selection	Specification of a target object to which a command will be applied. Includes picking the object and confirming.
Manipulation	Change the attributes of a target object by directly manipulating its state (e.g., switch state on/off, pressing a 3D button)
Transformation	Apply a direct geometrical transformation to a virtual object (i.e., position, rotation, and scale)
Creation	Change the number of active scene objects by creating or destroying objects
Modification	Change the intrinsic properties of an object (e.g. appearance)
Navigation	Change the user viewpoint (e.g., walking, teleporting)
System Control	Issuing commands to the application
Symbolic Input	Character input or editing

Table 2. Inclusion and exclusion criteria for the studies.

Criteria	Description
IC1	The study title, abstract or keywords matches the search query and is indexed on one of the queried databases (refer to Sect. 2.2 for a detailed description).
IC2	The study is relevant for this systematic review and is not a result of the search query (i.e., manually added).
EC1	The document is a duplicate
EC2	The document is not a research paper
EC3	The study's full-text is not available
EC4	The study is not written in English
EC5	The study's subject does not consider HCI
EC6	The study's subject does not consider a valid interaction task (as defined in Table 1, excluding navigation)
EC7	The study does not consider a hands-free interaction method
EC8	The study does not consider immersive VR
EC9	The study does not consider an implementation of the interaction method

OR "hmd" OR "head mount*" OR "headset") AND TITLE-ABS-KEY ("human computer interaction" OR "HCI" OR "human machine interaction" OR "HMI" OR "man machine interaction" OR "MMI" OR "computer human interaction" OR "CHI") AND TITLE-ABS-KEY ("multimodal" OR "hand free" OR "novel" OR "natural user interface" OR "NUI" OR "natural interface" OR "natural interaction" OR "voice" OR "speech" OR "spoke" OR "VUI" OR "brain computer" OR "BCI" OR "brain machine" OR "BMI" OR "neural control" OR "NCI" OR "mind machine" OR "MMI" OR "direct neural" OR "DNI" OR "eye" OR "gaze" OR "optical" OR "*myograph" OR "*myography" OR "myographic" OR "neuromuscular" OR "gesture" OR "pose"). The queries for the other databases were similar while taking into consideration the search characteristics of those databases.

2.3 Study Selection

Considering the studies gathered by the inclusion criteria and search strategy described above, eligibility assessment and data abstraction were performed independently in a conventional unblinded standardized manner by a total of three reviewers. Each study was screened by two reviewers, with studies being randomly distributed between the reviewers. The studies were only rejected when there was a consensus between both reviewers for the rejection, otherwise, the study was eligible for full-text analysis. The full-text analysis was performed by one reviewer.

2.4 Data Collection Process

The selected studies for full-text analysis were reviewed, and the data collection process was conducted using piloted forms. To support the goals of this literature review, three types of variables were collected from the studies: interaction properties, the study scope, and evaluation data. For the interaction properties variable, data regarding the type of interface (e.g., head tracking or eye-tracking), the interaction task (e.g., selection, system control, symbolic input), and the devices used for interaction were collected. The goal and main findings of the studies related to the hands-free interaction we collected for the study scope variable. Lastly, related to evaluation data, if an evaluation is provided, we collect the metrics used (e.g., which questionnaires) and, when a comparison of interfaces is provided, the significant results of the interaction interfaces comparison under those metrics. The collected data allows for a qualitative analysis of the selected studies.

2.5 Quantitative Analysis and Quality Assessment

Based on the data gathered from the data collection process, a quantitative analysis was performed, characterizing the distribution of studies by the interface, interaction task, and evaluation metrics. To provide a better overview of the most used terms in the titles of the analyzed studies, a word cloud was generated by using the Display (<https://displayr.com>) and Word Art (<https://wordart.com>) online tools. Moreover, an analysis of scientific production over the years is also provided to better show the evolution of studies over the years.

The quality of the studies was also assessed in the context of this review. A scoring system was adopted based on the approaches for the quality assessment used by [73]. The scores were attributed to certain factors that reflect the expressiveness of the studies in the context of the systematic review and ranged from 1 to 3 points for each factor and were summed into an overall quality score that ranged from 3 to 9 points, with 3 points being the lowest quality and 9 the highest quality. The following factors and respective scores were considered:

QA1 – Focus How relevant is the focus of the study for addressing the questions of this review? Is the hands-free interface the focus of the study? (1 - Somewhat relevant when it has data regarding one of the goals; 2 - Relevant when it has data for two or three of the goals; 3 - Very relevant when it has data for all four goals)

QA2 – Comparison Within the scope of this systematic review, is the hands-free interface compared to another interface? (1 - No; 2 - Compared to a non-hands-free interface; 3 - Compared to a hands-free interface)

QA3 – Methodology To what extent can the study methodology and findings be trusted in answering the study question(s)? (1 - Somewhat trusted when the sample is low, or no data reported; 2 - Trusted when there is a sufficient sample and lacking on validated metrics or robust statistical analysis; 3 - Completely trusted when it has a significant sample, results based on adequate metrics and a robust statistical analysis)

3 RESULTS AND DISCUSSION

3.1 Study Selection

According to the previously defined search strategy, a total of 1337 records were identified and collected from the queried databases, of which 230 were identified as being duplicate studies and were consequently removed. Additionally, 12 studies were manually added to the record database due to their relevance for this survey, resulting in a record database of 1119 unique records. The title and abstracts of these unique records were screened, and 785 records were removed due to matching the exclusion criteria, resulting in 334 records eligible for full-text analysis. From the full-text analysis, 254 records were removed since they also matched the exclusion criteria, which in return resulted in 80 studies accepted for data extraction¹. Please refer to Fig. 1 for an overview of the study selection.

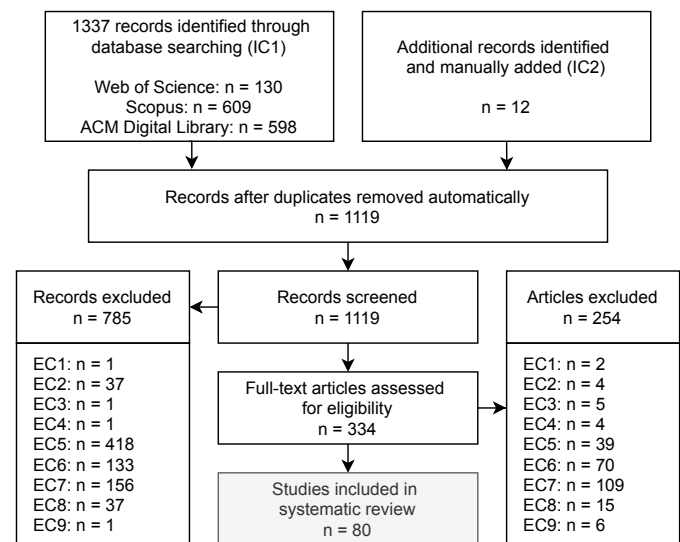


Fig. 1. Flow diagram of study selection.

3.2 Quantitative Analysis

3.2.1 Word Cloud

A word cloud from the most used terms in the study titles was generated (Fig. 2). The terms list was carefully reviewed to remove common language words and merge synonym terms (e.g., “head-mounted” was merged with “HMD” and “brain” with “BCI”). The terms *Virtual Reality* (54), *Interaction* (25), and *Immersive* (18) overload the word cloud as they reflect the main scope of the studies. If we exclude those terms, the 10 most frequent ones and their respective frequencies are: *eye* (15), *HMD* (12), *interface* (12), *voice* (10), *natural* (9), *evaluation* (9), *BCI* (7), *multimodal* (7), *tracking* (7), and *user* (7). Considering the surveyed interaction interfaces (Table 4), the most frequent terms represent the most researched interfaces.

Interestingly, the frequency of the term “voice” does not manifest the abundance of studies that used that interface (12 of 44 studies). Note that these terms reflect the titles of the studies and, although they should represent their content, such might not be always the case. This might explain the difference in occurrences between the top three terms

¹ Full data extraction table is available as a supplemental file.

Table 4. Taxonomy of interaction interfaces from the surveyed studies. Interfaces are subdivided by the source of interaction, modality, and hardware.

Interaction Source	Modality	Hardware	N	Studies
Voice	Commands	Microphone	20	[3, 10–13, 17, 21, 27, 30, 53, 55, 76, 78, 87, 88, 90, 102, 103, 110, 117]
	Speech	Microphone	14	[2, 3, 13, 22, 24, 26, 45, 56, 57, 63, 70, 81, 92, 116]
	Input	Microphone	6	[2, 12, 66, 84, 98, 106]
	Non-verbal "Blow"	Microphone	4	[25, 99, 119, 120]
Eye	Gaze	N/IR-ET	10	[4, 8, 46, 47, 53, 67, 75, 83, 85, 95]
		Generic ET	4	[38, 68, 69, 82]
		RGB Sensor	4	[1, 36, 39, 100]
		EEG	1	[109]
		EOG	1	[62]
	Blink	EOG	3	[62, 94, 113]
		RGB Sensor	2	[1, 39]
	Relative Movement	EEG	1	[108]
		EOG	1	[94]
Head	Gaze	HMD	14	[4, 8, 15, 42, 46, 60, 69, 75, 78, 83, 85, 86, 114, 115]
		M-HMD	5	[31, 32, 35, 98, 103]
		Gyro	1	[56]
	Relative Movement	M-HMD	2	[10, 59]
		HMD	1	[37]
Brain	SSVEP	EEG	3	[37, 68, 104]
	P300	EEG	2	[23, 71]
	Concentration	EEG	1	[60]
	N/A	EEG	1	[72]
Face	Expression	EEG	2	[108, 109]
	Mouth State	RGB-D Sensor	1	[16]
Foot	Movement	Mouse	2	[4, 75]
Body	Position	RGB Sensor	1	[33]
		RGB-D Sensor	1	[48]
Arm	Force	EMG	1	[83]

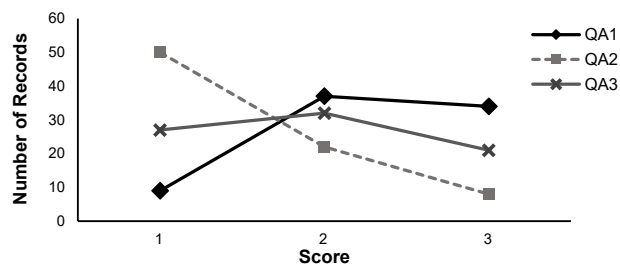


Fig. 6. Quality assessment scores by quality factor.

the system can recognize sentences and detecting a command with parameters while giving the user real-time feedback of the sentence being recognized. Alternatively, the voice can also be used for text input, as in [84, 98], allowing the users to input text by speaking the words. Lastly, there are a few studies that make use of non-verbal actions to issue commands to the VR systems, detecting when the user “blows” to confirm a selection (e.g., [120]).

The high number of studies focusing on the voice might be explained by nearly every HMD having an integrated microphone or adding a microphone to a VR setup is easy with a little negative impact on the experience. Moreover, speech recognition services are widely available, such as the ones from Google and Microsoft, that provide seamless integration with the game engines and allow developers to easily enable speech interfaces. These services mitigate the problem of microphones

tending to pick more than just the voice of the user (e.g., background noise, other people’s voice) which usually results in more errors during the recognition process. Moreover, the use of activation keywords is also common, allowing systems to recognize when the user is issuing a command [3].

Tracking the eyes can be done with optic sensors, such as infrared (IR) or near IR (NIR) eye trackers (ET) and RGB cameras, or by measuring the electrical activity of the eye (electrooculography (EOG)) or of the brain (electroencephalography (EEG)). By using these methods, it is possible to determine the user’s eye gaze (i.e., where the user is looking at in the VE), enabling them to point or aim at virtual elements (e.g., buttons) without moving the head [8]. However, it is also possible to determine relative eye movements (e.g., the eyes turned to the right) and issue a “to right” command [108], and detect when users blink [62, 113] or have their eyes closed [59] to confirm selections or control the VR systems. Additionally, using the eyes for interaction suffers from what is described as the “Midas Touch” problem [50]. Since people naturally use their eyes to explore the environment that surrounds them, they expect to look at things without that look meaning anything, meaning that if the eyes are also used to interact, there is an overload of the visual channel. Practically, this can be solved by coupling these interfaces with other interaction techniques (e.g., select objects by pointing with eye gaze, but require a voice confirmation).

Head tracking is also highly used, being available in practically all the immersive VR systems that use an HMD or mobile smartphone-based HMD (M-HMD), since these need to track the users’ head to allow 360° viewing of the VE. Similarly to eye interaction, one can

obtain the head gaze (i.e., the direction of the head forward vector), and by overlaying a reticle or point drawn in the middle of the image, users can point at objects with their heads, enabling symbolic input by pointing the head to a keyboard (e.g., [42]), and most often, to perform selection tasks (e.g., [114, 115]). Less commonly, it is also possible to detect head gestures by using its relative movements, which can be used to, for instance, instruct the system to reset the current action [37]. Interaction with head gaze is useful for minimal VR setups since it does not require additional hardware. This makes it an excellent tool for interaction in a multitude of commercial applications and games. For this reason, it is unexpected that the number of studies that use the head for interaction is lower than the use of the voice or the eye, which are relatively more novel interfaces. Possible reasoning for this is, despite being an interaction source that can be used straightforwardly, the HMD also requires head movement to view the VE. This can conflict with additional head movements designed to interact with the world, leading to involuntary interactions (“Midas Touch”). It is therefore useful to use an alternative confirmation method (e.g., button [46]) for selection with the head. Moreover, using head movements for the sole purpose of interacting can be fatiguing for users.

The use of EEG devices has allowed advancements for interaction in the context of accessibility, allowing the use of the brain as a source of interaction through Brain-Computer Interfaces (BCI). A few studies use this method to allow, mainly, system control tasks and symbolic input. With steady-state visually evoked potentials (SSVEP), it is possible to read a signal response on the brain by providing a visual stimulation at a certain frequency (i.e., elements flashing at a specific rate). This can be used, to determine if the user is focusing its visual attention on a particular key in a virtual keyboard [37, 68], allowing symbolic input. Besides SSVEP, Event-related potentials (ERP) also account for other types of stimulation, such as the P300 wave which denotes a reaction to a stimulus. In the surveyed literature, the P300 was used as a reaction to a visual stimulus of elements flashing on a matrix, allowing users to control the active tools [71] or select objects [23]. A particular study also used the brain by measuring the users’ concentration to perform actions on the VR system [60]. A major challenge of using BCI is the training of such systems which is required for the correct capture and processing of emotions and motor imagery by EEG devices [107]. Emotion recognition is a complex field since emotions are subjective and are generated in different parts of the brain. There is also a limitation on the amount of motor imagery that systems can recognize, and which is directly influenced by the amount of time spent on calibration.

By using an EEG device, a couple of studies assessed the users’ facial expressions and associated them with specific actions on the VR application, for instance, make the character jump when raising the eyebrows [108]. Alternatively, by using an RGB depth camera sensor (RGB-D), it was possible to detect the mouth state (mouth stretch and expression) and control a character based on those states [16]. In these cases, it is required to train the users, since performing actions with the face is not a natural form of interaction. Using RGB and RGB-D camera sensors, it is also possible to determine the users’ body location in space and use that information to select menu items or objects that are closer to users [48] or in front of the users [33]. Two studies used the users’ feet to control a platform that acted like a mouse depending on the direction and angle of the feet on the platform, allowing the control of the direction of a virtual pointer [4, 75]. Lastly, a single study was found where the users were able to perform the selection by eye gazing to an element and confirming the selection contracting the muscles of their arms [83], which was made possible using an electromyograph (EMG).

For these novel interfaces, the challenge is to seamlessly integrate them into the VR setups to support immersive experiences. In the case of BCI and the face, they directly compete for the same space on the users’ heads, while the former compete for the space used by HMD straps to secure themselves to the heads, the latter competes for the space occupied by the HMD, occluding a part of the face. Using the body position for interactions other than navigation is also a challenge since this technique is mainly used for that purpose.

3.5 Interaction Tasks

To provide an overview of which tasks are performed in a hands-free manner, Table 5 systematizes and summarizes the studies based on the interaction tasks (Table 1). Although most of the interaction interfaces (apart from voice) involve the use of menus that require the selection of buttons, in this systematic review the studies classified under selection are the ones where selections are performed for the sole purpose of objects being selected and not when selecting buttons with other purposes. Please also note that a few studies [1, 36, 94, 100] did not specify the interaction tasks being performed, generally stating that the hands-free interfaces were used to interact with the VE. Fig. 7 shows the distribution of interaction sources per interaction task.

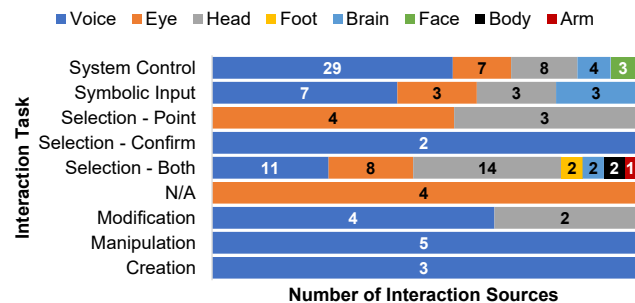


Fig. 7. Number of studied interaction sources per interaction task.

Besides navigation, which is out of the scope of this review, selection and system control are the two most performed interaction tasks in a VR application, being this reflected in the number of studies with those tasks. Given the number of tasks that can technically fit under the definition of system control, there is not a particular focus on the literature in a single task, with studies exploring a diversity of system control operations. Despite the diversity of actions and interfaces, for system control tasks, the most used interface is the voice by using speech commands, for instance, to perform actions like zooming on the data [3], to change the active game mode [87], and to navigate through menus [24, 55]. Regarding other methods, in [99] the players were able to use the breath to change the base functionality of the game. In [76] users could issue voice commands to start or stop the music playback. By using the eye states and head movements, in [59] users were able to perform zoom operations, scroll through data, and reset the VR system.

Regarding the selection tasks, the studies were divided by what kind of the selection process is being performed, i.e., fixating on one object (point), confirming the selection of the fixated object (confirm), or the full selection process (both). As the selection objective is usually the full process (fixating on an object and “selecting” it), most of the studies fall under this category. The most used interfaces for performing a complete selection are the head gaze and voice. In head gaze selections (e.g., [31, 69, 75, 114]) users point a reticle at the object they want to select which activates a timer. After the timer finishes, i.e., after pointing at an object for a certain amount of time (dwell time) the object that is being pointed at will become selected. using the voice for selection usually comprises a command that both fixates and confirms the selection in one action. For example, in [3, 13, 92] it is possible to say the name of the objects to be selected.

For the pointing task eye and head gaze are used. The confirming process needs to then be made using another method. For instance, in [85] users needed to point with either the head or eye but needed to press a key to confirm the selection. In [46], the eyes and head would be used to acquire a target and a controller button was used to shoot the target. In [8] users could also use their head and eye gaze for pointing at elements on a UI, with this study focusing only on this part of the selection process. The nature of these two interfaces makes them good candidates for this task as it is natural and easy to point at things with the eyes/head. For the confirmation task alone, only the use of the voice was found in the literature with a non-verbal “blow” action [119, 120] used for confirming the selection of an object being targeted with a

Table 5. Interaction tasks found on the surveyed studies. Task types based on the Table 1.

Interaction Task	N	Studies
System Control	42	[2, 3, 10–12, 16, 17, 21, 22, 24, 26, 27, 35, 47, 53, 55, 57, 59, 60, 62, 66, 70–72, 76, 78, 81, 82, 86–88, 92, 98, 99, 102–104, 108–110, 116, 117]
Selection	Both	29 [3, 4, 13, 23, 31–33, 35, 39, 45, 48, 59, 63, 66, 67, 69, 72, 75, 78, 83, 86, 90, 92, 98, 103, 113–115, 117]
	Point	4 [8, 38, 46, 85]
	Confirm	2 [119, 120]
Symbolic Input	13	[2, 12, 37, 42, 45, 66, 68, 72, 84, 95, 98, 106, 113]
Manipulation	5	[11, 25, 45, 63, 90]
Modification	5	[11, 15, 30, 45, 56]
Creation	3	[45, 90, 117]
N/A	4	[1, 36, 94, 100]

controller.

When performing symbolic input hands-free there are two alternatives studied in the literature. The first most used approach is by using speech recognition in conjunction with a text-to-speech service, which allows users to directly input text by spelling the words (e.g., [2, 84, 106]). The second is to use a virtual keyboard in the VE and use the selection methods (e.g., eye gaze [95] and head gaze [42]) to choose which letters to input. In [37] a novel approach was used, with a keyboard similar to a T9 keyboard consisting of three main keys that when selected would consecutively show a subset of the previous letters available until only one letter was selected, being the keys selected by a BCI with SSVEP. In [68] a combination of eye-tracking and BCI SSVEP was used to accurately select the keys.

Regarding other interaction tasks, in [25] it was possible to interact with a cloud of particles by “blowing” at them to manipulate their state, in this case, their displacement values. In [56] speech was used to modify the colors of objects by head gazing at them and then spelling the intended color. In the context of machine maintenance training [90], users could use speech to remove or replace machine parts and to open or close the machine views. With a broader focus, in [45] is also possible to modify colors of objects with speech after pointing at them with a non-hands-free interface as well as manipulating the on/off state of a lamp and create or destroy new objects in the VE by issuing the respective speech commands. No studies were found that used hands-free interaction for transformation tasks, which is expected since these tasks are better performed with NUI.

3.6 Evaluation Metrics

Table 6 shows the surveyed studies distributed by the respective metrics used for evaluation (note that some studies evaluate more than one variable, thus being represented more than once). From the 80 studies, 33 ($\approx 41\%$) did not provide an evaluation of the interfaces used. This represents a gap on the state-of-the-art as, without the evaluation of interfaces, it is not possible to ensure that the interfaces are adequate for interaction.

Usability metrics were subdivided by their components (i.e., satisfaction, efficiency, and efficacy), with most evaluation studies focusing on these metrics. The satisfaction component is measured with user questionnaires. Most of the studies ask users on their sentiment towards the interfaces being evaluated and their feedback on them (e.g., [37, 114, 115, 119]) and which interfaces they preferred to use (e.g., [31, 117]) through custom made questionnaires, some of which (e.g., [2, 85]) based on ISO 9241 [49]. A few validated questionnaires for user sentiment and preference were the User Experience Questionnaire [64] used by [98], the Positive and Negative Affect Schedule Survey [19] used by [71, 72], and the AttrackDiff Questionnaire [44] used in [25]. A few studies (e.g., [98, 120]) used the System Usability Scale Questionnaire [54], while others (e.g. [8, 67]) evaluated the interfaces cognitive load with the NASA Task Load Index [43].

Regarding efficiency, the most used evaluation metric is the time users take to complete tasks using the interfaces (e.g., [8, 35, 98]) and

respective speed [114, 115]. Performance through the number of tasks users can accomplish during a fixed period is also a preferred evaluation metric (e.g., [75, 113]), especially for symbolic input (e.g., [84]), which is also evaluated using the Information Transfer Rate ([37, 68]). Somewhat less used is Fitt’s Law [28] to calculate the performance throughput of selection tasks ([85, 119, 120]). Some studies also measure the users’ reaction times regarding interface operation [2, 8, 120]. Additionally, for pointing tasks, the movement and pointing times can also be measured [46, 85], as well as pointer trajectories [114].

Most studies that measure efficacy, do it by recording the number of errors or successes in performing the interaction tasks (e.g., [8, 31, 45, 85, 98, 120]). Some also measure it by reporting the accuracy of the interaction interfaces (e.g., [37, 56, 68, 115]). A study with evaluation based on ISO 9241 [2] recorded the number of times users asked for help during the tasks and, lastly, one study directly measured the precision of the interface [36].

Some studies evaluate the interfaces by the users’ physical state. Subjectively, some studies [11, 67, 76] evaluate the simulator sickness of the interfaces using the Simulator Sickness Questionnaire [58] while another [98] uses the Motion Sickness Assessment Questionnaire [34]. In [47], researchers have measured eye fatigue with a questionnaire and eye vergence objectively with the ET data. ET data was also used by [8] to measure eye movements and the HMD data for head movements. Finally, pupil dilation was also measured in [4, 75].

Regarding the mental state of users, custom made questionnaires are used to measure the sense of agency [110], the users’ emotional response [81], the concentration [23], and the sense of presence [11, 45, 99]. The sense of presence is also measured using the Slater-Usch-Steed Presence Questionnaire [105] in [98], the Presence Questionnaire [112] in [72, 90], and the Measurement, Effects, Conditions Spatial Presence Questionnaire [111] in [110]. Objectively, in [60] the relaxation time of users was measured with EEG data.

Finally, some studies also evaluated the performance of the interaction systems, where for instance, in [16] the accuracy in recognizing of mouth states was recorded, and in [113] the time until the system responded to a blink of the eye. Moreover, the recognition rates and system response times were also recorded in the detection of speech [56] and SSVEP [104].

3.7 Evaluation Results

Considering the studies where a comparison of interfaces was made (30 of 80), only 8 compared two or more hands-free interfaces, and 22 compared hands-free interfaces to one or more non-hands-free interfaces. From all studies, statistically significant results were found in 25, being these results recorded to understand how the interfaces perform. From the 7 studies that compared hands-free interfaces, and which had significant results, 5 evaluated selection tasks comparing head and eye gaze. On these, head gaze was consistently more efficient in [46, 75, 85] while eye gaze was more efficient in [8]. Regarding satisfaction, eye gaze was better for selection in [75, 83] and head gaze in [85]. These mixed results found when comparing head and eye gaze for selection,

Table 6. Evaluation metrics found on the surveyed studies.

Evaluation Metric		N	Studies
Usability	Satisfaction	31	[2,8,10,11,17,25,26,31,35,37,45,55,62,66,67,70–72,78,83–85,90,95,98,99,114,115,117,119,120]
	Efficiency	25	[2,8,10,31,35,37,45–47,55,60,66–68,75,83–85,95,98,113–115,119,120]
	Efficacy	23	[2,8,10,23,24,31,36,37,45,55,56,62,66–68,85,95,98,104,113,115,119,120]
Mental State	Subjective	9	[11,23,45,72,81,90,98,99,110]
	Objective	1	[60]
Physical State	Subjective	5	[11,47,67,76,98]
	Objective	4	[4,8,47,75]
System Performance		4	[16,56,104,113]
None		33	[1,3,12,13,15,21,22,27,30,32,33,38,39,42,48,53,57,59,63,69,82,86–88,92,94,100,102,103,106,108,109,116]

indicate that the best interface is dependent on the study design and on how the interfaces are configured, despite favoring head gaze for efficient selection. From the 13 studies that compared hands-free and non-hands-free interfaces, and which had significant results, 7 evaluated selection tasks. The non-hands-free interfaces (mainly controllers) were consistently better in most usability metrics in [67, 115, 119]. However, for system control voice was the preferred interface in terms of mental and physical state [76, 81, 99].

Comparing head gaze with speech for selection, symbolic input, and system control, in [98] speech scored better in all the usability metrics, despite head gazing being considered a more novel interface by users.

In [114], when comparing three head gaze selection confirmation techniques (one implemented by researchers, pass through the target and dwell time) against non-hands-free interfaces (controller and hand gestures), the implemented head gaze technique was more efficient than pass, but less than dwell time and the controller. At the same time, head gaze had better efficacy and satisfaction results, while hand gestures were more fatiguing.

For selection purposes, head gaze had better user satisfaction when compared to a controller [35], wherein [67] eye gaze was better in efficiency and satisfaction for the same task, while the controller was more accurate. In [115], head gaze was better in satisfaction, but the controller had better efficacy. In the context of smartphone VR, touching a capacitive button was better at efficiency and user satisfaction than head gazing with dwell time [31].

Comparing non-verbal “blow” with a controller and hand gestures, “blow” scored better in satisfaction and sense of presence [99], as well as efficiency and efficacy [119, 120]. When comparing controller with speech for system control, the speech interface was better in both having less simulator sickness [76] and engagement [81], while being worse in the sense of agency [110].

Additionally, in [45], speech was better in satisfaction and sense of presence and the controller was better in efficacy for selection, and at the same time speech was better across metrics for symbolic input. For symbolic input, it was found in [84], that when comparing speech to a controller or smartphone-based VR interaction, speech was better in efficiency and user satisfaction, and that when comparing eye gaze with a button and a keyboard, the eye gaze was worse in efficiency [95].

3.8 Opportunities

When surveying the literature, it was possible to understand that there are many opportunities for the research of hands-free interaction for immersive VR applications, regarding the interaction interfaces, their use cases, and most importantly their evaluation.

When analyzing the characteristics of the interfaces, there are still some limitations for them to be of practical use with VR. For instance, BCI still requires a fair amount of time and effort before being used to allow for calibration of the devices and respective data, so more research should be conducted to either reduce and improve the calibration phase or to make the calibration more appealing and enjoyable to users. For speech interfaces, there is the need to properly isolate the voice of users, improving the recognition rates, and reducing the false positives. Much

research is done to improve this aspect, for instance, with the use of activation triggers (e.g., button presses) and noise filters. However, in the context of hands-free VR, much of the studies assume users' have a non-hands-free interface available to act as a trigger, so further research should be done to address the use of hands-free triggers.

Contrary to what was expected, and to the best of our knowledge, only one study was found in the scope of this review that made use of a myographic interface. This indicates a gap in the literature as, with such interfaces, it is possible to associate different muscles or force amounts to specific actions in the VE. Additionally, very few studies made use of full-body tracking and body gestures, and while this is expected because of their use for navigation tasks, exploratory research should be conducted on how to bring this interaction method to other tasks (e.g., selection and system control). Moreover, the couple of studies that did use the body as an interface did so through the raw use of RGB/RGB-D camera sensors, requiring developers to implement the whole image processing pipeline. As such, the use of modern motion capture systems should provide better body tracking and enable more interaction possibilities.

Because of the division of the selection process in pointing and confirming and because of the previously mentioned “Midas Touch” problem, the selection tasks would benefit the most from multimodal interaction, in which despite possibly introducing some cognitive load [14], the pointing part of the process can be performed with one interface and the confirmation is done with another interface. A prime example and opportunity would be to use eye or head gaze to point at items and use speech commands to confirm or act upon the pointed item. However, this multimodal interaction was not found in any of the selected studies.

Regarding the graphical user interfaces (GUI), few studies account for the specificities of hands-free interfaces to activate the buttons and heavily rely on non-hands-free interfaces (e.g., controllers with 6 degrees of freedom tracking) to point and confirm button selections. Further research should be made to ensure those GUI can be properly interacted with in a hands-free manner. Not only that, but also ensure that interacting with GUI hands-free does respect the principles of interaction [80] and that GUI adapts to the different sources of interaction, providing users with intuitive and meaningful interactions.

Accessibility wise, in a very small number of studies it was found that hands-free interaction can bring VR experiences to people with specific motor disabilities that may still allow the use of, for example, speech or eye interaction. Hands-free interaction enables those people to properly interact with VR systems and take advantage of them. As such, despite not being the focus of this systematic review, we recognize a gap in the literature, regarding this topic.

Regarding evaluation, as few studies properly compare the hands-free interfaces, the main research opportunity is the creation of a testing framework that could account for the different interaction tasks and interfaces, enabling a formal evaluation and comparison of the interfaces using objective metrics and scientifically validated questionnaires (i.e., there was a formal validation of the psychometric properties of the instrument). Since the focus is also on VR, more studies should include evaluation of common VR metrics (i.e., the sense of presence and sim-

ulator sickness). The formal evaluations and proper data report would allow for future meta-analysis and to better conclude which interfaces are better for each task, benefiting researchers and VR developers.

Finally, complementary to this systematic review that provides a broad overview of the hands-free interfaces, a more detailed review should be made for each of the hands-free interaction sources to provide a better in-depth overview of those sources. This would help find details about how the interfaces are implemented to tackle specific problems and find their strengths and disadvantages.

4 CONCLUSIONS

This systematic review explored the current literature on the use of hands-free interaction techniques for immersive VR for interaction tasks other than navigation. The advantage of such techniques is that they allow users to still perform tasks in VR while the users have their hands available for grasping and directly interact with objects (virtual and tangible).

It was found that the most studied interaction method is the use of voice, mostly through speech commands and for system control tasks. This can be a result of the availability of microphones in common VR setups, as well as the fact that speech recognition systems are evermore better at detecting the intended commands. Secondly, eye gazing is the next most studied technique, as eye trackers can be seamlessly integrated into HMD and provide accurate eye data, it is natural to use the information of where users are looking mainly for selection purposes. Similarly, head gaze is also a good method for this since HMD inherently provide head rotation data. EEG devices allow the use of brain signals as a method of interaction with VR, particularly through the measurement of reactions to events happening on the VE. However, there is still great potential for researchers to explore new interaction techniques and improve the usability of current interfaces, making them more natural to users.

The system control and selection tasks are the most studied interaction tasks for performance evaluation as they reflect the most common tasks performed besides the excluded navigation tasks, allowing users to control the systems and navigate through virtual menus. Some studies focused on symbolic input, especially regarding voice interfaces and input through virtual keyboards, however, as more complex applications require users to input text, more studies must be carried for hands-free interfaces to reach the performance of keyboards. The versatility of the voice is verified, with it being used across all interaction tasks. This versatility also positions the voice as one of the best interfaces for hands-free interaction.

Satisfaction, efficiency, and efficacy are the most used evaluation metrics in studies to assess the usability of interfaces, usually through custom questionnaires that gather feedback from users about their preferences and use of the interfaces, with few studies relying on validated questionnaires. For efficiency measurement, interaction times are the most used method for evaluating performance, while accuracy and number of errors are the most used for efficacy. Although few studies consider the sense of presence and simulator sickness, the majority do not, so researchers should include these metrics when evaluating interfaces as they have an impact on the VR experiences.

Finally, there is still the need to address the interaction problems found with current hands-free interfaces, for instance, better integration of the interfaces with the VR setups, and to better adapt the virtual GUI to the properties these interfaces. Moreover, most studies do not compare the hands-free interfaces with other interfaces, and the results from the ones that do so present opposite results for the same interface and task. To allow future interface evaluation analysis, studies should provide comparisons and the respective usage context of the interfaces. This would also allow for better interface use by VR developers.

ACKNOWLEDGMENTS

This work was done under the project SFRH/BD/147813/2019 entitled Hand-free Interaction Framework for Virtual Reality Applications financed by the FCT - Fundação para a Ciência e a Tecnologia. This work is also financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and

Internationalisation - COMPETE 2020 Programme and by National Funds through the FCT within project POCI-01-0145-FEDER-028618 entitled PERFECT - Perceptual Equivalence in virtual Reality For authEntiC Training.

REFERENCES

- [1] K. Ahuja, R. Islam, V. Parashar, K. Dey, C. Harrison, and M. Goel. Eyespyvr: Interactive eye sensing using off-the-shelf, smartphone-based vr headsets. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2):1–10, jul 2018. doi: 10.1145/3214260
- [2] L. Alfaro, R. Linares, and J. Herrera. Scientific articles exploration system model based in immersive virtual reality and natural language processing techniques. *International Journal of Advanced Computer Science and Applications*, 9(7), 2018. doi: 10.14569/ijacsa.2018.090736
- [3] R. Arangarasan and G. N. Phillips. Modular approach of multimodal integration in a virtual environment. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pp. 331–336. IEEE Comput. Soc, 2002. doi: 10.1109/icmi.2002.1167017
- [4] P. Bækgaard, J. P. Hansen, K. Minakata, and I. S. MacKenzie. A fitts' law study of pupil dilations in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19. ACM, New York, NY, USA, jun 2019. doi: 10.1145/3314111.3319831
- [5] S. Baraldi, A. D. Bimbo, L. Landucci, and N. Torpei. Natural interaction. In *Encyclopedia of Database Systems*, pp. 1880–1885. Springer US, 2009. doi: 10.1007/978-0-387-39940-9_243
- [6] L. P. Berg and J. M. Vance. Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality*, 21(1):1–17, sep 2016. doi: 10.1007/s10055-016-0293-9
- [7] J. Bertram, J. Moskaliuk, and U. Cress. Virtual training: Making reality work? *Computers in Human Behavior*, 43:284–292, feb 2015. doi: 10.1016/j.chb.2014.10.032
- [8] 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 *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18. ACM Press, New York, NY, USA, 2018. doi: 10.1145/3206343.3206349
- [9] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. An introduction to 3-d user interface design. *Presence: Teleoperators and Virtual Environments*, 10(1):96–108, feb 2001. doi: 10.1162/105474601750182342
- [10] D. Brun, S. M. Ferreira, C. Gouin-Vallerand, and S. George. A mobile platform for controlling and interacting with a do-it-yourself smart eyewear. *International Journal of Pervasive Computing and Communications*, 13(1):41–61, apr 2017. doi: 10.1108/ijpcc-02-2017-0011
- [11] D. W. Carruth, C. R. Hudson, C. L. Bethel, M. Pleva, S. Ondas, and J. Juhar. Using HMD for immersive training of voice-based operation of small unmanned ground vehicles. In J. Y. Chen and G. Fragomeni, eds., *Virtual, Augmented and Mixed Reality. Applications and Case Studies*, pp. 34–46. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-21565-1_3
- [12] M. Cavallo, M. Dholakia, M. Havlena, K. Ocheltree, and M. Podlaseck. Dataspace: A reconfigurable hybrid reality environment for collaborative information analysis. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 145–153. IEEE, mar 2019. doi: 10.1109/vr.2019.8797733
- [13] S. Chabot, J. Drozdal, Y. Zhou, H. Su, and J. Braasch. Language learning in a cognitive and immersive environment using contextualized panoramic imagery. In C. Stephanidis, ed., *HCI International 2019 - Posters*, pp. 202–209. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-23525-3_26
- [14] F. Chen, N. Ruiz, E. Choi, J. Epps, M. A. Khawaja, R. Taib, B. Yin, and Y. Wang. Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems*, 2(4):1–36, dec 2012. doi: 10.1145/2395123.2395127
- [15] J. Choi, S. Jeong, and J. Ko. Xremul: An emulation environment for xr application development. In *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications*, HotMobile '19, p. 187. ACM, New York, NY, USA, feb 2019. doi: 10.1145/3301293.3309569
- [16] U. Ciftci, X. Zhang, and L. Tin. Partially occluded facial action recognition and interaction in virtual reality applications. In *2017 IEEE Interna-*

- tion Conference on Multimedia and Expo (ICME), pp. 715–720. IEEE, jul 2017. doi: 10.1109/icme.2017.8019545
- [17] L. Corrado, T. Pedro, J. D'Alpuim, S. Eloy, and M. S. Dias. VIAR-MODES Visualization and Interaction in Immersive Virtual Reality for the Architectural Design Process. In B. Martens, G. Wurzer, T. Grasl, W. Lorenz, and R. Schaffranek, eds., *ECAADE 2015: REAL TIME - EXTENDING THE REACH OF COMPUTATION, VOL 1*, pp. 125–134. Autodesk; Bentley; TU Wien, Fac Architecture & Reg Planning; Arch-Ing Akademie, 2015. 33rd International Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe), TU Wien, Vienna, AUSTRIA, SEP 16-18, 2015.
- [18] G. Cortes, E. Marchand, J. Ardouin, and A. Lecuyer. Increasing optical tracking workspace of VR applications using controlled cameras. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 22–25. IEEE, 2017. doi: 10.1109/3dUI.2017.7893313
- [19] J. R. Crawford and J. D. Henry. The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 43(3):245–265, sep 2004. doi: 10.1348/0144665031752934
- [20] R. Dachsel and A. Hübner. A Survey and Taxonomy of 3D Menu Techniques. In M. Lin and R. Hubbard, eds., *Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2006. doi: 10.2312/EGVE/EGVE06/089-099
- [21] M. Daily, M. Howard, J. Jerald, C. Lee, K. Martin, D. McInnes, and P. Tinker. Distributed design review in virtual environments. In *Proceedings of the Third International Conference on Collaborative Virtual Environments*, CVE '00, p. 57–63. ACM Press, New York, NY, USA, 2000. doi: 10.1145/351006.351013
- [22] A. Dingli and F. Haddod. Interacting with intelligent digital twins. In A. Marcus and W. Wang, eds., *Design, User Experience, and Usability. User Experience in Advanced Technological Environments*, pp. 3–15. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-23541-3_1
- [23] M. Donnerer and A. Steed. Using a p300 brain-computer interface in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 19(1):12–24, feb 2010. doi: 10.1162/pres.19.1.12
- [24] D. Dorozhkin and J. Vance. Implementing speech recognition in virtual reality. In *Mechanical Engineering Conference Presentations, Papers, and Proceedings*, vol. 30, Sept. 2002.
- [25] S. Eroglu, S. Gebhardt, P. Schmitz, D. Rausch, and T. W. Kuhlen. Fluid sketching—immersive sketching based on fluid flow. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 475–482. IEEE, mar 2018. doi: 10.1109/vr.2018.8446595
- [26] A. Ferracani, M. Faustino, G. X. Giannini, L. Landucci, and A. D. Bimbo. Natural experiences in museums through virtual reality and voice commands. In *Proceedings of the 25th ACM International Conference on Multimedia*, MM '17, p. 1233–1234. ACM Press, New York, NY, USA, 2017. doi: 10.1145/3123266.3127916
- [27] S. S. Fisher, M. McGreevy, J. Humphries, and W. Robinett. Virtual environment display system. In *Proceedings of the 1986 Workshop on Interactive 3D Graphics*, I3D '86, p. 77–87. ACM Press, New York, NY, USA, 1987. doi: 10.1145/319120.319127
- [28] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General*, 121(3):262–269, 1992. doi: 10.1037/0096-3445.121.3.262
- [29] P. Fuchs, G. Moreau, and P. Guitton, eds. *Virtual Reality: Concepts and Technologies*. CRC Press, Boca Raton, FL, July 2011.
- [30] L. Gallo, G. D. Pietro, A. Coronato, and I. Marra. Toward a natural interface to virtual medical imaging environments. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '08, p. 429–432. ACM Press, New York, NY, USA, 2008. doi: 10.1145/1385569.1385651
- [31] P. Ganapathi and K. Sorathia. 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*, MobileHCI '18, p. 166–173. ACM Press, New York, NY, USA, 2018. doi: 10.1145/3236112.3236136
- [32] F. Garzotto, M. Gelsomini, V. Matarazzo, N. Messina, and D. Occhiuto. XOOM: An end-user development tool for web-based wearable immersive virtual tours. In J. Cabot, R. De Virgilio, and R. Torlone, eds., *Web Engineering*, pp. 507–519. Springer International Publishing, Cham, 2017. doi: 10.1007/978-3-319-60131-1_36
- [33] M. Gelsomini, G. Leonardi, and F. Garzotto. Embodied learning in immersive smart spaces. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. ACM, New York, NY, USA, apr 2020. doi: 10.1145/3313831.3376667
- [34] P. J. Gianaros, E. R. Muth, J. T. Mordkoff, M. E. Levine, and R. M. Stern. A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space, and Environmental Medicine*, 72(2):115–119, Feb 2001.
- [35] I. Giannopoulos, A. Komninos, and J. Garofalakis. Natural interaction with large map interfaces in vr. In *Proceedings of the 21st Pan-Hellenic Conference on Informatics*, PCI 2017. ACM Press, New York, NY, USA, 2017. doi: 10.1145/3139367.3139424
- [36] S. W. Greenwald, L. Loreti, M. Funk, R. Zilberman, and P. Maes. Eye gaze tracking with google cardboard using purkinje images. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, p. 19–22. ACM Press, New York, NY, USA, 2016. doi: 10.1145/2993369.2993407
- [37] R. Grichnik, M. Benda, and I. Volosyak. A VR-based hybrid BCI using SSVEP and gesture input. In I. Rojas, G. Joya, and A. Catala, eds., *Advances in Computational Intelligence*, pp. 418–429. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-20521-8_35
- [38] A. Haffegge, V. Alexandrov, and R. Barrow. Eye tracking and gaze vector calculation within immersive virtual environments. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*, VRST '07, p. 225–226. ACM Press, New York, NY, USA, 2007. doi: 10.1145/1315184.1315232
- [39] H. Hakoda, W. Yamada, and H. Manabe. Eye tracking using built-in camera for smartphone-based HMD. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 15–16. ACM, New York, NY, USA, oct 2017. doi: 10.1145/3131785.3131809
- [40] C. Hand. A Survey of 3D Interaction Techniques. *Computer Graphics Forum*, 16(5):269–281, dec 1997. doi: 10.1111/1467-8659.00194
- [41] M. Handrahan. Oculus rift development kits pass 100,000 sales, Jul 2014.
- [42] J. P. Hansen, A. Alapetite, M. Thomsen, Z. Wang, K. Minakata, and G. Zhang. Head and gaze control of a telepresence robot with an hmd. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18. ACM, New York, NY, USA, jun 2018. doi: 10.1145/3204493.3208330
- [43] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Advances in Psychology*, vol. 52 of *Advances in Psychology*, pp. 139–183. Elsevier, 1988. doi: 10.1016/s0166-4115(08)62386-9
- [44] M. Hassenzahl, S. Diefenbach, and A. Göritz. Needs, affect, and interactive products – facets of user experience. *Interacting with Computers*, 22(5):353–362, sep 2010. doi: 10.1016/j.intcom.2010.04.002
- [45] 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. *Computers & Graphics*, 82:321–331, aug 2019. doi: 10.1016/j.cag.2019.06.003
- [46] 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 *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8. IEEE, sep 2019. doi: 10.1109/vs-games.2019.8864589
- [47] Y. Hirata, H. Soma, M. Takimoto, and Y. Kambayashi. Virtual space pointing based on vergence. In M. Kurosu, ed., *Human-Computer Interaction. Recognition and Interaction Technologies*, pp. 259–269. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-22643-5_21
- [48] M. Hohl. Radiomap: Experiential interactive environment. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, MULTIMEDIA '04, p. 995–996. ACM Press, New York, NY, USA, 2004. doi: 10.1145/1027527.1027767
- [49] Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. Standard, International Organization for Standardization, Geneva, CH, July 2019.
- [50] R. J. K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems (TOIS)*, 9(2):152–169, apr 1991. doi: 10.1145/123078.128728

- [51] A. Jaimes and N. Sebe. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*, 108(1-2):116–134, oct 2007. doi: 10.1016/j.cviu.2006.10.019
- [52] J. Jerald. *The VR Book: Human-Centered Design for Virtual Reality*. Association for Computing Machinery and Morgan & Claypool, 2015.
- [53] L. Jie, C. Jian, and W. Lei. Design of multi-mode UAV human-computer interaction system. In *2017 IEEE International Conference on Unmanned Systems (ICUS)*, pp. 353–357. IEEE, oct 2017. doi: 10.1109/icus.2017.8278368
- [54] P. Jordan. *Usability evaluation in industry*. Taylor & Francis, London Bristol, Pa, 1996.
- [55] R. Jota, B. R. de Araújo, L. C. Bruno, J. M. Pereira, and J. A. Jorge. Immiview: a multi-user solution for design review in real-time. *Journal of Real-Time Image Processing*, 5(2):91–107, 2010. doi: 10.1007/s11554-009-0141-1
- [56] E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th International Conference on Multimodal Interfaces, ICMI '03*, p. 12–19. ACM Press, New York, NY, USA, 2003. doi: 10.1145/958432.958438
- [57] S. Kakez, V. Conan, and P. Bisson. Virtually documented environments: A new interface paradigm for task-oriented access to information. *Computer Graphics Forum*, 16(3):C319–C327, sep 1997. doi: 10.1111/1467-8659.00169
- [58] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, jul 1993. doi: 10.1207/s15327108ijap0303_3
- [59] J. Kim, J. Cha, H. Lee, and S. Kim. Hand-free natural user interface for VR HMD with IR based facial gesture tracking sensor. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, VRST '17*. ACM, New York, NY, USA, nov 2017. doi: 10.1145/3139131.3143420
- [60] R. K. Kosuru, K. Lingelbach, M. Bui, and M. Vukelić. Mindtrain: How to train your mind with interactive technologies. In *Proceedings of Mensch und Computer 2019 on - MuC'19, MuC'19*, p. 643–647. ACM Press, New York, NY, USA, 2019. doi: 10.1145/3340764.3344884
- [61] A. Kronqvist, J. Jokinen, and R. Rousi. Evaluating the authenticity of virtual environments: Comparison of three devices. *Advances in Human-Computer Interaction*, 2016:1–14, 2016. doi: 10.1155/2016/2937632
- [62] D. Kumar and A. Sharma. Electrooculogram-based virtual reality game control using blink detection and gaze calibration. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2358–2362. IEEE, sep 2016. doi: 10.1109/icacci.2016.7732407
- [63] M. Latoschik, M. Frohlich, B. Jung, and I. Wachsmuth. Utilize speech and gestures to realize natural interaction in a virtual environment. In *IECON '98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No.98CH36200)*, vol. 4, pp. 2028–2033 vol.4. IEEE, 1998. doi: 10.1109/iecon.1998.724030
- [64] B. Laugwitz, T. Held, and M. Schrepp. Construction and evaluation of a user experience questionnaire. In A. Holzinger, ed., *HCI and Usability for Education and Work*, pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-89350-9_6
- [65] 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. Addison-Wesley, 2017.
- [66] R. Linares, J. Herrera, and L. Alfaro. AliciaVR: Exploration of scientific articles in an immersive virtual environment with natural user interfaces. In *2016 IEEE Ecuador Technical Chapters Meeting (ETCM)*, pp. 1–6. IEEE, oct 2016. doi: 10.1109/etcm.2016.7750829
- [67] F. L. Luro and V. Sundstedt. 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, ETRA '19*. ACM, New York, NY, USA, jun 2019. doi: 10.1145/3317956.3318153
- [68] X. Ma, Z. Yao, Y. Wang, W. Pei, and H. Chen. Combining brain-computer interface and eye tracking for high-speed text entry in virtual reality. In *23rd International Conference on Intelligent User Interfaces, IUI '18*, p. 263–267. ACM Press, New York, NY, USA, 2018. doi: 10.1145/3172944.3172988
- [69] D. Mardanbegi and T. Pfeiffer. Eyemrtk: A toolkit for developing eye gaze interactive applications in virtual and augmented reality. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, ETRA '19*. ACM, New York, NY, USA, jun 2019. doi: 10.1145/3317956.3318155
- [70] V. F. Martins, P. N. M. Sampaio, F. da S. Mendes, A. S. Lima, and M. de Paiva Guimarães. Usability and functionality assessment of an oculus rift in immersive and interactive systems using voice commands. In S. Lackey and R. Shumaker, eds., *Virtual, Augmented and Mixed Reality*, pp. 222–232. Springer International Publishing, Cham, 2016. doi: 10.1007/978-3-319-39907-2_21
- [71] W. McClinton, D. Caprio, D. Laesker, B. Pinto, S. Garcia, and M. Andujar. P300-based 3d brain painting in virtual reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA '19*, p. 1–6. ACM, New York, NY, USA, may 2019. doi: 10.1145/3290607.3312968
- [72] W. McClinton, S. Garcia, and M. Andujar. An immersive brain painting: The effects of brain painting in a virtual reality environment. In *Augmented Cognition*, pp. 436–445. Springer International Publishing, 2019. doi: 10.1007/978-3-030-22419-6_31
- [73] M. Melo, G. Goncalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa. Do multisensory stimuli benefit the virtual reality experience? a systematic review. *IEEE Transactions on Visualization & Computer Graphics*, (01):1–1, jul 2020. doi: 10.1109/TVCG.2020.3010088
- [74] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum*, 38(1):21–45, apr 2018. doi: 10.1111/cgf.13390
- [75] K. Minakata, J. P. Hansen, I. S. MacKenzie, P. Bækgaard, and V. Rajanna. Pointing by gaze, head, and foot in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, ETRA '19*. ACM, New York, NY, USA, jun 2019. doi: 10.1145/3317956.3318150
- [76] M. A. Mirzaei, J.-R. Chardonnet, F. Mérienne, and A. Genty. 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, VRIC '14*. ACM Press, New York, NY, USA, 2014. doi: 10.1145/2617841.2620716
- [77] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6(7):e1000097, jul 2009. doi: 10.1371/journal.pmed.1000097
- [78] A. Mower, R. Nguyen, and K. Frank. Evaluation of technology accessibility and user sentiment in learning through virtual reality modality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20*, p. 1–8. ACM, New York, NY, USA, apr 2020. doi: 10.1145/3334480.3375205
- [79] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *i-Perception*, 8(3):204166951770820, may 2017. doi: 10.1177/2041669517708205
- [80] D. Norman. *The design of everyday things*. Basic Books, New York, New York, 2013.
- [81] H. Osking and J. A. Doucette. Enhancing emotional effectiveness of virtual-reality experiences with voice control interfaces. In D. Beck, A. Peña-Rios, T. Ogle, D. Economou, M. Mentzelopoulos, L. Morgado, C. Eckhardt, J. Pirker, R. Koitz-Hristov, J. Richter, C. Gütl, and M. Gardner, eds., *Immersive Learning Research Network*, pp. 199–209. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-23089-0_15
- [82] R. Ott, M. Gutiérrez, D. Thalmann, and F. Vexo. Advanced virtual reality technologies for surveillance and security applications. In *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications, VRCIA '06*, p. 163–170. ACM Press, New York, NY, USA, 2006. doi: 10.1145/1128923.1128949
- [83] Y. S. Pai, T. Dingler, and K. Kunze. Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality*, 23(2):119–131, nov 2018. doi: 10.1007/s10055-018-0371-2
- [84] S. Pick, A. S. Puika, and T. W. Kuhlen. SWIFTER: Design and evaluation of a speech-based text input metaphor for immersive virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 109–112. IEEE, mar 2016. doi: 10.1109/3dui.2016.7460039
- [85] 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

- Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 91–98. ACM Press, New York, NY, USA, 2017. doi: 10.1145/3131277.3132182
- [86] C. D. Rojas Ferrer, H. Shishido, I. Kitahara, and Y. Kameda. Read-the-game: System for skill-based visual exploratory activity assessment with a full body virtual reality soccer simulation. *PLOS ONE*, 15(3):1–29, mar 2020. doi: 10.1371/journal.pone.0230042
- [87] K. Satter and A. Butler. Competitive usability analysis of immersive virtual environments in engineering design review. *Journal of Computing and Information Science in Engineering*, 15(3), sep 2015. 031001. doi: 10.1115/1.4029750
- [88] M. Schirski, A. Gerndt, T. van Reimersdahl, T. Kuhlen, P. Adomeit, O. Lang, S. Pischinger, and C. Bischof. Vista flowlib - framework for interactive visualization and exploration of unsteady flows in virtual environments. In *Proceedings of the Workshop on Virtual Environments 2003*, EGVE '03, p. 77–85. ACM Press, New York, NY, USA, 2003. doi: 10.1145/769953.769963
- [89] W. Schoor, F. Bollenbeck, T. Seidl, D. Weier, W. Weschke, B. Preim, U. Seiffert, and R. Mecke. VR Based Visualization and Exploration of Plant Biological Data. *JVRB - Journal of Virtual Reality and Broadcasting*, 6(2009)(8), 2010. doi: 10.20385/1860-2037/6.2009.8
- [90] B. L. Schroeder, S. K. T. Bailey, C. I. Johnson, and E. Gonzalez-Holland. Presence and usability do not directly predict procedural recall in virtual reality training. In C. Stephanidis, ed., *HCI International 2017 - Posters' Extended Abstracts*, pp. 54–61. Springer International Publishing, Cham, 2017. doi: 10.1007/978-3-319-58753-0_9
- [91] M. J. Schuemie, P. van der Straaten, M. Krijn, and C. A. van der Mast. Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201, apr 2001. doi: 10.1089/109493101300117884
- [92] R. Sharma, T. S. Huang, and V. I. Pavlović. *A Multimodal framework for Interacting with Virtual Environments*, pp. 53–71. Springer US, Boston, MA, 1996. doi: 10.1007/978-1-4613-1447-9_5
- [93] W. Sherman. *Understanding virtual reality : interface, application, and design*. Morgan Kaufmann, San Francisco, CA, 1 edition ed., Sept. 2003.
- [94] J. Shimizu and G. Chernyshov. Eye movement interactions in google cardboard using a low cost eog setup. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, p. 1773–1776. ACM, New York, NY, USA, sep 2016. doi: 10.1145/2968219.2968274
- [95] N. Sidorakis, G. A. Koulteris, and K. Mania. Binocular eye-tracking for the control of a 3d immersive multimedia user interface. In *2015 IEEE 1st Workshop on Everyday Virtual Reality (WEVR)*, pp. 15–18. IEEE, mar 2015. doi: 10.1109/wevr.2015.7151689
- [96] M. Slater and M. Usoh. Body Centred Interaction in Immersive Virtual Environments. In N. Magnenat-Thalmann and D. Thalmann, eds., *Artificial Life and Virtual Reality*, chap. 9. Wiley, 1 ed., 1994.
- [97] M. Slater and S. Wilbur. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, dec 1997. doi: 10.1162/pres.1997.6.6.603
- [98] M. Speicher, S. Cucerca, and A. Krüger. Vrschop: A mobile interactive virtual reality shopping environment combining the benefits of on- and offline shopping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–31, sep 2017. doi: 10.1145/3130967
- [99] M. Sra, X. Xu, and P. Maes. Breathvr: Leveraging breathing as a directly controlled interface for virtual reality games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. ACM Press, New York, NY, USA, 2018. doi: 10.1145/3173574.3173914
- [100] T. C. Stratmann, U. Gruenefeld, and S. Boll. Eyemr: Low-cost eye-tracking for rapid-prototyping in head-mounted mixed reality. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18. ACM, New York, NY, USA, jun 2018. doi: 10.1145/3204493.3208336
- [101] T4. Vr headset market share, Aug 2020.
- [102] J. T. C. Tan and T. Inamura. Embodied and multimodal human-robot interaction between virtual and real worlds. In *2013 IEEE RO-MAN*, pp. 296–297. IEEE, aug 2013. doi: 10.1109/roman.2013.6628431
- [103] M. Teofilo, J. Nascimento, J. Santos, Y. Albuquerque, A. L. Souza, and D. Nogueira. Bringing basic accessibility features to virtual reality context. In *2016 IEEE Virtual Reality (VR)*, pp. 293–294. IEEE, mar 2016. doi: 10.1109/VR.2016.7504769
- [104] H. Touyama and M. Hirose. Brain computer interface via stereoscopic images in CAVE. In *Human Interface and the Management of Information. Methods, Techniques and Tools in Information Design*, pp. 1004–1007. Springer Berlin Heidelberg, 2007. doi: 10.1007/978-3-540-73345-4_113
- [105] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, oct 2000. doi: 10.1162/105474600566989
- [106] J. Valentin, P. Torr, V. Vineet, M.-M. Cheng, D. Kim, J. Shotton, P. Kohli, M. Nießner, A. Criminisi, and S. Izadi. Semanticpaint: Interactive segmentation and learning of 3d world. In *ACM SIGGRAPH 2015 Talks*, SIGGRAPH '15. ACM Press, New York, NY, USA, 2015. doi: 10.1145/2775280.2792589
- [107] G. A. M. Vasiljevic and L. C. de Miranda. Brain-computer interface games based on consumer-grade EEG devices: A systematic literature review. *International Journal of Human-Computer Interaction*, 36(2):105–142, jun 2019. doi: 10.1080/10447318.2019.1612213
- [108] K.-J. Wang, Q. Liu, Y. Zhao, C. Y. Zheng, S. Vhasure, Q. Liu, P. Thakur, M. Sun, and Z.-H. Mao. Intelligent wearable virtual reality (VR) gaming controller for people with motor disabilities. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 161–164. IEEE, dec 2018. doi: 10.1109/aivr.2018.00034
- [109] K.-J. Wang, K. You, F. Chen, P. Thakur, M. Ulrich, S. Vhasure, and Z.-H. Mao. Development of seamless telepresence robot control methods to interact with the environment using physiological signals. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, p. 44. ACM, New York, NY, USA, mar 2018. doi: 10.1145/3173386.3177843
- [110] P. Winkler, P. Stiens, N. Rauh, T. Franke, and J. Krems. How latency, action modality and display modality influence the sense of agency: a virtual reality study. *Virtual Reality*, sep 2019. doi: 10.1007/s10055-019-00403-y
- [111] W. Wirth, T. Hartmann, S. Böcking, P. Vorderer, C. Klimmt, H. Schramm, T. Saari, J. Laarni, N. Ravaja, F. R. Gouveia, F. Biocca, A. Sacau, L. Jäncke, T. Baumgartner, and P. Jäncke. A process model of the formation of spatial presence experiences. *Media Psychology*, 9(3):493–525, may 2007. doi: 10.1080/15213260701283079
- [112] B. G. Witmer, C. J. Jerome, and M. J. Singer. The factor structure of the presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 14(3):298–312, jun 2005. doi: 10.1162/105474605323384654
- [113] J. Xiao, J. Qu, and Y. Li. An electrooculogram-based interaction method and its music-on-demand application in a virtual reality environment. *IEEE Access*, 7:22059–22070, 2019. doi: 10.1109/access.2019.2898324
- [114] Y. Yan, Y. Shi, C. Yu, and Y. Shi. Headcross: Exploring head-based crossing selection on head-mounted displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–22, mar 2020. doi: 10.1145/3380983
- [115] Y. Yan, C. Yu, X. Yi, and Y. Shi. Headgesture: Hands-free input approach leveraging head movements for hmd devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–23, dec 2018. doi: 10.1145/3287076
- [116] G. Yu, J. Liang, and L. Guo. Multi-modal interaction for space tele-science of fluid experiments. In *Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality*, AIVR 2018, p. 35–41. ACM Press, New York, NY, USA, 2018. doi: 10.1145/3293663.3293672
- [117] T. Zarraonandia, P. Díaz, A. Montero, and I. Aedo. Exploring the benefits of immersive end user development for virtual reality. In C. R. García, P. Caballero-Gil, M. Burmester, and A. Quesada-Arencibia, eds., *Ubiquitous Computing and Ambient Intelligence*, pp. 450–462. Springer International Publishing, Cham, 2016. doi: 10.1007/978-3-319-48746-5_46
- [118] M. A. Zayer, P. MacNeilage, and E. Folmer. Virtual Locomotion: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 26(6):2315–2334, jun 2020. doi: 10.1109/tvcg.2018.2887379
- [119] D. Zielasko, S. Freitag, D. Rausch, Y. C. Law, B. Weyers, and T. W. Kuhlen. Blowclick: A non-verbal vocal input metaphor for clicking. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, SUI '15, p. 20–23. ACM Press, New York, NY, USA, 2015. doi: 10.1145/2788940.2788953
- [120] D. Zielasko, N. Neha, B. Weyers, and T. W. Kuhlen. A reliable non-verbal vocal input metaphor for clicking. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 40–49. IEEE, 2017. doi: 10.1109/3dUI.2017.7893316