



Systematic Review of Comparative Studies of the Impact of Realism in Immersive Virtual Experiences

GUILHERME GONÇALVES, HUGO COELHO, and PEDRO MONTEIRO,

INESC TEC, Portugal and UTAD, Portugal

MIGUEL MELO, INESC TEC, Portugal

MAXIMINO BESSA, UTAD, Portugal and INESC TEC, Portugal

The adoption of immersive virtual experiences (IVEs) opened new research lines where the impact of realism is being studied, allowing developers to focus resources on realism factors proven to improve the user experience the most. We analyzed papers that compared different levels of realism and evaluated their impact on user experience. Exploratorily, we also synthesized the realism terms used by authors. From 1,300 initial documents, 79 met the eligibility criteria. Overall, most of the studies reported that higher realism has a positive impact on user experience. These data allow a better understanding of realism in IVEs, guiding future R&D.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; Human computer interaction (HCI); • **General and reference** → **Surveys and overviews**; • **Computing methodologies** → Computer graphics; • **Software and its engineering** → **Virtual worlds software**;

Additional Key Words and Phrases: Virtual reality, mixed reality, augmented reality, realism, immersive

ACM Reference format:

Guilherme Gonçalves, Hugo Coelho, Pedro Monteiro, Miguel Melo, and Maximino Bessa. 2022. Systematic Review of Comparative Studies of the Impact of Realism in Immersive Virtual Experiences. *ACM Comput. Surv.* 55, 6, Article 115 (December 2022), 36 pages.

<https://doi.org/10.1145/3533377>

1 INTRODUCTION

Over the years, the ability to recreate reality in virtual experiences has inspired several science fiction movies and series (e.g., *The Matrix* or *Star Trek*). Beyond the realm of science fiction, we have examples such as the Link trainer in 1937 (a device created to simulate flight conditions and train pilots) [59] and Sensorama in 1962 [53] (an entertainment device capable of delivering multisensory stimuli). Although the quality of the experiences is far from what is achievable with

This work was financed by National and European Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project SFRH/BD/148714/2019 entitled Automated Creation of Authentic Virtual Experiences. This work was also partially funded by the European Union's Horizon 2020 - The EU Framework Programme for Research and Innovation 2014-2020, under grant agreement No. 833573.

Authors' addresses: G. Gonçalves (corresponding author), H. Coelho, and P. Monteiro, INESC TEC, R. Dr. Roberto Frias, Porto, Porto, Portugal, 4200-465, UTAD, Quinta de Prados, Vila Real, Portugal, 5001-801; emails: guilhermeg@utad.pt, hugo.r.mendes@inesctec.pt, monteiro.p@outlook.pt; M. Melo, INESC TEC, R. Dr. Roberto Frias, Porto, Porto, Portugal, 4200-465; email: mcmelo@inesctec.pt; M. Bessa, UTAD, Quinta de Prados, Vila Real, Portugal, 5001-801, INESC TEC, R. Dr. Roberto Frias, Porto, Porto, Portugal, 4200-465; email: maxbessa@utad.pt.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2022 Copyright held by the owner/author(s).

0360-0300/2022/12-ART115

<https://doi.org/10.1145/3533377>

today's knowledge and technology, both aim to create realistic virtual experiences, even though with different end goals. In 1965, Ivan Sutherland introduced the concept "The ultimate display" [115]. The author states that the ultimate display would be "*a room within which the computer can control the existence of matter.*" Such a display would allow users to interact with what they see as real. Ivan Sutherland even gives examples such as "a chair being materialized allowing users to sit on it and even bullets could prove to be fatal." This type of display is very similar to the "Holodeck" depicted in the science fiction series *Star Trek*. Although very futuristic and almost unthinkable today, Ivan Sutherland proposed this display when no area-filling picture displays were available commercially for human use. Ivan Sutherland further speculated that new displays with such capabilities would eventually appear and that we would have much to learn about how to take advantage of them.

The higher computational power available today, at lower costs [56] and better scientific knowledge, allows us to have experiences that were considered science fiction years ago. Directly or indirectly, we are slowly approaching the ultimate display concept. Contrasting with Ivan Sutherland's reality in 1965 [115] "no one seriously proposes computer displays of smell or taste," we now have systems capable of delivering other stimuli beyond audiovisual, such as wind, temperature, scents, taste, or tactile feedback. These stimuli can help create experiences that replicate real-life experiences [74].

Although the computational technology capabilities are now higher than before (Moore's Law [81]), the simulation of reality in its most profound and most intrinsic detail is not conceivable yet. Trying to simulate everything in a virtual experience with the highest possible level of realism will waste computational resources and overload developers. Besides, the majority of those efforts might not be perceivable to humans. For example, Itti and Koch [43] discussed that only a small fraction of the stimulus received by our eyes is processed in a way that directly influences our behavior. The scene parts that elicit a stronger response ("visually salient") are the ones that get targeted first to be processed by the brain. Whether part of a scene provokes more or less response is thought to depend on the context of the environment. In other words, it is dependent on what else is presented in the other parts of the scene. Redirecting the attention towards other parts requires a voluntary "effort." Studies [12, 114] have also proven that tasks can grab the user attention in such a way that they are not able to perceive differences in quality on unrelated objects to the task. There are also the cross-modal effects of multisensory stimuli, where the perception of details of a given stimulus might go unnoticed when other more dominant stimuli are present [14].

These studies prove that user perception is not able to perceive the whole complexity of the virtual environment, because they can only process part of the stimuli at a given time. Therefore, increasing the environment realism as a whole might be a waste of resources. Instead, and by exploiting user perception, developers can shift their efforts and improve realism where it can be actually perceived. So, depending on the virtual experience's objective and the degree to which a user can interact, optimizing the resources available will always be crucial. Furthermore, understanding how users perceive realism and how it affects their virtual experience is critical in developing applications that require a faithful simulation of specific real-life conditions [20].

The quality of **immersive virtual experiences (IVEs)** is improving as the technological advancements happen. Adoption is increasing, as well as the easiness to create such IVEs through cost-effective hardware and straightforward authoring tools. We are now able to create IVEs so realistic and cost effective that they are being used as complements to professional training. However, we are still in the early days of investigation and limitations, whether human resources, system capabilities, or budget costs are a reality. Optimization of resources is thus important to extract the best IVEs possible within the given constraints. Therefore, it is crucial to understand which factors of objective realism do improve the user experience within IVEs, how many studies support the

results, and which variables are still unexplored. This systematic review will help developers to focus their efforts on objective realism factors that are proven to improve the user experience. In contrast, researchers can further corroborate and fill the literature gaps.

To properly perform a systematic review that focuses on realism in IVEs, we must understand several key components of an IVE. We describe two concepts that are widely used in this research topic: presence and immersion. Alongside these concepts, we further explore how different terms associated with realism are defined and used. To find these associated terms, we have looked up synonyms of realism and performed a preliminary literature search to check if the terms returned results in the scope of this study while also recording other terms used by authors.

This theoretical background will allow us to understand the relations between concepts better and provide a high-quality discussion.

1.1 Virtual Reality, Augmented Reality, and Mixed Reality

Some early definitions of **Virtual Reality (VR)** include: “Virtual Reality is electronic simulations of environments experienced via head-mounted eye goggles and wired clothing enabling the end-user to interact in realistic three-dimensional situations,” by Coates [16], and “Virtual Reality is an alternate world filled with computer-generated images that respond to human movements. These simulated environments are usually visited with the aid of an expensive data suit which features stereophonic video goggles and fiber-optic data gloves,” by Greenbaum [32]. A more recent definition of VR from Fuchs et al. [26] is: “Virtual reality is a scientific and technical domain that uses computer science and behavioural interfaces to simulate in a virtual world the behaviour of 3D entities, which interact in real-time with each other and with one or more users in pseudo-natural Immersion via sensorimotor channels.” In this work, we consider Fuchs et al. definition. However, because our work focuses on the higher end of immersion, it is constrained to include stereoscopy devices such as HMD or CAVEs. IVEs can originate from different immersive technology beyond VR. Thus, we also considered the **Mixed Reality (MR)** spectrum. Considering the *virtuality continuum* [76], MR consists of merging virtual and real-world stimuli [26, 76]. AR consists of overlaying virtual elements over the real world, positioning it closer to the real-world end of the continuum, while VR is on the opposite side of the continuum (purely virtual). MR, therefore, encompasses everything in between, including the known **Augmented Reality (AR)**.

1.2 Presence

The sense of presence seems to be one of the main results of virtual experiences. It has been used in the literature as a metric to evaluate the user experience and consequently the virtual environments, guiding future research. Several authors defined presence in different ways throughout the literature. Gibson [28] referred to presence as “the feeling of being in an environment”; Witmer and Singer [124] considered presence as being the “the subjective experience of being in one place or environment, even when one is physically situated in another”; Slater and Wilbur [112] defined presence as “a state of consciousness, the (psychological) sense of being in the virtual environment.” According to the authors, users who feel higher levels of presence in a virtual experience are also more prone to behave similarly to how they would in a real situation. Harter et al. [37] consider that the same autonomic reactions users have on real situations may also happen in virtual experiences as long as there is the suspension of disbelief (even though users know the virtual experience is not real, they suspend disbelief). David et al. [86] in his work about VR training with firefighters, considers that this mental suspension of disbelief is related to higher levels of presence. These results suggest that presence is particularly crucial in VR training scenarios, requiring users to behave in the IVE and real life in the most similar way possible. However, note that realistic behavior is not presence but rather a sign of it [104].

Several conceptualizations of presence have emerged throughout the literature. An explanation of several of those can be found in Lombard and Ditton's work [60]. Lombard et al. introduce presence as realism: "a medium can produce seemingly accurate representations of objects, events, and people—representations that look, sound, and/or feel like the "real" thing" [60]. The authors also argue that this concept can be confused with "social realism," which refers to the extent to which a media is plausible by reflecting events that could occur in a non-mediated world. Although presence as realism can include "social realism," it also includes "perceptual realism." For example, in a sci-fi scene, some events are unlikely to happen in reality, and therefore the social presence can be low. Nevertheless, if people and objects behave as expected in the same scene, then perceptual realism would have a high value even if the social presence is low.

In 2009, Mel Slater [105] introduced two constructs: "**place illusion**" (PI) and "**plausibility illusion**" (PsI). PI is considered the "strong illusion of being in a place despite the sure knowledge that you are not there," which is equivalent to what is usually called presence. The author considers PI as being binary; either there is an illusion or not, there cannot be a partial illusion. PsI is the "correlations between external events not directly caused by the participant and his/her sensations (both exteroceptive and interoceptive)." PsI does not require physical realism as people can still show signs of anxiety and other emotions depending on what is happening in the virtual experience, even though virtual realism is low. An example is the Stanley Milgram obedience experiment [77], where users would show signs of anxiety when causing pain to a non-realistic virtual avatar [107]. According to Mel Slater, if we experience both PI and PsI, then we are more likely to respond as if the situation was real.

Although presence is closely related to realism, Baños et al. [3] argue that reality judgment ("the belief that our experiences are real") is a construct different from presence. This means that users can consider an IVE real even if they do not feel any sense of presence and vice versa. Under this definition, Skarbez et al. [103] further suggest that reality judgment is synonymous with PsI.

1.3 Immersion

The term immersion in the context of VR is often confused with presence, being used sometimes interchangeably [103]. According to Slater et al. [112], immersion refers to the technology (objective characteristics of the systems that make a virtual experience possible) giving an inclusive (extent to which the user is isolated from physical reality), extensive (number of sensory stimuli being used), surrounding (field of view), and vivid (richness, information content, resolution, and quality of the displays) experience to users. The authors [112] also state that immersion requires a virtual body as it is part of the perceived environment and the one that is doing the perceiving. This immersion definition can also be referred to as "perceptual immersion" [6]. In their work from 2009, Slater et al. [105] further extended the immersion definition, characterizing it by the supported **sensorimotor contingencies (SCs)**. SCs are actions we perform to perceive the environment around us (e.g., looking around, crouching, bending down, jumping). Valid sensorimotor actions are actions allowed by the system, resulting in the images being updated accordingly. Valid effectual actions are actions users can take that result in changes in the environments. Both sets of actions are called valid actions, which users can perform, resulting in changes in their perception or modifications to the virtual environment. For example, if users try to bend down, but the HMD only allows for rotational movements, then the image will not be updated accordingly; therefore, this action would not be a valid sensorimotor action. Similarly, if the user tries to grasp a virtual object but there is no hand tracking, this action would not be a valid effectual action. Slater et al. [105] state that an ideal immersive system should be able to simulate a non-immersive system fully. For example, an immersive system comprising an HMD, could simulate a non-immersive system such as a desktop monitor, mouse, and keyboard. Still, the other way around would not be possible.

Note that Slater defines immersion as purely objective and dependent on the physical properties of the system.

A contrasting definition of immersion comes from Witmer and Singer [124]. They define immersion as “a psychological state characterised by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences.” This definition can also be called “psychological immersion” [61]. We consider Mel Slater’s immersion definition for this work’s scope because it clearly separates presence (subjective feeling) from immersion (objective characteristics of the system).

1.4 Fidelity

Several definitions have been given to fidelity. Franzluebbers and Johnsen [24] considered fidelity as how well the object physical properties are replicated. Meyer et al. [75] state that fidelity is a measure of how well the simulation represents the real world. To objectively evaluate fidelity, the authors affirm that a “referent” is required—“an abstract description of the real world that defines reality in a level of detail and format that makes a meaningful evaluation possible.” Because the real world is too complex to be compared, a referent would define reality through an abstraction of the real world, to which the simulation system can be compared to and evaluated.

Bowman and McMahan [20] discuss that realism comes from high-fidelity sensory stimuli in successful immersive VR applications. McMahan et al. [72] divide a VR environments’ overall fidelity into display fidelity (exactness in which the real-world stimuli is represented) and interaction fidelity (exactness that the real-world interaction can be reproduced). The authors also concluded that both constructs are important to determine performance, presence, engagement, and usability.

Alexander et al. [2] define fidelity as how well the real world is emulated. The author also defines various subcategories of fidelity, which can be divided into three subcategories. Physical Fidelity refers to how the simulation is physically similar to the real environment being replicated. However, a virtual environment can look similar to the real one until we try to interact with it. Functional Fidelity defines how similar the functionality of the virtual environment is to the real one. Last, Alexander et al. [2] describes Psychological Fidelity, which, as the name describes, defines how close the simulation can replicate the psychological factors that happen in the direct real environment.

Schricker et al.’s [99] definition is similar to Meyer et al. [75], as they describe fidelity as the extent to which a simulation represents its referent. The referent is an abstraction that is as close to the real world as possible, with no concern for design issues or resource limitations while a model usually will not be as close to the real world because it must make these considerations. However, the model is what it’s actually developed, and therefore, will not be as close to the real world, because it must consider the inherent limitations.

On the fidelity referent topic, Roza et al. [97] considers it as a “formal specification of all knowledge about reality plus indicators to determine the uncertainty levels and quality of this knowledge to judge the confidence level of this referent data.” The authors state that for Fidelity to be measured it must consider a common referent to be compared to.

Similarly to Schricker et al. [99], Meyer et al. [75], Roza et al. [97], and Hughes and Rolek [41] define fidelity as how well the attributes and behaviors of the referent are reproduced. Regarding fidelity measurement, Vincenzi et al. [83] describe two main methods to do it. We should note that fidelity for his work context is related to simulations (e.g., airplane cockpits). One method to measure fidelity can be a mathematical calculation of how many identical elements are shared between the real and virtual worlds. A higher simulation fidelity is then the result of a greater number of shared identical elements. The other methods refer to the trainee’s performance. By

evaluating the transfer of knowledge through comparing his/her performance in the simulation with the real world, it is possible to measure fidelity indirectly.

1.5 Coherence, Authenticity, and Credibility

Coherence is defined by Skarbez et al. [102] as “the set of objectively reasonable circumstances that the scenario can demonstrate without introducing objectively unreasonable circumstances.” The authors link coherence to PsI, as both depend on the users’ expectations and past experiences and highly depend on the specific virtual scenario and/or software. It does not require the virtual experience to replicate the real world. The authors further divide coherence in physical coherence (laws of physics) and narrative coherence (how virtual agents and scenario meets the users’ expectations as they know it from everyday life) which seems to overlap with Lombard et al.’s perceptual realism [60]. Skarbez et al. [103] also discuss that coherence can not be purely objective, as it will always depend on the user’s past knowledge and experiences. However, they consider that it would be helpful to consider as if it was objective, because developers cannot control the user’s prior knowledge. Still, they can control whether or not the virtual experience events are internally consistent. Furthermore, developers can further minimise how dependent coherence is on their prior knowledge by hinting the user to expect certain behaviors.

The term authenticity is considered by Gilbert [29] as how well the virtual experience is within the expectations of the user consciously and unconsciously. Malliet et al. [66] referred to authenticity as a constituent of perceived video realism, which was based on Hall’s [35] typicality (events and behaviors related to everyday life) and plausibility (events and behaviors that can potentially occur in the real world) notions. Malliet et al. [66] consider the sense of realism as “authenticity” where it “relates to the belief gamers have in a game designer’s intention and ability to express something real.” For example, the simulation of human emotion or the virtual characters and storylines of humanness is considered the authenticity construct. Skarbez et al. [103] discussed how Gilbert’s authenticity is similar to coherence but distinct as it assumes the virtual experience is trying to replicate the real world on the contrary to coherence.

Credibility is considered by Boucaud et al. [10] as the degree to which the user felt the virtual agent behaved in a human-like way. The scope of his study was about social touch in human-agent interaction. Building on their definition, the more overall meaning of credibility would be if the virtual world behaves similarly to the real world, which ultimately is equivalent to Gilbert’s authenticity and similar to the definition of Malliet et al. Bishop et al. [7] stated that the virtual experience would become more credible as data, computer display environments, and haptic feedback improve. Although the authors did not directly define credibility, it seems to be a function of the system capabilities (immersion/fidelity). Gonçalves et al. [30] used a dictionary definition to define credibility as “the quality of inspiring belief,” which is necessary to develop a sense of presence. This definition is equivalent to coherence and PsI, where there are no assumptions that the virtual environment is trying to replicate reality, but instead if the events and the environment itself follows its internal logic and meets the users’ expectations for the given context.

1.6 Realism

In the scope of visual stimulus, Chalmers and Ferko [13] considered realism in real-time the “holy grail” of computer graphics. VR systems aim at providing users with both realism and real-time update of visual stimuli. Creating a single realistic computer rendered image uses fewer resources than producing the same level of realism at a high and constant frame rate required by IVEs. There is a tradeoff between quality and system performance. Usually, realism is reduced to achieve decent performance. Authors state that no one can precisely define what realism is. They introduce a way of considering realism in virtual environments—**Levels of Realism (LoR)**. This is the level necessary to achieve the same experience in the virtual environment as in the real environment.

These can be particularly useful in training simulations to guarantee that trainees do not adopt different strategies just because they are in a virtual environment, which would ultimately harm their real-world performance in the same task. The authors also introduce the concept of believable realism. This is achieved by introducing imperfection in the virtual environment. Usually, computer-rendered images look intact and clean, which does not happen in the real world. Stains, dust, scratches, and other imperfections can affect the sense of perceived realism [62].

Ferwerda [23] defined three types of realism in computer graphics: physical (“image provides the same visual stimulation as the scene”), photorealism (“image produces the same visual response as the scene”), and functional realism (“image provides the same visual information as the scene”). Physical realism requires that virtual scenes be objectively the same as the real scene. Such would require the geometry, textures, lighting, and so on, to be done with extreme accuracy. A photorealistic image is an image indistinguishable from a real photograph of a real scene. Note that a photograph is a representation of reality and not reality itself. Introducing certain artefacts present in real photographs in computer-generated ones can actually increase its photorealism [62]. Functional realism is an image that provides meaningful information about objects’ properties, allowing users to perform real tasks. An example would be instructions to put together pieces of furniture.

Chalmers and Ferko [13], using physics and believability as “dimensions,” classify realism into four quadrants: **Is Believable (IB)**, **Not Believable (NB)**, **Is Physics (IP)**, **Not Physics (NP)**. For example, a VR system for training should fall in the IP quadrant, as it should introduce real-world physics, but it can (IB) or not be believable (NB). Ferwerda’s Functional realism and Uncanny Valley Avatares (explained ahead) are placed into NBIP, because they are physically correct but are not believable. Photographs and Photorealism are placed in IBIP, whereas VR is placed between IBIP and NBIP.

Perroud et al. [88] consider five acceptations of realism:

- (1) Realistic looking - This realism factor is mainly linked to visuals such as textures and how light is modeled in similar ways to what we see in the real world.
- (2) Realistic construction of the virtual world - In virtual environments, physics rules can easily be broken through simplification of physics calculations or even by offsetting known constants to achieve specific behaviors. This realism factor describes the extent the behavior of the virtual environment is based on scientifically proven models.
- (3) Physiologic realism - Due to the extensive complexity of the real world and the human body, the stimuli received by the body in a virtual environment may be oversimplified, lacking, or too intense. Physiologic realism refers to this question by defining how similar the sensory input in the virtual environment is to what the user would receive in a real analogous situation.
- (4) Psychological realism - Subjectively to users, certain settings (such as field of view or gravity) may seem realistic to them even if their levels are below or above the ones found in the real world. Psychological realism considers what seems real to the users, whether or not the virtual environment parameters replicate the real-world conditions.
- (5) Presence - Even in simple rudimentary virtual environments, users can feel a sense of presence. As Perroud et al. [88] state, “even if the scene is only made of non-textured polygons, the maximum the presence, the better.”

Furthermore, Hoorn et al. [40] give another realism definition stating that if the simulation’s goal is achieved, then the simulation was realistic enough. They mention that there is a lack of objective methods to evaluate physiological realism. They proposed a model to evaluate physiological realism using a driving simulation context considering the human visual system skills. The authors point out that the model needs to go through more experimentation to validate the scoring system fully.

According to Slater et al. [108], visual realism can be divided into geometric realism (how similar the virtual object is to its real counterpart) and illumination realism (how realistic the lighting is). Their study concluded that a higher sense of presence is more likely with real-time ray tracing than with ray casting. In the follow-up study, Yu et al. [129] concluded that the previous results were due to dynamic shadows and reflections. Also, that illumination quality did not affect presence, but global illumination did result in greater plausibility. Such did affect participant responses leaving the authors to conclude that it is worth introducing global illumination. Hvass et al. [42] conducted a study comparing low vs. high visual realism (through polygon count and texture resolution). The results suggested that a stronger feeling of presence develops when participants are exposed to heightened visual realism. Furthermore, self-reports and physiological measures also indicated a stronger fear response in higher levels of realism, which leads authors to conclude that such may reveal a stronger presence. The sense of presence seems to be a well-established metric and a relevant factor in evaluating perceived realism. Studies seem to indicate its increase usually leads to a higher sense of presence. However, there are cases where realism might not lead to presence and where realism without presence is enough. Bowman and McMahan [20] discussed one example: the oil and gas industry, where users need to perceive complex 3D structures to make the right decisions, not requiring a high feeling of presence but rather a high level of realism. Another possible issue with higher realism is the Uncanny Valley [82]—the emotional response of a person over an avatar reassembling a real human being. Widely used across the robotics and computer graphics fields, it explains the phenomena where users' feeling of empathy increases with the degree that an avatar reassembles an actual human until this feeling shifts abruptly to revulsion in the so-called "valley." However, if the reassembling keeps increasing, then the feeling of empathy returns. When designing virtual experiences with human-like elements, such as virtual agents or self-avatars, the uncanny valley might negatively affect the virtual experience if their realism is brought up to the point where it is eerily similar to a human being. In these cases, it could be better to use a less-realistic avatar or virtual agent.

1.7 Summarizing Realism's Terms

Considering author definitions of realism, we theorize that realism can be both subjective and objective. Subjective realism depends on how users perceive the environment. This means that for the same system and experience, users can perceive different levels of realism, depending on several variables such as past life experiences and physiological differences. However, objective realism is how well the system (hardware and software) provides the same stimuli as the real world, whether users perceive it as real or not. There is a consensus that fidelity defines how a system can replicate the real-world experience. However, because reality is too complex, authors use a "referent" (an abstraction of reality) to make a comparison between the virtual and real experience possible. Slater's definition of immersion is closely related to fidelity. However, we distinguish between both terms where immersion is the system potential to provide a realistic virtual experience (defined by the supported SCs), and fidelity is how well that system is used in this regard. For example, a highly immersive system could allow full-body tracking with extreme precision, but if the IVE application only makes use of head tracking with a bad software implementation, then such application will be of lower fidelity compared to an application where full-body tracking is used together with a highly optimized software implementation within the same immersive system. Following this definition, fidelity is therefore equal to objective realism.

Ferwerda's physical realism is oriented to the visual system. Nevertheless, if we extend its definition to encompass the rest of the human senses—the virtual scene produces the same stimuli as the real counterpart scene—then it is similar to our definition of objective realism. Likewise, Perroud et al.'s dimensions of physiologic realism and realistic construction of the virtual world

[88] also fit within our notion of objective realism. To objectively evaluate an IVE realism, the scene must be possible to be replicated in the real world. For example, the realism of a partial fantasy world (an experience that can only be partially reproduced in the real world) could be objectively evaluated only in the parts shared between the real world (e.g., gravity or light behavior). The rest of the fantasy elements (e.g., humans being able to fly on their own) cannot be replicated in the real world and, therefore, have no realism is zero, even though they can still be coherent/credible within the internal logic of the experience and provide a sense of presence.

Gilbert's [29] authenticity, Skarbez et al.'s [102, 103] coherence, Slater's PsI [105], Lombard et al.'s [60] perceptual realism, and Perroud et al.'s psychological realism [88] are closely related to subjective realism—how well the virtual experience meets expectations of users of what they consider to be real. Subjective realism can therefore be high even if the IVE does not replicate the real world. This means that users can perceive a fantasy world to be realistic even if it depicts events that would be impossible to happen in real life, as long as they make sense within the logic of the scenario context. As per Gilbert's definition, the exception being authenticity makes assumptions that the IVE is trying to replicate the real world.

In short, we divide realism into subjective and objective realism, which we will use throughout the rest of the article. Subjective realism considers how the users' expectations are met and ultimately how they perceive the experience as real. Objective realism considers how close to reality the virtual experience was, discarding whether or not the users thought it was real or not. It is objectively defined by the system characteristics and how they are used.

We found different terms and definitions used to describe realism; thus, to avoid possible misinterpretations between different studies, we reviewed the terms and definitions from this systematic review of selected papers.

2 RESEARCH QUESTIONS

Two **research questions (RQ)** were developed to tackle the issues identified in the previous sections. Various authors have different interpretations of the same realism related terms that can cause misinterpretations. The first research question (RQ1 - How do authors define realism and related terms in this systematic review selected papers?) will address terms related to realism in IVEs in an exploratory way. It will give readers an overview of terms and definitions used in the scope of this article to better orient future work and foster systematic use of such terms.

The second research question (RQ2 - How does objective realism impact the user experience?) aims to provide knowledge on which realism factors were studied, which categories of user experience were covered, and the impact of such realism factors on user experience in IVEs. This RQ will enable future development to create realistic IVEs to better focus efforts on factors known to increase the user's subjective perception of realism and overall experience. It will also suggest new lines of research regarding unstudied factors.

3 SYSTEMATIC REVIEW METHODOLOGY

The systematic review methodology used was based on the PRISMA methodology [79]. It encompasses four phases (Figure 1(a): identification, screening, eligibility, and included documents. It provides transparency in the paper selection through extensive documentation of all steps taken.

3.1 Eligibility Criteria

The eligibility criteria (Table 1) allow identifying related studies that can answer the RQ and discard studies that do not. Because the criteria are defined before the document search, the chance of bias is lowered [48]. The criteria were defined to restrict the number of documents to the most relevant ones to answer the RQ.

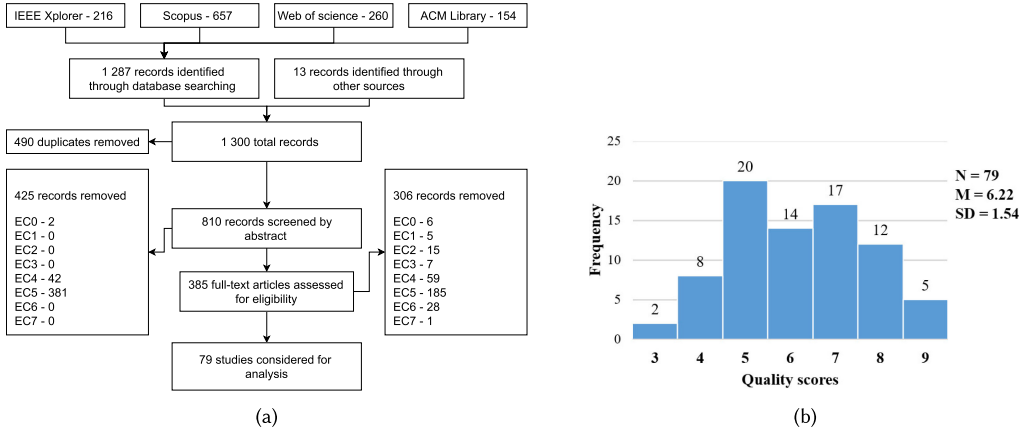


Fig. 1. Study selection flow diagram (a) and Quality assessment scores histogram (b).

Table 1. Inclusion (IC) and Exclusion Criteria (EC)

Criteria	Description
IC0	The title, abstract, or keywords match the search query (Section 3.2).
IC1	Search results from manual search.
IC2	Work published in refereed journal or conference.
IC3	The paper is written in English.
EC0	Duplicate studies.
EC1	Work not published in refereed journal or conference.
EC2	Text is not available.
EC3	The paper is not written in English.
EC4	Does not consider immersive VR/AR/MR.
EC5	The document does not consider the evaluation of realism of the virtual experience in a comparative study.
EC6	Does not compare two or more conditions between the same visual immersive setup.
EC7	Out of scope.

3.2 Identification Phase

Four well-established databases were included in the search phase: IEEE Xplore, Elsevier Scopus, Clarivate World of Science, and ACM Digital Library. The research team considered that these four databases cover all the venues where relevant studies for the systematic review scope are published. The search was conducted through Titles/Abstracts/Keywords. The query was built with the research questions in mind. Because the survey focuses on immersive VR, AR, and MR, the search term “virtual reality,” “augmented reality,” and “mixed reality” had to be found together with one of the following terms: immersive, HMD, variations of head-mounted, CAVE, or headset. Related terms had to be included to narrow the obtained results of the search and orient it towards realism. After several iterations, the final version of the query was the following:

(“virtual reality” OR “augmented reality” OR “mixed reality” OR “VR” OR “AR” OR “MR”) AND (“immersive” OR “cave” OR “hmd” OR “head mounted” OR “head mount” OR “head-mounted” OR “head-mount” OR “headset” AND (“Authentic” OR “Authenticity” OR “Credibility” OR “Coherent” OR “Credible” OR “Believable” OR “Coherence” OR “Fidelity” OR

Table 2. Data Extracted for Each Research Question

RQ	Data	Description
RQ1	Term	Which term authors used when referring “close to reality.”
RQ1	Term Definition	The definition of such term (if provided).
RQ2	Dependent Variables	Variables being tested and measured.
RQ2	Independent Variables	Variables that were being controlled.
RQ2	Results	How the dependent variables were influenced by the experimental groups.
RQ2	Highlights	The main conclusions of the study.

“Aesthetics” OR “Realism” OR “Realness”) AND (“Measure” OR “Measurement” OR “Evaluate” OR “Evaluation” OR “Quantify” OR “Quantification” OR “Estimate” OR “Estimation” OR “Assess” OR “Assessment” OR “Methodology” OR “Method” OR “Framework”)

The search was conducted on April 6, 2020, and had no time range limitations. A total of 1,287 documents were added to the screening phase. In addition, a search in additional sources was done to verify if any more documents to be included were not picked up in the selected databases’ search. As a result, the search identified and added 13 more documents, resulting in a total of 1,300.

3.3 Screening Phase

From the initial 1,300 documents, 490 were duplicates, resulting in 810 unique works for the screening phase. Papers with the same title would be further investigated as possible duplicates, and if the abstract was identical, it was considered a duplicate. In the case of similar abstracts, the full-text analysis would reveal if the studies were duplicates or not. The screening was performed individually by three researchers in an unblinded standard way. It consists of reading the abstracts to determine which ones are relevant and which ones are not. Two researchers read each abstract, and if there was a consensus between researchers, the document was accepted or rejected from further analysis. In case there was no consensus, a third researcher was assigned to decide by the majority. The screening phase excluded 425 documents (through exclusion criteria), resulting in 385 documents for full-text reading.

3.4 Eligibility Phase

The resulting papers underwent a full-text assessment guided by the exclusion criteria to assess their eligibility. From the 385 documents left from the screening phase, 306 were further excluded through exclusion criteria. The papers that fulfilled the eligibility criteria were processed to extract all the data necessary to answer the RQs and conduct a quality assessment. Thus, from the initial 1,300 documents, 79 were deemed fit for data extraction for the scope of this systematic review.

3.5 Data Retrieval

After reading the full-text documents, data was gathered through predefined forms. The variables included were aimed at answering each research question (Table 2). Data records were extracted to answer RQ1, the author’s terms and definition (if available). For RQ2, the dependent and independent variables will be considered as well as the results and study highlights. The authors’ study limitations and future work will be considered to discuss this systematic review results better.

3.6 Document Quality Assessment

The document quality assessment allows for an objective evaluation of how reliable studies are and to what extent they can be trusted. During the full-text analysis of the documents, a quality assessment was conducted. The scoring system was similar to the approaches by Connolly [17],

Feng [22], and Melo et al. [74]. Similarly to the study selection, and to avoid bias, the scoring was conducted by two researchers. If in consensus, then the rating would be closed; if not, then a third reviewer would moderate the scoring and provide closure. The scoring system consists of three items rated from 1 (lowest) to 3 (highest). The final score is constituted by the sum of the scores from the three items. The items cover the type of document, sample size, and how robust and relevant the methodology was for the study itself:

Type of document: conference papers get one point, papers published in the third quartile or inferior journals get two points, papers published in second- or first-quartile journals get three points (quartiles at publication time).

Sample size: sample sizes lower than 6 per group get one point, between 6 to 11 get two points, higher than 11 get three points (sample size recommendation based on Macefield [64], considering the lower end of the baseline range of problem discovery group size and higher end of the baseline range comparative study baseline range).

Methodology: evaluated through the analysis of the instruments, materials and procedures, limitations and possibility of biases, the extent to which the study is valid and replicable. Studies that present severe issues (e.g., issues that can compromise the entirety of the results without being presented as limitations or flaws in describing the methodology to the point that cannot be replicated) in this matter get one point. Studies that present issues that do not critically affect the trustworthiness of the results (e.g., flaws that can affect the results but are well documented as study limitations) get two points, studies that only present minor issues that do not compromise the results (all studies that did not fall into the previous categories) get three points.

4 RESULTS AND DISCUSSION

For a better organization of the systematic review, first, we disclose the analysis and discussion of document quality synthesis followed by how the categorization of variables and results were done and the results and discussion of each RQ.

4.1 Document Quality Synthesis

The mean quality score from all 79 documents considered for full-text was 6.22 points, with a standard deviation of 1.54 (Figure 1(b)). Documents with a score equal to or higher than seven are considered high quality. As such, 34 documents (43%) are considered high quality and provide an overall high-quality peer-review, sample count, and robust methodology. There is a limitation in the quality score field “document type,” because some conferences can have higher-quality documents than some journals. This limitation, however, is mitigated through the other two quality scores (methodology and sample). If the paper is of high quality, then it will still be considered high quality even if at a conference (1 point from conference plus 6 points from methodology and sample). Since several works are considered high quality (almost half), we conclude that the work seems to be well supported and in a mature phase. We also conclude that because of the quality research being done around realism in IVEs, both authors and journals/conferences seem to recognize this research field’s potential.

4.2 Categorization of Variables and Results

Due to the lack of existing taxonomies that can cover such a wide range of factors, it was decided to create one for this study. The taxonomy proposed allowed us to integrate every variable found throughout the studies (within the scope of the study) into categories. The categorization of the dependent variables (regarding the user experience) can be found in Table 3 and independent variables (from now on called Realism Factors, corresponding to objective realism) in Table 4. User Experience (Task Satisfaction) and User Performance (Effectiveness/Efficiency) factors are based

Table 3. Dependent Variable Categorization (Dependent Factors)

Dependent Categorization	Description
User Experience (Embodiment)	User evaluation of their self-avatar.
User Experience (Perceived Environment Realism)	Subjective perception of realism of the virtual environment.
User Experience (Task Satisfaction)	Subjective user satisfaction about the task.
User Experience (Virtual Agents)	User perception of other entities.
User Experience (Involvement)	How the user is involved (e.g., pleasure, stress, engagement, boredom).
User Experience (User preference)	Subjective preference of users.
User Experience (Presence)	User's feeling of Presence.
User Performance (Effectiveness/Efficiency)	The effectiveness and efficiency of the user when performing tasks.
User Behavior	How users behave in the experience.
Physiological Responses	User's physiological response (e.g., heart rate, skin conductance)

Table 4. Independent Variable Categorization (Realism Factors)

Independent Categorization	Description
IVE Content (Visual - Avatar)	Content of avatars (how they look and behave).
IVE Content (Visual - Environment)	Content of the virtual environment (e.g., texture, geometry, visual cue).
IVE Content (Audio)	Audio content (e.g., soundscapes, audio cues, footsteps).
IVE Content (Haptic)	Haptic content (e.g., presence of vibration, wind, passive haptics).
IVE Content (Olfactive)	Olfactory content (e.g., presence of scents).
IVE System (Audio)	How audio was synthesized and delivered (e.g., headphones, speaker array).
IVE System (Haptic)	How the haptic was synthesized and delivered (e.g., different haptic devices, synchronicity).
IVE System (Interaction)	Interaction between the user and virtual environment (e.g., locomotion, joysticks, gestures).
IVE System (Camera)	Virtual camera settings (e.g., field of view, filters, interpupillary distance).
IVE System (Lights)	Light rendering (e.g., ray casting, ray tracing, reflections).
IVE System (Physics)	World physics engine (e.g., object behavior, gravity).

on the satisfaction and effectiveness/efficiency metrics, respectively, from ISO/IEC 9126-4 recommended usability metrics [1].

Although we did our best to segment each independent and dependent variable in different Realism Factors and categories, in specific contexts, some variables can be ambiguous and may fit in more than one Realism Factor or category (in part the result of the absence of a proper taxonomy). In these situations, the researchers discuss where the variable would fit the best according to the study context. For example, changing the tangible device that the participant uses to interact in the IVE could both fit in IVE System (Interaction) as well as IVE Content (Haptic) or IVE Content (System). However, if the study is oriented to research interaction (e.g., studying different levels at which the user can provoke changes in the virtual environment) and not the haptic feedback effect (e.g., compare different haptic devices/modes while the extent to which the users can provoke changes in the virtual environment remains the same), then the study would be placed in the IVE System (Interaction). This way, we avoid increasing complexity by simultaneously having the same variable in different factors/categories. The Realism Factors are divided into two main categories:

- **IVE Content:** Related to the content of the virtual experience (e.g., self-avatar, virtual agents, audio cues, texture quality, mesh quality).
- **IVE System:** Related to the equipment capabilities to synthesize the virtual experience (e.g., type of haptic device, different audio setups, illumination models such as ray-tracing).

Both IVE Content and IVE System have sub-categories that were created to accommodate the studies' independent variables. This categorization enabled us to systematize the data better and find patterns in what was studied and yet have to be researched. During the analysis, the meaning of the variables and what they measured was taken into account when fitting them into a category.

Inside of each Realism Factor there are different levels of objective realism that are compared. Usually, authors state which level (or condition) is the most realistic. However, some authors do not

give this information; in these situations, the research team considered the most realistic condition the one that better mimics (or has a better potential of mimicking) real-life situations. Based on this methodology, we report that a given factor's Better/Worse level of realism had a Positive/Negative/Mixed/Neutral impact in a given dependent category. Note that Positive/Negative does not necessarily refer to higher or lower scores, as there are variables where higher scores could be interpreted as worse (and vice versa). Note that an increase or decrease of scores in a given dependent variable does not necessarily mean that a given Realism Factor directly affected that same variable but rather indirectly by influencing an unconsidered variable, which by its turn influenced the one being measured. The results had to be significantly different for the impact to be considered Positive or Negative. For qualitative studies, the results were based on the authors' conclusions.

Consider the following example: A study investigates how different levels of visual detail can influence the sense of presence, user memory recall, and cybersickness [36]. In this case, the visual detail is included in the Realism Factor of IVE Content (Visual - Environment), as it refers to a virtual environment's visual content. The sense of presence is included within the User Experience (Presence) category. The memory recall is considered a metric of User Performance (Effectiveness/Efficiency), and therefore it is included in this dependent category, while cybersickness is included in Physiological Responses. In the case that presence scores increases in the higher visual detail condition while memory recall and cybersickness are not influenced, we report it as follows: Better IVE Content (Visual - Environment) resulted in a Positive impact in User Experience (Presence) and a Neutral impact in User Performance (Effectiveness/Efficiency) and Physiological Responses.

There are also situations where authors do not isolate all variables, resulting in conditions that differ in more than one factor (e.g., Condition A has better lighting and sounds, whereas Condition B has worse lighting and no sounds). Therefore the independent variable cannot be categorized solely in one Realism Factor. In these cases, we report the following: Better/Worse IVE System (Camera) + IVE Content (Audio) resulted in a positive/neutral/negative/mixed/indeterminate result.

Another situation that can occur is when a particular Realism Factor has no impact (Neutral) by itself on the experience, but the interaction with other Realism Factors has. In these situations, we report it as follows: Better/Worse IVE System (Camera) x IVE Content (Audio) results in a positive/neutral/negative/mixed/indeterminate result. In this same situation, there is the possibility of impacting the experience when the level of objective realism of a giving factor is low and the level of objective realism of other factors is high. It would be erroneous to state that a better level of realism in both factors would result in a specific impact. In these cases, we report that IVE System (Camera) + IVE Content (Audio) results in a positive/neutral/negative/mixed/indeterminate result.

The Realism Factor's impact within the same dependent category on user's experience was also categorized into five possible outcomes:

- **Positive:** The Realism Factor showed an improvement in the user experience.
- **Negative:** The Realism Factor worsened the user experience.
- **Mixed:** The Realism Factor showed both a negative and positive effect.
- **Neutral:** The Realism Factor showed no effect.
- **Indeterminate:** The Realism Factor showed an effect, but it cannot be categorized as better or worse.

There is a difficulty in synthesizing the types of impact on physiological responses. For example, an increase in heart rate might be neither positive nor negative, just a manifestation of the human body to the stimuli received in the IVE. However, in some contexts, we may attribute an increase

of heart rate as positive or negative if the virtual experience depicted manages to replicate the same physiological reaction of a real analogous situation. In a hypothetical scenario where a user would experience a very stressful event (where a heart rate increase would be expected in a similar real-life situation), we consider it a positive impact if the heart rate increases. In conclusion, the type of impact registered at physiological responses will be highly context-dependent. Similarly to physiological responses, user behavior's also presents a challenge to synthesize. Participants have different backgrounds and life experiences and personalities (between other variables) that can lead them to behave differently to the same stimuli. However, some behaviors (such as proxemics [50]) can still be studied. In the case of User Behavior, the type of impact will be considered positive if the user reacts as expected (i.e., similarly to a real-world condition) as per the authors' conclusions. However, if the authors do not conclude whether user behavior was closer or not to reality, then our research team would analyze the context of the study and decide the impact type based on how we expected the user to behave under the same conditions in the real world.

4.3 Terms and Definitions (RQ1)

The first RQ is centered on understanding which terms authors used (regarding realism and similar terms) and how they define them. Considering all documents that fit the criteria of the systematic review, there is a clear use of the terms realism (50 documents –53%) and fidelity (39 documents –42%) over credibility (2 documents –2%), coherence (1 document –1%), congruent (1%), and believable (1 document –1%).

Note that some authors used more than one term. The terms that the authors are defining are the ones considered for the analysis. If no definition is given, then all terms used are considered (e.g., a document may contain realism and fidelity, but no definitions, the document is considered for both realism and fidelity count). Considering only documents that contain definitions (34 documents), we have: 18 documents (53%) defining fidelity, 13 documents (38%) defining realism, 2 documents (6%) defining credibility, and 1 document (3%) defining coherence. All these data and definitions can be visualized in the supplementary material.

Almost all authors presented their specific views on the subject, resulting in different approaches. Some authors only defined one dimension of the term they used, others divided it into several dimensions, and some considered only one dimension of the term (e.g., visual realism [109] or visual fidelity [31]) and further divided it into sub-dimensions.

Regarding authors who use the term realism, one took on a more objective approach [54]—“Realism is the Immersion, which is the ‘the objective level of sensory fidelity produced by a IVE System’”; another a more subjective approach [18]: “closeness between the viewer’s perception of the virtual environment and an identical real one”; and others considered both aspects [88] when dividing it into sub-dimensions. Here, we can verify the importance of dividing realism in objective and subjective to avoid misinterpretations of the same across multiple studies. The objective approach by Laha et al. [54] and Perroud et al. [88], “Realistic looking” and “Realistic construction of the virtual world” and “Physiologic realism” constructs, are closely related to our notion of objective realism. However, we do distinguish in which immersion is not realism but rather the capabilities of the system within which a realistic virtual experience could take place. Conti et al. [18] and Perroud et al. [88] “Psychological realism” approaches are similar to our notion of subjective realism. However, on the contrary to Conti et al. [18], we do not assume if the virtual experience tries to replicate the real world, as users might consider a virtual experience realistic even if it depicts events impossible to happen in the real world as long as they are coherent [102, 103] within the rules of the environment.

Regarding fidelity, the majority of authors define it objectively to how well a virtual experience replicates a real analogous one [5, 19, 24, 31, 36, 38, 58, 72, 89, 101, 131], while one author takes

on a more subjective approach [68] by implying perceptual fidelity as “not necessarily equivalent to physical simulation.” We can conclude that fidelity is in fact the same as our objective realism definition, independent of users, except to Mania et al. [68] “perceptual fidelity,” which fits the subjective realism where users might consider a virtual experience real even if it does not properly simulate the real world.

Overall, there seems to be a mixed-use of the terms realism and fidelity, mainly the fact that some authors refer to realism as either objective or subjective or the same as other terms (which could also be subjective or objective for other authors), whereas some further divide realism in subjective and objective dimensions. For example, Slater et al. [109] divided visual realism into two components, geometric realism and illumination realism, where the latter one is defined as “the fidelity of the lighting model.” Rogers et al. [95] also seem to consider fidelity and realism as synonyms “fidelity (also naturalism, or realism)” using them interchangeably by defining display fidelity as “sensory realism, referring mainly to auditory and visual qualities” and interaction fidelity as “action realism, i.e., the degree of exactness with which user actions in VR resemble real-world actions in terms of biomechanical similarity, input, and control.”

Very few authors used and defined credibility. Boucaud et al. [10] defined it as, “We understand credibility here as the degree to which the participant feels the agent behaved itself in an adequate human-like way”, and another used a dictionary definition [30]: “defined by Merriam-Webster is the quality of inspiring belief.” Both definitions are equivalent to coherence [102], which are, by their turn, similar to our notion of subjective realism.

Only one author considered coherence defining it as [102]— “the set of objectively reasonable circumstances that can be demonstrated by the scenario without introducing objectively unreasonable circumstances”—while also stating it is different from fidelity—“Coherence as a construct is related to, but distinguishable from, fidelity,” which was already discussed in the introduction of this article.

We must note that we did not find works that based their definitions on terminology. Most of the authors included and defined certain dimensions/components that were of interest in the context of their studies. Therefore, they do not show us the whole picture. Although this systematic review only considered IVEs under the constraints of RQ2 criteria, each author seems to provide us different definitions, where although some similarities, may confuse readers. To add to this possible confusion, some authors do not provide definitions. These situations could be compared to the confounding use of immersion [124] and presence [106], where authors were using one’s definition instead of the other, which could originate confusion, mainly when no definitions are given.

Only 27 documents (34%) presented the authors’ definition of the term used. This suggests that the authors considered the definition as general culture. In these cases, the term’s definition should be considered as the one present in the dictionary, which sometimes might not be the definition authors considered for their paper.

Even though several terms with different definitions were used, when considering the ones that address the subjective side, they all come together regarding “how the IVE meet the users’ expectations of what is real within the internal rules of the virtual scene.” Different people have different realities due to cultural background, physiological, and/or other differences. For example, a person unfamiliar with airplanes might find a hypothetical flight simulator as real. However, an aircraft pilot may find the same flight simulator as unreal. This would be due to differences in both subjects’ life experiences. There are also mental illnesses with psychoses that can shift the perception of reality. There are also situations where unrealistic experiences (such as the Stanley Milgram experiment [77] replicated in VR [107]) might still be considered real despite being known that they are not. This is known as PsI [105], which could also be considered the same as reality judgment [103], coherence, and credibility.

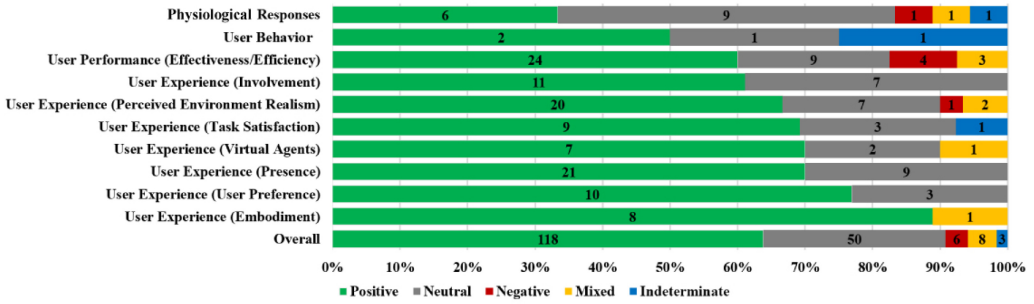


Fig. 2. Impact of increasing objective realism on dependent categories. The number of occurrences is shown inside the chart bars. Note that this graph only represents the number of studies supporting each impact and should be interpreted as such.

Then, we have the objective side of the definitions, which commonly means “how a system was able to replicate the real-world conditions,” independent of the users’ perception. Due to these circumstances, a unified definition of realism is important.

4.4 The Impact of Different Realism Factors in IVEs (RQ2)

After listing every variable from data extraction, the research team found patterns in both independent and dependent variables, allowing us to categorize them into several groups. Results from three studies could not be categorized (check supplementary material) and were excluded from RQ2 statistics, resulting in 76 documents being considered. A detailed description of the impact of each Realism Factor per document can be found in the supplementary materials. Overall, the impact of realism was positive (Figure 2): 64% positive, 27% neutral, 3% negative, 4% mixed results, and 2% were indeterminate.

4.4.1 Dependent Categories. In Figure 2, we can verify how objective realism impacted each dependent category and how many occurrences support it. The most positively impacted dependent categories were embodiment (88.9%), followed closely by user preference (76.9%), presence (70%), virtual agents (70%), task satisfaction (69.2%), perceived environment realism (67.7%), involvement (61.1%), user performance (60%), user behavior (50%), and physiological responses (33%).

User performance: The impact of objective realism in user performance is widely researched, featuring 40 occurrences. User performance was shown to be influenced positively by an increase in objective realism in 60% of the cases, neutral in 22.5%, negative in 10%, mixed in 7.5%. However small, there were also negative impacts found. There is a clear search for how well users perform under different levels of objective realism in IVE. The rise of virtual simulators could explain this number. Because IVEs have the potential to recreate and/or replicate real scenarios, they are good candidates for simulators in the most different fields. Thus, user performance is an important metric to evaluate if a given simulator can reproduce the same conditions as reality. Even though almost all of the documents considered in this study were of general purpose, several of those studies produced insightful knowledge that could be applied to IVE in different fields.

Considering one of the consequences of the sense of presence, similar user behavior to a real analogous situation, this metric would make sense to be used together with user performance to provide a better discussion about why differences in user performance may have happened. We should note that better user performance in virtual environments should not always translate to better performance in real life. If the difficulty of performing a task in a virtual space is easier than performing in real life, then IVEs in the context of training (e.g., sports, medicine, military) might

not be efficient at their purpose. Possible reasons for such difficulty differences can be anxiety [113], stimuli cues meant to help the users [19], visual complexity [70], the object/tool representation fidelity and interaction [25, 78], absence/limitation of physics [5], or even its incorrect use [102]. Note that in specific situations, lower user performance caused by higher objective realism might be a good indicator that the virtual environment is replicating the real-world conditions and not facilitating the user by having a simpler and lesser realistic experience. Nevertheless, in this work, we consider a negative impact on user performance if the users perform worse, as per the authors' conclusions, disregarding if it is the desired result or not for that specific situation.

Presence: Next follows the presence dependent category with 30 registered. This metric is widely popular in IVEs and was expected to be one of the most used. Presence was positively influenced by objective realism in 70% of the cases, whereas in 30.0% there was no impact (neutral). No negative, mixed, and indeterminate impacts were found. This led us to conclude that objective realism can be an effective way to increase presence. Although, in some contexts, the efforts to create more-realistic experiences might not influence the user's feeling of presence.

Perceived Environment Realism: Right next to presence comes the perceived environment realism with 30 occurrences. Both are usually connected in their interpretation [3, 103], and some presence questionnaires even have subscales/questions to evaluate experienced realism (**Igroup Presence questionnaire (IPQ)** [100] and **Presence Questionnaire (PQ)** [125]). The increase of objective realism was proven to be perceptible by users (perceived environment realism) in 66.7% of the cases. The other 23.3% were neutral, 3.3% negative, and 6.7% mixed. It is worth mentioning that one can feel present in an environment with low objective realism. The close connection between both categories might justify why they have roughly the same number of occurrences.

Involvement: There were 18 occurrences in the involvement-dependent category, suggesting that researchers aim to understand how realism affects how much and how well participants are involved in the experience. Involvement includes many different metrics (e.g., intuitiveness, emotions, comfort, pleasure, or engagement) but all related to the user's involvement during the experience. Involvement was positively impacted by realism in 61.1% of the cases and neutral in 38.9%. Also, no negative, mixed, or indeterminate impacts were found, suggesting that an increase of realism is unlikely to decrease the user's involvement. This dependent category could be further broken down into subcategories due to various variables, but such detail is out of this article's scope.

Physiological Responses: We found 18 occurrences regarding the impact of realism on physiological responses. One would expect that physiologic responses would have a higher percentage of positive impacts (33.3%). However, differently from other categories, the largest part of the impact was considered neutral (50%). The rest of the impact is divided into 5.6% as negative, 5.6% as mixed, and 5.6% as indeterminate. The large percentage of neutral results does not directly mean that physiological response was not affected entirely, but could also mean that the physiological responses researchers were measuring were not the ones being affected; or if affected, the increase of objective realism was not impactful enough to change physiological responses significantly.

Physiological responses are considered objective and can help researchers better understand how participants react to different levels of realism that otherwise could not be picked up in other instruments. We speculate that physiological measurements may not be easy to use, as they require proper equipment and analysis and may be intrusive during the IVE. The fact that users can have the liberty to navigate and interact in the environment, as they would in reality, might present an obstacle for specific physiological measurements, such as electrocardiograms, introducing noise in captured data. Due to the HMD apparatus covering the head, an electroencephalogram, for example, might prove difficult to properly setup. However, it may present a powerful instrument to evaluate realism, as it is unbiased. We must note that, although the physiological responses do not depend on the user's opinion, they may still be prone to be influenced by other variables.

For example, in a stressful environment, a user with past experience in the field might present a different physiological response than a person experiencing the depicted event for the first time. Although, this situation might be mitigated by using a larger and diverse sample. However, it is still important to keep track of possible confounding variables to guarantee valid results.

Overall, physiological measurements consisted of heart rate, skin conductance, and electrocardiograms. Other physiological responses that were not measured by physiology equipment consisted of simulator sickness symptoms and motion sway. We argue that simulator sickness (cybersickness) is particularly important to be evaluated in specific contexts. It consists of motion sickness that evokes symptoms such as nausea, headaches, dizziness, eye strain, sweating, disorientation, or vomiting [21, 55]. Several theories exist regarding the symptoms' origin, but the most accepted is the sensory conflict theory. When a discrepancy happens between the visual and vestibular systems, cybersickness symptoms arise. Positional tracking error, lag, or flickering (the higher the field of view, the more the flickering is perceived, as peripheral vision is more sensitive to it) may contribute to cybersickness [21, 55]. Although the advancements of technology already mitigate these variables, other factors may still be at play. Individual factors such as age, gender, and illness may influence the sensitivity to cybersickness.

Regarding how the virtual experience was done, the user's position (sited/standing) and the amount of control they have in the experience may affect cybersickness. Therefore, if overlooked, cybersickness could interfere negatively with the participant's experience, which could compromise the results, depending on the severity of the symptoms. The results revealed that authors are very consistent in how cybersickness is measured, always using questionnaires, particularly the **Simulator Sickness Questionnaire (SSQ)** [46], although, from 1993, it is still used in recent studies.

User Preference: A total of 13 occurrences regarding the user's preferences were found. Users notably preferred more objective realism, with user preference scoring positively in 76.9% of cases, 23.1% as neutral. No negative, mixed, or indeterminate results were found. We suggest that user preferences should be taken into account as control variables to better understand the results and provide better discussions.

Task Satisfaction: We found 13 occurrences regarding the users' task satisfaction. Results indicated that it was positively impacted by realism in 69.2% of the cases, neutral in 23.1%, and indeterminate in 7.7%. The lack of negative and mixed results suggests that task satisfaction is likely to not decrease due to higher objective realism. This can be an interesting metric, as users might consider the task easier and comfortable to perform with different levels of objective realism. Consider that there was a mismatch where users wrongly thought that they performed well above their actual performance. Such could indicate a possible excess of confidence (users might feel at ease and confident they performed well), low understanding of the tasks (wrongly thinking the task is being performed right), or lack of feedback from the application. If the application aims to train users to execute tasks in the real world, then a higher objective realism might be desirable [80, 86] even if the task satisfaction is lower because it properly replicates the real-world conditions. However, if the application objective is to lead users to perform tasks specifically in the IVE, then it might be desirable to increase their task satisfaction ratings.

Virtual Agents: We found 10 occurrences regarding the impact of realism on how users rate virtual agents. Eventually, a virtual experience will need other entities, either controlled by humans or autonomous, to replicate reality properly. Social interactions are an intrinsic aspect of human life. As such, how realism impacts our perception of virtual agents is an important element to consider. Virtual agents were usually better rated when realism was increased, gathering 60% of positive impacts, 20% neutral, and 10% mixed. Despite the one mixed impact, no negative impact was found.

Embodiment: A total of nine occurrences were found considering the embodiment (substituting the real body by a virtual one [103]), lower than what was expected. It was the most benefited dependent category from an increase in objective realism, with 88.9% of the cases being a positive impact and only 11.1% having mixed results.

One of the key challenges of IVEs is to trick the human brain into considering the self-avatar as their real body [103]. If such an illusion is successful, then users will behave differently physically and cognitively than if there was no illusion. This is called the “Proteus Effect” [126]. This mental illusion can even be traced back to studies not using IVE, such as the rubber-hand experience [9]. This experiment proved that participants behave as if the rubber hand was indeed their hand, even though they knew it was not, in the face of a threat. Due to the plasticity of the human brain in accepting different avatars as its own [110] and having diverse results on human psychology [126], we expected a higher number of studies documenting the realism role in this illusion. We suspect that the difficulty to properly track the human body, as it requires complex tracking systems, can be a barrier for several studies. Also, self-avatars are mainly used in immersive VR and MR [110], as in AR the real body can be seen at all times unless the virtual body is overlaid on top of the real body.

User Behavior: With the least occurrences (four), we have user behavior. This was unexpected, as user behavior can be a strong objective indicative (if objective evaluations are used) that users consider the environment as real and therefore behave accordingly. However, to properly evaluate user behavior, it is necessary to determine the ground truth behavior. An example of this would be using methodologies to record facial expressions and body movements under real-world conditions and compare them to those recorded under virtual conditions. Although, this requires a real-world variant to be possible. When such is undoable, we would need to rely on the literature (how users should behave under specific situations) or/and use conditions as close as possible to the one being replicated in an IVE as the ground truth, or even resort to machine learning.

Therefore, we speculate that one of the reasons for a lower study count could be the lack of viable methodologies to evaluate user behavior properly. From the identified studies, only two studies evaluated user behavior, and both used objective measurements but without a real-world baseline. The study conducted by Habibnezhad et al. [34] researched how users behaved in IVE when exposed to heights vs. ground level with and without virtual legs. Their approach to measure behavior was through tracking, measuring the user’s stride. The other study was done by Krum et al. [50] where user proxemic behavior was studied by registering the distance users kept from a virtual agent. Overall, user behavior was not negatively impacted in any of the studies, resulting in 50% of positive impact, 25% neutral, and 25% indeterminate. However, we should also note that the sample was too small (four occurrences) to properly present any conclusions. Therefore, we suggest that more work should be directed at studying how users behave when confronted with different levels of objective realism in IVE.

4.4.2 Individual Realism Factor Impact. This section discusses the impact of each Realism Factor individually. An overall view of the Realism Factors impact can be visualized in Figures 3 and 4.

IVE Content (Visual - Avatar) impact was studied in every dependent category except involvement. The most studied was User Performance, accounting for 11 occurrences, more than in any other Realism Factor, with the majority of the impact being positive (58.3%).

Some examples of what was evaluated in this Realism Factor were: virtual agents behavioral realism [52], communicative realism through smile [33], body parts such as legs [34] or hands [91], avatar texture fidelity [117], anthropometric fidelity [120], crowd behavior [51, 94], hand proportion [87], different head/arms/forearm/hands configurations [98].

From the results, we can see a search for understanding how avatars (which could be in the form of self-avatar or virtual agents) influence user performance. This is an important topic of research when considering training simulators. Several professions such as firefighters [90], surgeons [15],

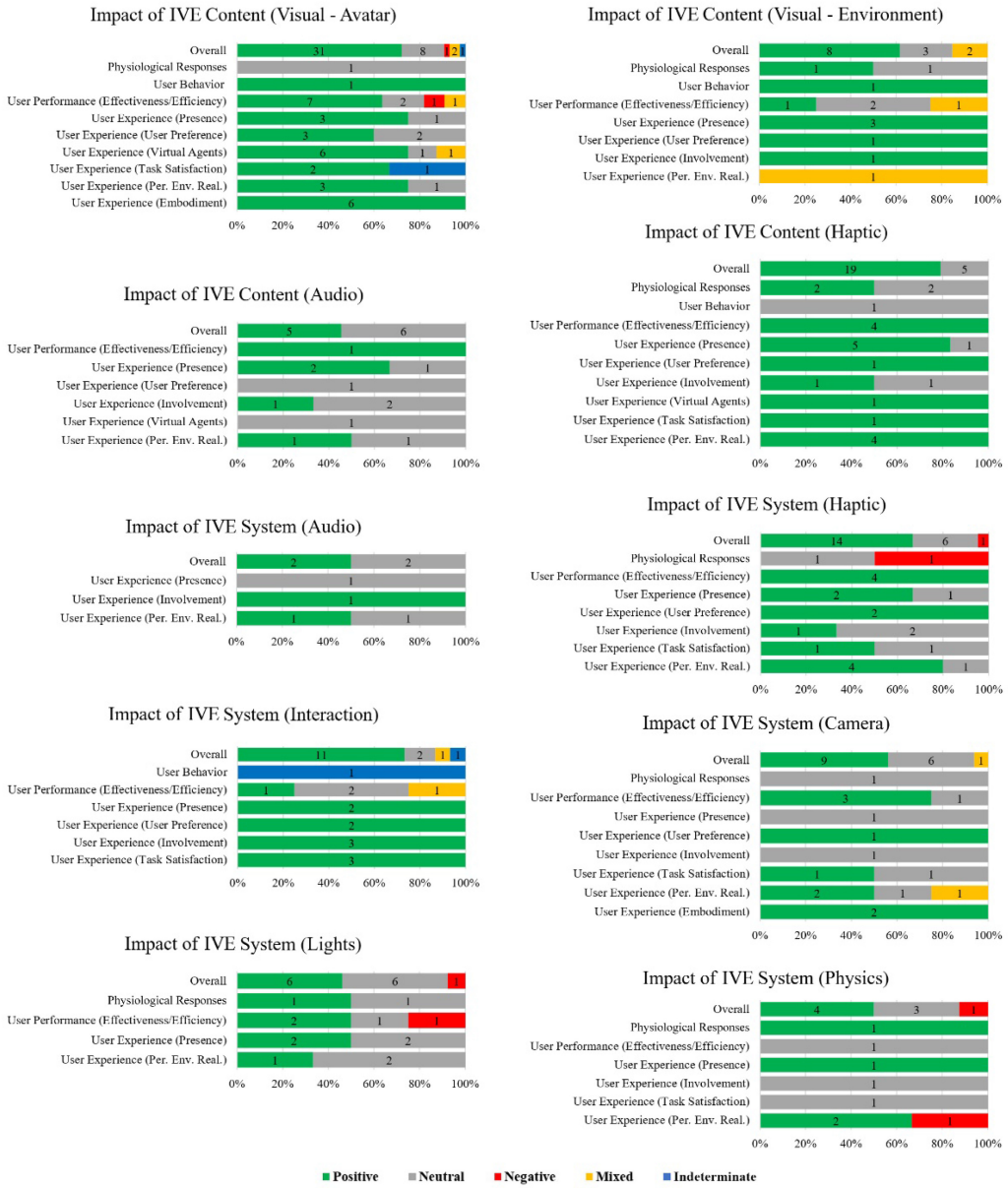


Fig. 3. The impact of increasing each Realism Factor individually on the user's virtual experience. The number of occurrences is shown inside the chart bars. Note that this graph only represents the number of studies supporting each impact and should be interpreted as such.

or military [116] have tasks that require teamwork. Thus, a self-avatar might prove useful to give users a sense of embodiment as well as virtual agents to which they must collaborate/interact [57, 128].

As expected, the following two most researched categories are virtual agents (eight occurrences) and embodiment (six occurrences). As researchers are investigating avatars (self-avatars and

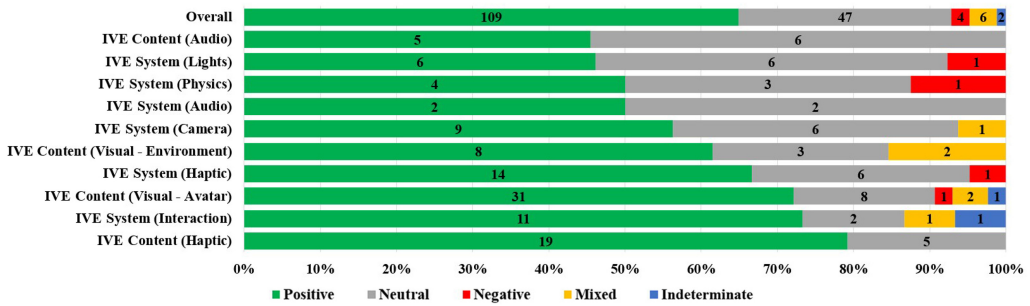


Fig. 4. The overall impact of increasing each Realism Factor. Note that this graph only represents the number of studies supporting each impact and should be interpreted as such.

virtual agents), it is expected that they also apply metrics to evaluate them. Virtual agents' scores usually improved (75% of the cases) with higher objective avatar realism. All of the six occurrences of embodiment showed a positive impact when increasing the objective avatar's realism. Overall, IVE Content (Visual - Avatar) positively influenced 72.1% of studied cases, 18.6% were neutral results, 2.3% were negative, 4.7% mixed, and 2.3% indeterminate.

Curiously, there was no mention of the Uncanny Valley. The negative effect found was on user performance, which resulted from an increase in difficulty to perform a task (as discussed in 4.4.6).

IVE Content (Visual - Environment) impact was researched in seven categories. No studies were found regarding its impact on Task Satisfaction and Virtual Agents.

Variables studied in this Realism Factor consisted in: visual complexity (amount of detail, clutter, and objects on scene) [92], texture quality and **levels of detail (LOD)** [36], texture repetition [11], and polygon/triangle count [42, 84, 118].

Involvement research was expected to be more extensive, as one would expect that an increase of environmental realism should be directly linked to a better involvement. Similarly to IVE Content (Visual - Avatar), the most researched impact of the factor was on User Performance. Overall, IVE Content (Visual - Environment) positively impacted 61.5% of the studied cases, 23.1% were neutral, and 15.4% were mixed results.

IVE Content (Audio) impact was investigated in six categories. Non-researched categories were: Embodiment, Task Satisfaction, User Behavior, and Physiological Responses. Some variables studied in the context of this Realism Factor were: task-appropriate sounds [19], steps and sound-scape [47], ambient noises [95], audio directivity [123], and **Head-Related Transfer Function (HRTF)** [44].

This Realism Factor presents only 11 total occurrences with a maximum of 3 occurrences on presence and involvement. It would be interesting to focus more research on how audio content realism could improve user behavior. We suggest this particular research due to the premise that, in specific contexts, more-realistic audio cues could change the users' behavior in dangerous situations. Monteiro et al. [80] studied critical stimuli in decision-making in VR training and concluded that trainees should experience the same critical stimuli in VR as they would experience in the real-world scenario for a VR simulator to be a valid alternative to real-world training. For example, when crossing the road, the presence of a vehicle sound approaching the user could influence how the user would behave. Likewise, in a mechanic simulator, heavy machinery's audio fidelity could modify how users behave around them due to safety reasons. We also suggest more work on user performance in IVEs, as some tasks might require faithful audio to be properly performed. For example, a simulation where mechanics have to diagnose engine problems through their sound, a medic trying to hear a patient's heart through a stethoscope, or firefighters listening for possible

gas leaks. Please note that this does not mean that realistic audio is not already used in IVEs, only that its impact is not thoroughly investigated.

Curiously, none of the studies explored the audio occlusion, Doppler Effect, or reverberation. Although two studies explored HRFT [44] and speech directivity [123], overall, little attention was given to sound propagation. Audio occlusion, for example, could prove useful for firefighters to locate a gas leak. Because the frequency of occluded audio would be lower by a wall or door, firefighters might use such audio cues to locate the leak. In the architecture field, a hypothetical simulator where users could build infrastructures (such as theaters) would also benefit from a proper simulation of sound propagation. Users could experiment with several objects and materials and modify the overall 3D shape to preview how sound would feel. Also, no studies were found researching audio-compression quality, such as bit-rate or sample rate. Instead, studies were more focused on the presence/absence of audio cues. Overall, it presented a positive impact in 45.5% of the studied cases with the rest being neutral (54.5%), making it the only Realism Factor where a positive impact was the less prominent.

IVE Content (Haptic) was investigated in all dependent categories, except Embodiment. From the categories researched, the one that featured more occurrences was presence (six). Gonçalves et al. [30] placed a simple wooden board in the ground to serve as a haptic stimulus of going up a stair step. Joyce and Robinson [45] used a blank panel so users would have haptic feedback when touching a virtual button. The studies above show relatively simple yet effective methods of increasing objective realism through passive haptics with good positive results overall. However, some problems could arise due to safety reasons, such as users not expecting a stair degree and possibly tripping and falling [30], or properly synchronize the real object with the virtual one [25, 27]. Some example of variables studied in active haptics were: wind [30, 93] and thermal feedback [93], being touched by avatar [50], presence of vibration [19, 63], and presence of force-feedback [127].

Notably, no negative, mixed or indeterminate results were found. Overall, IVE Content (Haptic) impact was positive on 79.2% of the studied cases and neutral in 20.8%.

IVE System (Audio) was investigated in three categories. Non-researched categories were: Embodiment, Task Satisfaction, Virtual Agents, User Preference, User Performance, User Behavior, and Physiological Responses. No negative, mixed, and indeterminate results were found. Half of the results were positive and half were neutral, with the most researched dependent category (Perceived Environment Realism) only having two occurrences. However, more research is needed in this Realism Factor to corroborate existing results as only two documents researched this topic. One of the studies compared types of acoustic environment (FOA-static binaural, FOA-tracked binaural, FOA-2D octagonal speaker array) [39] and the other explored the use of noise-cancelling headphones [47]. The first study concluded that FOA-2D octagonal array outperformed the rest of the acoustic environment. The second concluded that noise-cancelling did not affect presence, involvement, and subjective realism but reduced user's distraction.

IVE System (Haptic) was investigated in six categories. Non-researched categories were: Embodiment, Virtual Agents, User Behavior. One negative impact was found on the Physiologic Response. No mixed or indeterminate results were registered. Examples of variables studied in this Realism Factor were: tangible fidelity (different ways to represent the same stimulus) [85], different prototypes to synthesize texture [4], using a motion platform vs. a real vehicle [127], and using a tracked real putter instead of a controller in a golf-related task [25]. Overall, the IVE System (Haptic) impact was positive on 66.7% of the studied cases, neutral in 28.6%, and negative in 4.8%.

IVE System (Interaction) was investigated in six categories. Non-researched categories were: Embodiment, Perceived Environment Realism, Virtual Agents, and Physiological Responses. Some examples of variables studied under this Realism Factor were: locomotion technique [50], different

grasping object techniques [119], and ability to interact with virtual object [78, 131]. More studies investigating interaction were expected. One of the essential points of being in an IVE is the ability to interact with it to provoke changes and received feedback from those changes [65, 130]. Although few documents were found that studied different levels of interaction objective realism, it does not mean other studies were not using interaction at all. Overall, the IVE System (Interaction) impact was positive on 73.3% of the studied cases, neutral in 13.3%, mixed in 6.7%, and indeterminate in 6.7%.

IVE System (Camera) was investigated in eight categories. Non-researched categories were: Virtual Agents and User Behavior. There were no negative and indeterminate impacts. However, a mixed result was found. Examples of variables studied under this Realism Factor were: field of view [92] and field of regard [54], intra-camera distance [18], depth of field [18], foveated rendering with different foveal regions sizes [122], ocular parallax [49], processed video feedback from the real world [121], and first-person and third-person perspectives [71, 73]. Overall, IVE System (Camera) impact was positive on 56.3% of the studied cases, neutral in 37.5%, and indeterminate in 6.3%.

IVE System (Lights) was investigated in four categories (Physiological Responses, User Performance, Presence, and Perceived Environment Realism). Non-researched categories were: Embodiment, Task Satisfaction, Virtual Agents, Involvement, User Preference, and User Behavior. There were no mixed or indeterminate results. Some examples of variables studied in the Realism Factor were: radiosity [67, 69, 70, 84], raytracing and raycasting [109], shading models (unlit shader, Lambert diffuse shader, flat shader) [68–70, 118, 129], and **high dynamic range (HDR)** [101]. Overall, the IVE System (Lights) impact was positive on 46.2% of the studied cases, neutral in 46.2%, and negative in 7.7%.

IVE System (Physics) was investigated in six categories (Physiological Responses, User Performance, Presence, Involvement, Task Satisfaction, and Perceived Environment Realism). Non-researched categories were: Embodiment, Virtual Agents, User Preference, and User Behavior. There was one negative impact but no mixed or indeterminate results. Examples of variables studied in this Realism Factor were: object physics coherency [102], gravity [5], and simulated collisions [8]. We expected more studies in this regard. Every day, we experience physics laws, which led us to get used to how objects and other elements should behave. This is important for certain IVE, such as soccer players training their kicks or golf players training their putt. Both examples are widely sensitive to the physics engine's faithfulness in replicating real-world conditions so users can properly transfer their skills to real-world situations. For example, it is not natural to launch an object and see it floating in the air, ignoring earth's gravity. But, like any other Realism Factor, it is all context-dependent. If one is led to believe it is in space, then lack of gravity is expected and justified in the IVE, as the same would happen in the same circumstances in reality. If one is led to believe it is in space but with artificial gravity (e.g., centrifugal force), then experiencing gravity in a space simulated environment would be objectively realistic. Overall, the impact of IVE System (Physics) was positive on 50% of the studied cases, neutral in 37.5%, and negative in 12.5%. We must note that physics was one of the less-studied Realism Factors, with only three documents researching it. Curiously, presence was researched in every Realism Factor (except in some Realism Factor combinations), which indicates it as a very popular metric in IVE studies.

4.4.3 Realism Factors Combination Impact. Regarding the impact of the combinations of Realism Factors, each combination only has one study to support it. Also, these combinations do not allow us to understand if the impact was due to the combination, or only one of the Realism Factors. Overall, nine documents, with a total of 17 occurrences, explored combinations of these Realism Factors: 52.94% of the results were positive, 17.65% neutral, 11.75% negative, 11.75% mixed, and 5.88% were indeterminate.

4.4.4 Impact of Realism on User's Perceived Realism. We have analyzed how an increase of objective realism would impact the user experience, but there is yet another unexplored perspective: "What impacts the subjective realism?" The dependent category most capable of providing an answer is the Perceived Environment Realism factor, as this factor agglomerates variables addressing how the users evaluate the overall realism of the virtual environment. All independent factors (except IVE System (Interaction)) addressed this dependent category. Overall, from the 64% positive results on perceived environment realism, 22.22% were from IVE Content (Haptic), 22.22% were from IVE System (Haptic), 16.67% were from IVE Content (Visual-Avatar), 11.11% were from IVE System (Camera), 11.11% were from IVE System (Haptic), 5.56% were from IVE Content (Audio), 5.56% were from IVE System (Audio), and 5.56% were from IVE System (Lights). As visualized in Figure 3, IVE Content (Haptic) and IVE System (Haptic) had the most positive impact on the user's subjective environmental realism with a positive impact of 100% and 80%, respectively. Such leads us to suggest the use of haptic stimuli to improve the user's subjective realism. IVE Content (Visual-Avatar) follows as the next factor that registered the most positive impacts, indicating that it can lead users to rate the overall environment as more real.

4.4.5 Unsynthesized Studies. Three studies results were not possible to synthesize like other documents due to their methodologies [10, 88, 111] and therefore were not included in the RQ2 statistics. The first study, conducted by Slater et al. [111], explored how illumination type, field of view, display type (simulated powerwall or HMD), and extent of self-representation would affect PI and PsI. Participants were divided into two groups, one for PI and another for PsI. Participants started the experiment in the highest settings of the immersive system. In the following five trials, the participant had to change the settings (from low to high) for each factor until they felt the same level of PI or PsI they felt in the highest settings. The results indicated that the PI group tended to choose a wide field of view and HMD first and then a proper self-representation that moved as they did. The PsI groups focused more on attaining a higher level of illumination realism but still attributed importance to the self-representation like the PI group.

Perroud et al. [88] proposed a new scoring system to evaluate the realism of an IVE by objectively quantifying the visual perception characteristics. The research items were divided into vision cues (contrast and luminosity, frames per second, number of different colors achievable, field of view, monoscopic and stereoscopic acuities) and Immersion cues (latency, field of regard, stereoscopy, tracking, uniformity, and camera convergence). To compare the authors' theoretical model to scores given subjectively by participants, a user study was conducted. Participants performed an 8 min drive without any other task besides looking around while driving. In the end, they were asked to rate the different criteria of the authors' model by comparing their virtual experience with real life. Authors conclude that measured frames per second fit well the theoretical value, monoscopic acuity and color values are far greater than the theoretical ones. Field of view values were higher but close to the theoretical ones, while the field of regard theoretical values were very lower.

Boucaud et al. [10] studied the credibility of Virtual Agents when conveying emotions to users as well as when users are tasked to express emotions to virtual agents. Three emotions were considered: anger, sadness, and sympathy. Participants were equipped with a haptic sleeve in their arm so they could feel the virtual agent's touch. Results showed a difference in the credibility, as virtual agents appeared more credible when they touched the participants to express anger and sympathy and much less credible when it reacted to being touched or when they tried to express sadness. However, expressing anger, sadness, and sympathy to the virtual agent, and vice versa, are situations equally realistic that can happen in real life. Even so, some conditions were still perceived as more credible than others, which could be due to unknown factors or study limitations

such as participants not considering the haptic sleeve vibrations appropriate to simulate touch or that the virtual agent programmed reaction was not made credible enough. Therefore, we cannot state that a higher/lower realism resulted in a positive/neutral/negative/mixed/indeterminate impact, because the conditions were equally real for the depicted context.

4.4.6 Discussing the Negative Impact. Only a few studies registered a negative impact. It is important to understand why these studies, in particular, reported such results. We verified that the negative impact is only present in Physiological Responses (provoked by IVE System (Haptic)), User Performance (caused by IVE System (Lights) + IVE Content (Visual-Environment), IVE System (Lights), IVE System (Camera) + IVE Content (Visual-Environment) and IVE Content (Visual-Avatar)), and Perceived Environment Realism (caused by IVE System (Physics)).

Regarding Physiological responses, there was only one document that reported a negative impact (of IVE System (Haptic)) [127], which was in the form of simulator sickness. In this study, users felt more simulator sickness when going from a motion platform to an actual vehicle in a virtual driving simulation using an HMD. However, the participants were not in control of the real vehicle. Although the authors state that the visual stimuli were in sync with the real vehicle movement, the lack of control may be one of the causes that increased simulation sickness [96].

Regarding User Performance, we found a total of four documents reporting a negative impact of realism. One of the works was from Petti et al. [89], where higher IVE System (Lights) + IVE Content (Visual-Environment) realism resulted in worse User Performance. Authors speculate that the higher level of detail distracted participants, leading them to spend more time in the virtual environment, resulting in increased motion sickness (several display issues were reported by the authors, with emphasis on high-fidelity condition), which reduced their scores and lowered their performance. If we consider this justification, then objective realism was not the direct cause of the negative impact, but rather the hardware limitations (as display issues were reported) provoking cybersickness symptoms the longer users spent in the IVE. Such led us to suggest that an increase of objective realism should only be done if the equipment allows it, or it may cause the exact opposite results from those expected.

Another work was from Mania et al. [67], where a negative impact of IVE System (Lights) was found over User Performance, specifically on memory performance. Although memory performance itself was not affected by viewing condition (flat-shaded vs. radiosity), confidence scores (the certainty of users had in their responses) were lower in the most realistic condition. Authors suggest that this is due to the less-realistic environment being more distinctive than the real one. The less realistic the environment is, the more distinctive it is when judging it against reality. Authors link this difference in distinctiveness to psychological research, which has shown that distinctive experiences led to more awareness states related to ‘remembering.’ The research team has two views over this: In a way, the increase in realism diminished user performance. However, and by following the author’s justification, such happened due to indirect consequences of realism and/or due to methodology limitations. By looking through a different scope, purposely lowering realism to increase performance can be beneficial if there is no intent to transmit the experience to a real-life situation. Therefore, if the objective is to translate the experience/knowledge to real life, then one should increase realism to avoid a disconnection between what users experience virtually and what reality is. For example, a student conducting an online test in immersive VR could benefit from a less-realistic scenario, possibly making it less distracting, increasing user performance. Another example would be a virtual training situation where firefighters would have to remember a building’s layout. By using a less-realistic environment, their performance might be higher. Albeit, there could be a discrepancy when going to a real-life location, as the complexity of the environment would be much higher than what they experienced in simulation.

The results from Stinson et al.'s [113] study demonstrated a negative impact of increased IVE System (Camera) + IVE Content (Visual - Environment) realism on User Performance. The authors investigated how anxiety triggers, the field of regard, and simulation fidelity would impact the user's performance (save percentage) and task satisfaction as a goalkeeper in a football free-kick simulation. Simulation fidelity had two levels, low (only the field, net, kicker, and ball) and high (the addition of other characters, large stadium with a crowd, more sounds, better animations), which represented a more complex and closer experience to reality. The results indicated that the save percentage in the lowest field of regard condition decreased with higher simulation fidelity. Authors suggest that peripheral awareness can be helpful only if it is not too distracting. Therefore, the higher peripheral vision allowed by the higher field of regard could increase the potential for distraction in more complex audiovisual simulations (such as the one portrayed in the high simulation fidelity condition). This distraction could have affected user's performance. However, we should note that it should be preferable to consider more-realistic environments for training, even if users perform worse than less-realistic environments, because a more-realistic experience is closer to real-life conditions. Users would then be better prepared to transfer that training into the real world (where all the stressful and distractive factors such as crowd, teammates, opponents, and others are present) [80, 86].

The last study that reported a negative effect of higher objective realism (IVE Content (Visual-Avatar)) on User Performance was conducted by Kyriakou et al. [51]. The authors studied crowd behavior realism, where the user task was to follow a child through the crowd. The conditions differed in how the crowd reacted to the user, from ignoring it altogether, avoiding collisions, and avoiding collisions while displaying basic social interactions. The results indicated that the more realistic the crowd reaction, the less the user performance. Authors suggest that the increase of realism of crowd led users also to behave more realistically. This is supported by the fact that when virtual characters were trying to avoid collisions, participants also started to avoid collisions at a higher rate. Furthermore, when virtual characters showed social interactions such as waving, users were found to also wave back to them. This led us to suggest that, similarly to Mania et al.'s study [67], low objective realism can be a way to "cheat." By downplaying the realism of a situation to the point where users do not need to follow social rules or obey physics, we are, in a way, decreasing the difficulty of "real life." However, we enforce that this is context-dependent, and higher objective realism over lower might not always be better.

Finally, the last reported negative case comes from Jeffrey Bertrand et al.'s study [5], where a negative effect of IVE System (Physics) was found on the Perceived Environment Realism. Authors studied the effect of the presence and absence of gravity. The results indicated that higher realism (gravity) condition resulted in an unexpected lower sensory fidelity factor score. They presented two possible justifications: one centered on a limitation of the questionnaire and another on a limitation of the experimental apparatus. The first one was due to questions that required the ability to examine objects closely, which was easier when no gravity is at play, and objects can float right in front of the users. The second was due to electromagnetic tracking controllers getting out of the boundary and introducing jittering in the tracking. Both factors could have influenced the results, causing a possible false negative (impact wise).

There seem to be some limitations regarding the equipment restrictions when trying to create objectively realistic IVEs. To virtually replicate the real world, enormous computational power and specialized equipment are often needed to provide users with faithful stimuli. Considering that VR usually requires higher refresh rates to reduce cybersickness [55] while also generally having higher resolutions than non-immersive setups, a higher load is sometimes put over the rendering pipeline saturating the hardware processing capacity. However, this type of equipment is not always possible to set up, and some researchers might opt to use available equipment to

conduct experiences that may not be up to the task. The authors can catch up with these limitations during the development phase and/or experimental phase. Still, sometimes they can pass unnoticed (mainly in situations with less robust methodologies), influencing the results without the author's knowledge. There may also be situations where authors are just conducting experiments already using cutting-edge technology and still encounter hardware/software limitations (depending on the context of the applications and how well coded and optimized they are).

We suggest studying equipment limitations thoroughly. What are their boundaries? What happens if users/authors push them too far? How can the study methodology mitigate these limitations? Is it worth it to push the equipment to its edge and risk biased results? Some of these points could be carefully accounted for while developing the study. If possible, then pilot studies could add an extra layer of testing. Users could behave in unexpected ways that could uncover hardware restrictions (e.g., leaving the tracking area, erratic head movements displaying the limits of lower refresh rates). There will always be boundaries in IVEs. Slater [105], in his introduction to PI and PsI, also discussed this problem: The more participants probe the virtual system, the more significant the change to break the PI. There seems to exist a tradeoff between giving participants freedom to explore and behave like in reality and breaking the illusion of "being there" when system limitations are met, or providing such a rich and detailed environment but with poorer performance.

4.5 Conclusions

This systematic review aimed at investigating, in an exploratory way, the terms and definitions being used in the literature to define realism in IVEs (RQ1) and the impact of objective realism in the user's virtual experiences (RQ2).

RQ1 results led us to conclude that there are two mainly used terms: realism and fidelity. Overall, we verified several authors confound use of realism-related terms, while some did not even provide a definition. Therefore, a table was created describing each term and its definition (if present) (see supplementary material). We have proposed segmentation of realism into subjective and objective realism for the scope of this study. Objective realism defines how well a virtual experience replicates the real world, determined by what the immersive system allows and how well we take advantage of it. Subjective realism relates to how users perceived the virtual experience as being real, whether the virtual experience depicts a scenario possible to happen in the real world or not.

RQ2 results indicated that objective realism had an overall positive impact and very low negative and mixed impacts. Realism had the highest positive impact on embodiment and less on physiological responses. Only three dependent categories were negatively affected by objective realism (although with very low occurrences count): physiological responses, user performance, and perceived environment realism. The three less researched dependent categories (with equal or less than 10 occurrences) were embodiment, virtual agents, and user behavior. We suggest further research on these less-evaluated factors to increase the validity of the results further.

Regarding the Realism Factors, the least independently studied ones (with equal or less than 10 occurrences) were IVE System (Audio), IVE System (Physics), and IVE Content (Scent). Therefore, we suggest further studies to verify the impact of such Realism Factors in IVEs, especially the IVE Content (Scent), where no occurrences were found. A table was created denoting all the results found for each document (see supplementary material).

Methodology or/and equipment limitations were found to be contributors to the negative impacts registered. Examples of the first one were methodologies that facilitated the users' performance by simplifying the scene with less-objective realism or methodologies that promoted a higher risk of simulator sickness. Hardware limitations contributed to an increase in motion sickness symptoms. We consider the following points as the main takeaways from this study:

- Many authors use their own definitions for the same terms, leading to possible misinterpretations between studies. A taxonomy is a need in this regard.
- Lack of taxonomy to define Realism Factors and the resulting user experience in IVEs.
- The majority of studies have shown that objective realism on IVEs has a positive impact on user experience, followed by a neutral impact and only specific negative, mixed, and indefinite impact studies were recorded.
- Haptic stimuli are the most supported to improve the perceived environment realism.
- Gaps were found where some Realism Factors' impact on specific parts of the user experience is still left to be studied or needs more corroborating studies.
- Some Realism Factors impacting user experience in IVEs, such as scent and taste, are still to be researched.
- The influence of some variables within Realism Factors is yet to be considered (e.g., IVE Content (Audio) reverberation or audio occlusion).
- Objective realism negative impact on user experience was found to be related mainly to methodological and technological limitations. Researchers should perform pilot studies to anticipate possible unexpected user reactions probing the system limitations while also keeping track of system performance.

5 LIMITATIONS AND FUTURE WORK

As with any study, this systematic review encountered some barriers. Some difficulties in the data synthesis were already discussed in Section 4.2.

Regarding the realism and related terms definitions, no taxonomy was found, resulting in a confused use of terms by many authors. We proposed dividing realism into subjective and objective realism for this study scope. However, a proper taxonomy should be created in future work, because the terms reviewed under RQ1 were under the inherent limitations of the selected papers for RQ2. Because there was no proper taxonomy where Realism Factors and the resulting user experience could be based, we have created one for this study. However, such was not the focus of this study, and the methodology to develop such taxonomy was not formal. In future work, a taxonomy for Realism Factors and resulting user experience in IVEs should be addressed. The category involvement merged several different variables that, although all related to the user involvement with the IVEs, could also fit in other new categories.

There is a limitation in the quality score field "document type," as some conferences can have higher-quality documents than some journals. This limitation is partially mitigated through the other two quality scores (Methodology and Sample). If the paper is of high quality, then it will still be considered high quality even if at a conference (1 point from conference plus 6 points from methodology and sample). There was no checklist for methodology quality rating, because each study is unique, and what works for a study may not work for another. Therefore, this metric is prone to reviewer subjectiveness, which is partially mitigated by having two reviewers rating their quality, and when there was no consensus, a third reviewer would help the final decision. The paper quality assessment scores were meant to describe the overall quality of the sample and were not taken into account in the realism impact scoring. This means that the result of a high-quality paper would have the same weight as the result of a lower-quality paper.

This study allows a global overview of how different Realism Factors affect the virtual experience, but we cannot conclude how much, as effect sizes were not taken into account. It also presents a substantial potential bias, as multiple studies of low statistical power can skew results. As future work, a meta-analysis should be considered to perform a more deep and robust analysis of the data. This study's results and statements only refer to IVEs and should not be generalized outside this systematic review scope (non-immersive experiences) without further investigation. The

“indeterminate” impact could not be properly discussed, as authors did not state which conditions were better/worse.

Some studies changed more than one variable between conditions. When the variables were of different Realism Factors, we could not conclude if the result was due to one of the Realism Factors or cross-effects. This resulted in nine different combinations, and due to each one being supported by only one study, we were not able to present proper conclusions about them.

Another limitation is the fact that we cannot directly compare studies results even if they evaluate the same Realism Factor, as the rest of the IVE objective realism is different in almost every aspect.

Some authors did not compare the user behavior or physiological responses to a real-world condition, which can present a limitation, as their baseline was done inside the IVE.

Publication bias and Hawthorne effect in the selected papers may have affected the results of this systematic review. Some authors may omit certain results from their paper because they consider it “undesirable,” which would influence the results of this article up to an extent.

We wanted to keep the immersive visual system consistent (discarding comparisons between a given HMD to another or HMD to CAVE’s). This allowed us to organize the review better and understand, using the same visual system, how several Realism Factors influence the user’s IVE. However, in part, we also consider it a limitation, because a determined type of HMD, CAVE, or Large Stereoscopic Screen characteristics could increase the IVE realism (e.g., higher resolution, field of view, or better color representation).

Although the search was performed to include as many realism-related studies in the context of IVEs, some might still be left out due to not matching the query keywords. For example, this could happen when authors are testing a particular set of variables indirectly or directly related to the IVE realism, but not mentioning it, as it was not within their scope. However, further increasing the query coverage would result in an incredible number of documents to be analyzed.

Some variables could fit more than one factor/category. However, to keep the results and discussion more straightforward to understand, the research team opted not to use a determinate variable in more than one factor simultaneously. Instead, these ambiguous variables would be introduced in the factor the research team recognized they would better fit for study context.

REFERENCES

- [1] 2018. ISO 9241-11:2018(en) Ergonomics of human-system interaction–Part 11: Usability: Definitions and concepts. <https://www.iso.org/standard/63500.html>.
- [2] Amy L. Alexander, T. T. Brunyé, Jason Sidman, and Shawn A. Weil. 2005. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games. *DARWARS Training Impact Group 5* (2005), 1–14.
- [3] Rosa María Baños, Cristina Botella, Azucena García-Palacios, Helena Villa, Concepción Perpiñá, and Mariano Alcaniz. 2000. Presence and reality judgment in virtual environments: A unitary construct? *CyberPsychol. Behav.* 3, 3 (2000), 327–335.
- [4] Hrvoje Benko, Christian Holz, Mike Sinclair, and Eyal Ofek. 2016. NormalTouch and textureTouch. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM. DOI : <https://doi.org/10.1145/2984511.2984526>
- [5] Jeffrey Bertrand, Ayush Bhargava, Kapil Chalil Madathil, Anand Gramopadhye, and Sabarish V. Babu. 2017. The effects of presentation method and simulation fidelity on psychomotor education in a bimanual metrology training simulation. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. DOI : <https://doi.org/10.1109/3dui.2017.7893318>
- [6] Frank Biocca and Ben Delaney. 1995. Immersive virtual reality technology. *Commun. Age Virt. Real.* 15, 32 (1995), 10–5555.
- [7] Ian D. Bishop, JoAnna R. Wherrett, and David R. Miller. 2001. Assessment of path choices on a country walk using a virtual environment. *Landsc. Urb. Plann.* 52, 4 (2001), 225–237.

- [8] K. J. Blom and S. Beckhaus. 2013. Virtual travel collisions: Response method influences perceived realism of virtual environments. *ACM Trans. Appl. Percept.* 10, 4 (2013). DOI : <https://doi.org/10.1145/2536764.2536772>
- [9] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands “feel” touch that eyes see. *Nature* 391, 6669 (1998), 756–756.
- [10] Fabien Boucaud, Quentin Tafiani, Catherine Pelachaud, and Indira Thouvenin. 2019. Social touch in human-agent interactions in an immersive virtual environment. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications. DOI : <https://doi.org/10.5220/0007397001290136>
- [11] Andrea Brogni, Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. 2006. Variations in physiological responses of participants during different stages of an immersive virtual environment experiment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. ACM Press. DOI : <https://doi.org/10.1145/1180495.1180572>
- [12] Kirsten Cater, Alan Chalmers, and Patrick Ledda. 2002. Selective quality rendering by exploiting human inattention blindness: Looking but not seeing. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST’02)*. Association for Computing Machinery, New York, NY, 17–24. DOI : <https://doi.org/10.1145/585740.585744>
- [13] Alan Chalmers and Andrej Ferko. 2008. Levels of realism: From virtual reality to real virtuality. In *Proceedings of the 24th Spring Conference on Computer Graphics (SCCG’08)*. Association for Computing Machinery, 19–25. DOI : <https://doi.org/10.1145/1921264.1921272>
- [14] Alan Chalmers, David Howard, and Christopher Moir. 2009. Real virtuality: A step change from virtual reality. In *Proceedings of the 25th Spring Conference on Computer Graphics (SCCG’09)*. Association for Computing Machinery, New York, NY, 9–16. DOI : <https://doi.org/10.1145/1980462.1980466>
- [15] Vuthea Chheang, Patrick Saalfeld, Tobias Huber, Florentine Huettl, Werner Kneist, Bernhard Preim, and Christian Hansen. 2019. Collaborative virtual reality for laparoscopic liver surgery training. In *Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR’19)*. 1–17. DOI : <https://doi.org/10.1109/AIVR46125.2019.00011>
- [16] George Coates. 1992. Invisible Site - A virtual show, a multimedia performance work presented by George Coates Performance Works. <https://doi.org/10.1111/j.1460-2466.1992.tb00812.x>
- [17] Thomas M. Connolly, Elizabeth A. Boyle, Ewan MacArthur, Thomas Hainey, and James M. Boyle. 2012. A systematic literature review of empirical evidence on computer games and serious games. *Comput. Educ.* (2012). DOI : <https://doi.org/10.1016/j.compedu.2012.03.004>
- [18] J. Conti, B. Ozell, E. Paquette, and P. Renaud. 2017. Adjusting stereoscopic parameters by evaluating the point of regard in a virtual environment. *Comput. Graph. (Pergamon)* 69 (2017), 24–35. DOI : <https://doi.org/10.1016/j.cag.2017.08.017>
- [19] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer. 2018. The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLoS One* 13, 2 (2018). DOI : <https://doi.org/10.1371/journal.pone.0191846>
- [20] D. A. Bowman and R. P. McMahan. 2007. Virtual reality: How much immersion is enough? *Computer* 40, 7 (2007), 36–43. DOI : <https://doi.org/10.1109/MC.2007.257>
- [21] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. 2014. A systematic review of cybersickness. In *Proceedings of the Conference on Interactive Entertainment*. ACM. DOI : <https://doi.org/10.1145/2677758.2677780>
- [22] Zhenan Feng, Vicente A. González, Robert Amor, Ruggiero Lovreglio, and Guillermo Cabrera-Guerrero. 2018. Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Comput. Educ.* (2018). DOI : <https://doi.org/10.1016/j.compedu.2018.09.002>
- [23] James A. Ferwerda. 2003. Three varieties of realism in computer graphics. In *Human Vision and Electronic Imaging VIII*, Vol. 5007. International Society for Optics and Photonics, 290–297.
- [24] Anton Franzluebbers and Kyle Johnsen. 2018. Performance benefits of high-fidelity passive haptic feedback in virtual reality training. In *Proceedings of the Symposium on Spatial User Interaction*. DOI : <https://doi.org/10.1145/3267782.3267790>
- [25] Anton Franzluebbers and Kyle Johnsen. 2018. Performance benefits of high-fidelity passive haptic feedback in virtual reality training. In *Proceedings of the Symposium on Spatial User Interaction*. ACM. DOI : <https://doi.org/10.1145/3267782.3267790>
- [26] Philippe Fuchs, Guillaume Moreau, and Pascal Guitton (Eds.). 2011. *Virtual Reality: Concepts and Technologies*. CRC Press, Boca Raton, FL.
- [27] Dominik Gall and Marc Erich Latoschik. 2018. The effect of haptic prediction accuracy on presence. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. DOI : <https://doi.org/10.1109/vr.2018.8446153>

- [28] James J. Gibson. 1978. The ecological approach to the visual perception of pictures. *Leonardo* (1978). DOI : <https://doi.org/10.2307/1574154>
- [29] Stephen B. Gilbert. 2016. Perceived realism of virtual environments depends on authenticity. 322–324 pages. https://dl.acm.org/doi/10.1162/PRES_a_00276.
- [30] G. Gonçalves, M. Melo, J. Vasconcelos-Raposo, and M. Bessa. 2019. Impact of different sensory stimuli on presence in credible virtual environments. *IEEE Trans. Visualiz. Comput. Graph.* 26, 11 (2019), 3231–3240. DOI : <https://doi.org/10.1109/TVCG.2019.2926978>
- [31] Geoffrey Gorisse, Olivier Christmann, Samory Houzangbe, and Simon Richir. 2018. From robot to virtual doppelganger. In *Proceedings of the International Conference on Advanced Visual Interfaces*. ACM. DOI : <https://doi.org/10.1145/3206505.3206525>
- [32] P. Greenbaum. 1992. The lawnmower man. *Film Video* 9, 3 (1992), 58–62. <https://doi.org/10.1111/j.1460-2466.1992.tb00812.x>
- [33] R. E. Guadagno, K. R. Swinth, and J. Blascovich. 2011. Social evaluations of embodied agents and avatars. *Comput. Hum. Behav.* 27, 6 (2011), 2380–2385. DOI : <https://doi.org/10.1016/j.chb.2011.07.017>
- [34] Mahmoud Habibnezhad, Jay Puckett, Mohammad Sadra Fardhosseini, and Lucky Agung Pratama. 2019. A mixed VR and physical framework to evaluate impacts of virtual legs and elevated narrow working space on construction workers gait pattern. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)*. International Association for Automation and Robotics in Construction (IAARC). DOI : <https://doi.org/10.22260/isarc2019/0141>
- [35] Alice Hall. 2003. Reading realism: Audiences' evaluations of the reality of media texts. *J. Commun.* 53, 4 (2003), 624–641.
- [36] Joel Harman, Ross Brown, Daniel Johnson, and Selen Turkay. 2019. The role of visual detail during situated memory recall within a virtual reality environment. In *Proceedings of the 31st Australian Conference on Human-computer-Interaction*. ACM. DOI : <https://doi.org/10.1145/3369457.3369469>
- [37] Derek Harter, Shulan Lu, Pratyush Kotturu, and Devin Pierce. 2011. An immersive virtual environment for varying risk and immersion for effective training. In *Proceedings of the World Conference on Innovative Virtual Reality*, Vol. 44328. 301–307.
- [38] Jason Hochreiter, Salam Daher, Gerd Bruder, and Greg Welch. 2018. Cognitive and touch performance effects of mismatched 3D physical and visual perceptions. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. DOI : <https://doi.org/10.1109/vr.2018.8446574>
- [39] J. Y. Hong, B. Lam, Z.-T. Ong, K. Ooi, W.-S. Gan, J. Kang, J. Feng, and S.-T. Tan. 2019. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. *Build. Environ.* 149 (2019), 1–14. DOI : <https://doi.org/10.1016/j.buildenv.2018.12.004>
- [40] Johan F. Hoorn, Elly A. Konijn, and Gerrit C. Van der Veer. 2003. Virtual reality: Do not augment realism, augment relevance. *Hum.-comput. Interact.: Overcom. Barriers* 4, 1 (2003), 18–26.
- [41] Tom Hughes and Evan Rolek. 2003. Fidelity and validity: Issues of human behavioral representation requirements development. In *Proceedings of the Winter Simulation Conference*. DOI : <https://doi.org/10.1109/wsc.2003.1261519>
- [42] Jonatan Hvass, Oliver Larsen, Kasper Vendelbo, Niels Nilsson, Rolf Nordahl, and Stefania Serafin. 2017. Visual realism and presence in a virtual reality game. In *Proceedings of the 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON'17)*. IEEE, 1–4. DOI : <https://doi.org/10.1109/3DTV.2017.8280421>
- [43] Laurent Itti and Christof Koch. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* 40, 10 (2000), 1489–1506. DOI : [https://doi.org/10.1016/S0042-6989\(99\)00163-7](https://doi.org/10.1016/S0042-6989(99)00163-7)
- [44] J. Y. Jeon and H. I. Jo. 2019. Three-dimensional virtual reality-based subjective evaluation of road traffic noise heard in urban high-rise residential buildings. *Build. Environ.* 148 (2019), 468–477. DOI : <https://doi.org/10.1016/j.buildenv.2018.11.004>
- [45] Richard D. Joyce and Stephen Robinson. 2017. Passive haptics to enhance virtual reality simulations. In *Proceedings of the AIAA Modeling and Simulation Technologies Conference*. American Institute of Aeronautics and Astronautics. DOI : <https://doi.org/10.2514/6.2017-1313>
- [46] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychology* 3, 3 (1993), 203–220. DOI : https://doi.org/10.1207/s15327108ijap0303_3
- [47] A. C. Kern and W. Ellermeier. 2020. Audio in VR: Effects of a soundscape and movement-triggered step sounds on presence. *Front. Robot. AI* 7 (2020). DOI : <https://doi.org/10.3389/frobt.2020.00020>
- [48] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. Keele University, UK and National ICT Australia. DOI : <https://doi.org/10.1.1.122.3308>
- [49] Robert Konrad, Anastasios Angelopoulos, and Gordon Wetzstein. 2020. Gaze-contingent ocular parallax rendering for virtual reality. *ACM Trans. Graph.* 39, 2 (Apr. 2020), 1–12. DOI : <https://doi.org/10.1145/3361330>

- [50] David M. Krum, Sin-Hwa Kang, and Thai Phan. 2018. Influences on the elicitation of interpersonal space with virtual humans. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. DOI: <https://doi.org/10.1109/vr.2018.8446235>
- [51] M. Kyriakou, X. Pan, and Y. Chrysanthou. 2016. Interaction with virtual crowd in immersive and semi-immersive virtual reality systems. *Comput. Anim. Virt. Worlds* 28, 5 (2016). DOI: <https://doi.org/10.1002/cav.1729>
- [52] C. Kyriltsias and D. Michael-Grigoriou. 2018. Asch conformity experiment using immersive virtual reality. *Comput. Anim. Virt. Worlds* 29, 5 (2018). DOI: <https://doi.org/10.1002/cav.1804>
- [53] H. M. L. 1962. *Sensorama Simulator*. Google Patents.
- [54] B. Laha, D. A. Bowman, and J. D. Schiffbauer. 2013. Validation of the MR simulation approach for evaluating the effects of immersion on visual analysis of volume data. *IEEE Trans. Visualiz. Comput. Graph.* 19, 4 (2013), 529–538. DOI: <https://doi.org/10.1109/TVCG.2013.43>
- [55] Joseph J. LaViola. 2000. A discussion of cybersickness in virtual environments. *SIGCHI Bull.* 32, 1 (Jan. 2000), 47–56. DOI: <https://doi.org/10.1145/333329.333344>
- [56] Charles E. Leiserson, Neil C. Thompson, Joel S. Emer, Bradley C. Kuszmaul, Butler W. Lampson, Daniel Sanchez, and Tao B. Schardl. 2020. There's plenty of room at the top: What will drive computer performance after Moore's law? *Science* 368, 6495 (2020).
- [57] Xiangyang Li, Qinhe Gao, Zhili Zhang, and Xianxiang Huang. 2012. Collaborative virtual maintenance training system of complex equipment based on immersive virtual reality environment. *Assem. Autom.* 32, 1 (2012), 72–85. <https://doi.org/10.1108/01445151211198737>
- [58] Q. Lin, J. Rieser, and B. Bodenheimer. 2015. Affordance judgments in HMD-based virtual environments: Stepping over a pole and stepping off a ledge. *ACM Trans. Appl. Percept.* 12, 2 (2015). DOI: <https://doi.org/10.1145/2720020>
- [59] Edwin A. Link Jr. 1937. *Trainer for Aviators*. Google Patents. <https://patents.google.com/patent/US2099857A/en>.
- [60] Matthew Lombard and Theresa Ditton. 2006. At the heart of it all: The concept of presence. *J. Comput.-mediat. Commun.* 3, 2 (2006), 0. DOI: <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x>
- [61] Matthew Lombard, Theresa B. Ditton, Daliza Crane, Bill Davis, Gisela Gil-Egui, Karl Horvath, Jessica Rossman, and S. Park. 2000. Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Proceedings of the 3rd International Workshop on Presence*, Vol. 240. 2–4.
- [62] Peter Longhurst, Patrick Ledda, and Alan Chalmers. 2003. Psychophysically based artistic techniques for increased perceived realism of virtual environments. In *Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (AFRIGRAPH'03)*. Association for Computing Machinery, 123–132. DOI: <https://doi.org/10.1145/602330.602354>
- [63] Cedric Di Loreto, Jean-Remy Chardonnet, Julien Ryard, and Alain Rousseau. 2018. WoaH: A virtual reality work-at-height simulator. In *Proceedings of the Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. DOI: <https://doi.org/10.1109/vr.2018.8448292>
- [64] Ritch Macefield. 2009. How to specify the participant group size for usability studies: A practitioner's guide. *J. Usabil. Stud.* (2009), 34–45. <https://dl.acm.org/doi/10.5555/2835425.2835429>
- [65] N. Magnenat-Thalmann, HyungSeok Kim, A. Egges, and S. Garchery. 2005. Believability and interaction in virtual worlds. In *Proceedings of the 11th International Multimedia Modelling Conference*. IEEE. DOI: <https://doi.org/10.1109/mmmc.2005.24>
- [66] Steven Malliet. 2006. An exploration of adolescents' perceptions of videogame realism. *Learn., Media Technol.* 31, 4 (2006), 377–394.
- [67] K. Mania, S. Badariah, M. Coxon, and P. Watten. 2010. Cognitive transfer of spatial awareness states from immersive virtual environments to reality. *ACM Trans. Appl. Percept.* 7, 2 (2010). DOI: <https://doi.org/10.1145/1670671.1670673>
- [68] K. Mania and A. Robinson. 2005. An experimental exploration of the relationship between subjective impressions of illumination and physical fidelity. *Comput. Graph. (Pergamon)* 29, 1 (2005), 49–56. DOI: <https://doi.org/10.1016/j.cag.2004.11.007>
- [69] K. Mania, A. Robinson, and K. R. Brandt. 2005. The effect of memory schemas on object recognition in virtual environments. *Pres.: Teleop. Virt. Environ.* 14, 5 (2005), 606–615. DOI: <https://doi.org/10.1162/105474605774918769>
- [70] K. Mania, D. Wooldridge, M. Coxon, and A. Robinson. 2006. The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Trans. Visualiz. Comput. Graph.* 12, 3 (2006), 396–404. DOI: <https://doi.org/10.1109/TVCG.2006.55>
- [71] A. Maselli and M. Slater. 2013. The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7 (2013). DOI: <https://doi.org/10.3389/fnhum.2013.00083>
- [72] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. 2012. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Trans. Visualiz. Comput. Graph.* 18, 4 (2012), 626–633. DOI: <https://doi.org/10.1109/TVCG.2012.43>

- [73] Daniel Medeiros, Rafael K. dos Anjos, Daniel Mendes, João Madeiras Pereira, Alberto Raposo, and Joaquim Jorge. 2018. Keep my head on my shoulders! In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM. DOI : <https://doi.org/10.1145/3281505.3281511>
- [74] M. Melo, G. Gonçalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa. 2020. Do multisensory stimuli benefit the virtual reality experience? A systematic review. *IEEE Trans. Visualiz. Comput. Graph.* (2020), 1–1. DOI : <https://doi.org/10.1109/TVCG.2020.3010088>
- [75] Georg F. Meyer, Li Ting Wong, Emma Timson, Philip Perfect, and Mark D. White. 2012. Objective fidelity evaluation in multisensory virtual environments: Auditory cue fidelity in flight simulation. *PLoS One* (2012). DOI : <https://doi.org/10.1371/journal.pone.0044381>
- [76] Paul Milgram and Fumio Kishino. 1994. Taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* E77-D, 12 (1994), 1321–1329. https://www.researchgate.net/publication/231514051_A_Taxonomy_of_Mixed_Reality_Visual_Displays.
- [77] Stanley Milgram. 1963. Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67, 4 (1963), 371–378. DOI : <https://doi.org/10.1037/h0040525>
- [78] X. Min, W. Zhang, S. Sun, N. Zhao, S. Tang, and Y. Zhuang. 2019. VPMoel: High-fidelity product simulation in a virtual-physical environment. *IEEE Trans. Visualiz. Comput. Graph.* 25, 11 (2019), 3083–3093. DOI : <https://doi.org/10.1109/TVCG.2019.2932276>
- [79] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and the PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. DOI : <https://doi.org/10.1371/journal.pmed.1000097>
- [80] Pedro Monteiro, Miguel Melo, António Valente, José Vasconcelos-Raposo, and Maximino Bessa. 2021. Delivering critical stimuli for decision making in VR training: Evaluation study of a firefighter training scenario. *IEEE Trans. Hum.-mach. Syst.* 51, 2 (Apr. 2021), 65–74. DOI : <https://doi.org/10.1109/THMS.2020.3030746>
- [81] Gordon M. Moore. 1965. Cramming more components onto integrated circuits with unit cost. *Electronics* 38, 8 (1965), 114.
- [82] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* 19, 2 (2012), 98–100. DOI : <https://doi.org/10.1109/MRA.2012.2192811>
- [83] William Moroney and Michael Lilienthal. 2008. Human factors in simulation and training. In *Human Factors in Simulation and Training*. CRC Press, 3–38. DOI : <https://doi.org/10.1201/9781420072846.sec1>
- [84] N. Mourkoussis, F. M. Rivera, T. Troscianko, T. Dixon, R. Hawkes, and K. Mania. 2010. Quantifying fidelity for virtual environment simulations employing memory schema assumptions. *ACM Trans. Appl. Percept.* 8, 1 (2010). DOI : <https://doi.org/10.1145/1857893.1857895>
- [85] Thomas Muender, Anke V. Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. 2019. Does it feel real? In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM. DOI : <https://doi.org/10.1145/3290605.3300903>
- [86] David Narciso, Miguel Melo, José Vasconcelos Raposo, João Cunha, and Maximino Bessa. 2020. Virtual reality in training: An experimental study with firefighters. *Multim. Tools Applic.* 79, 9 (2020), 6227–6245.
- [87] Nami Ogawa, Takuji Narumi, and Michitaka Hirose. 2018. Object size perception in immersive virtual reality: Avatar realism affects the way we perceive. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. DOI : <https://doi.org/10.1109/vr.2018.8446318>
- [88] B. Perroud, S. Régner, A. Kemeny, and F. Mérenne. 2019. Model of realism score for immersive VR systems. *Transport. Res. Part F: Traff. Psychol. Behav.* 61 (2019), 238–251. DOI : <https://doi.org/10.1016/j.trf.2017.08.015>
- [89] A. Petti, W. Hutabarat, J. Oyekan, C. Turner, A. Tiwari, N. Prajapat, X.-P. Gan, and N. Ince. 2016. Impact of model fidelity in factory layout assessment using immersive discrete event simulation, In *Proceedings of the 8th Operational Research Society Simulation Workshop (SW'16)*. 124–134.
- [90] Darque Pinto, Bruno Peixoto, Guilherme Gonçalves, Miguel Melo, Vasco Amorim, and Maximino Bessa. 2019. Developing training applications for hydrogen emergency response training. In *Proceedings of the International Conference on Graphics and Interaction (ICGI'19)*. 130–136. DOI : <https://doi.org/10.1109/ICGI47575.2019.8955091>
- [91] Andreas Pusch, Olivier Martin, and Sabine Coquillart. 2011. Effects of hand feedback fidelity on near space pointing performance and user acceptance. In *Proceedings of the IEEE International Symposium on VR Innovation*. IEEE. DOI : <https://doi.org/10.1109/isvri.2011.5759609>
- [92] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan. 2015. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Trans. Visualiz. Comput. Graph.* 21, 7 (2015), 794–807. DOI : <https://doi.org/10.1109/TVCG.2015.2403312>
- [93] Nimesha Ranasinghe, Pravara Jain, Shienny Karwita, David Tolley, and Ellen Yi-Luen Do. 2017. Ambiotherm. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM. DOI : <https://doi.org/10.1145/3025453.3025723>

- [94] T. Randhavane, A. Bera, and D. Manocha. 2017. F2Crowds: Planning agent movements to enable face-to-face interactions. *Pres.: Teleop. Virt. Environ.* 26, 2 (2017), 228–246. DOI: https://doi.org/10.1162/PRES_a_00294
- [95] Katja Rogers, Jana Funke, Julian Frommel, Sven Stamm, and Michael Weber. 2019. Exploring interaction fidelity in virtual reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM. DOI: <https://doi.org/10.1145/3290605.3300644>
- [96] Arnon Rolnick and ubow. 1991. Why is the driver rarely motion sick? The role of controllability in motion sickness. *Ergonomics* 34, 7 (1991), 867–879. DOI: <https://doi.org/10.1080/00140139108964831>
- [97] Manfred Roza, Jeroen Voogd, and Paul van Gool. 2000. Fidelity considerations for civil aviation distributed simulations. In *Proceedings of the Modeling and Simulation Technologies Conference*. DOI: <https://doi.org/10.2514/6.2000-4397>
- [98] Holger Salzmann and Bernd Froehlich. 2008. The two-user seating buck: Enabling face-to-face discussions of novel car interface concepts. In *Proceedings of the IEEE Virtual Reality Conference*. IEEE. DOI: <https://doi.org/10.1109/vr.2008.4480754>
- [99] B. C. Schricker, R. W. Franceschini, and T. C. Johnson. 2001. Fidelity evaluation framework. In *Proceedings of the 34th Annual Simulation Symposium*. 109–116. DOI: <https://doi.org/10.1109/SIMSYM.2001.922122>
- [100] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Pres.: Teleop. Virt. Environ.* 10, 3 (2001), 266–281. DOI: <https://doi.org/10.1162/105474601300343603>
- [101] Z. S. See, A. Dey, L. Goodman, Y. K. Ng, C. Hight, M. Billinghurst, and M. S. Sunar. 2019. Creating high fidelity 360° virtual reality with high dynamic range spherical panorama images. *Virt. Creat.* 9, 1–2 (2019), 73–109. DOI: https://doi.org/10.1386/vcr_00006_1
- [102] R. Skarbez, F. Brooks, and M. Whitton. 2020. Immersion and coherence: Research agenda and early results. *IEEE Trans. Visualiz. Comput. Graph.* (2020). DOI: <https://doi.org/10.1109/TVCG.2020.2983701>
- [103] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A survey of presence and related concepts. *Comput. Surv.* 50, 6 (11 2017), 1–39. DOI: <https://doi.org/10.1145/3134301>
- [104] Richard Skarbez, Solene Neyret, Frederick P. Brooks, Mel Slater, and Mary C. Whitton. 2017. A psychophysical experiment regarding components of the plausibility illusion. *IEEE Trans. Visualiz. Comput. Graph.* 23, 4 (2017), 1369–1378. DOI: <https://doi.org/10.1109/TVCG.2017.2657158>
- [105] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. Roy. Societ. B: Biol. Sci.* 364, 1535 (2009), 3549–3557. DOI: <https://doi.org/10.1098/rstb.2009.0138>
- [106] Mel Slater. 2013. A Note on Presence Terminology. https://www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm.
- [107] Mel Slater, Angus Antley, Adam Davison, David Swapp, Christoph Guger, Chris Barker, Nancy Pistrang, and Maria V. Sanchez-Vives. 2006. A virtual reprise of the Stanley Milgram obedience experiments. *PLoS One* 1, 1 (2006), e39. DOI: <https://doi.org/10.1371/journal.pone.0000039>
- [108] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. 2009. Visual realism enhances realistic response in an immersive virtual. *IEEE Comput. Graph. Applic.* 29, 3 (2009), 76–84. DOI: [10.1109/MCG.2009.55](https://doi.org/10.1109/MCG.2009.55)
- [109] M. Slater, P. Khanna, J. Mortensen, and I. Yu. 2009. Visual realism enhances realistic response in an immersive virtual environment. *IEEE Comput. Graph. Applic.* 29, 3 (2009), 76–84. DOI: <https://doi.org/10.1109/MCG.2009.55>
- [110] Mel Slater and Maria V. Sanchez-Vives. 2016. Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3 (2016). DOI: <https://doi.org/10.3389/frobt.2016.00074>
- [111] M. Slater, B. Spanlang, and D. Corominas. 2010. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graph.* 29, 4 (2010). DOI: <https://doi.org/10.1145/1778765.1778829>
- [112] Mel Slater and Sylvia Wilbur. 1997. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Pres.: Teleop. Virt. Environ.* 6, 6 (1997), 603–616. DOI: <https://doi.org/10.1162/pres.1997.6.6.603>
- [113] C. Stinson and D. A. Bowman. 2014. Feasibility of training athletes for high-pressure situations using virtual reality. *IEEE Trans. Visualiz. Comput. Graph.* 20, 4 (2014), 606–615. DOI: <https://doi.org/10.1109/TVCG.2014.23>
- [114] V. Sundstedt, K. Debattista, P. Longhurst, A. Chalmers, and T. Troschianko. 2005. Visual attention for efficient high-fidelity graphics. In *Proceedings of the 21st Spring Conference on Computer Graphics (SCCG'05)*. Association for Computing Machinery, New York, NY, 169–175. DOI: <https://doi.org/10.1145/1090122.1090150>
- [115] Ivan E. Sutherland. 1965. The ultimate display. *Proc. IFIP Congr.* 2 (1965), 506–508.
- [116] René Ter Haar. 2005. Virtual reality in the military: Present and future. In *Proceedings of the 3rd Twente Student Conference on IT*. Citeseer.
- [117] Jerald Thomas, Mahdi Azmandian, Sonia Grunwald, Donna Le, David Krum, Sin-Hwa Kang, and Evan Suma Rosenberg. 2017. Effects of personalized avatar texture fidelity on identity recognition in virtual reality. DOI: <https://doi.org/10.2312/EGVE.20171345>

- [118] Jacob Thorn, Rodrigo Pizarro, Bernhard Spanlang, Pablo Bermell-Garcia, and Mar Gonzalez-Franco. 2016. Assessing 3D scan quality through paired-comparisons psychophysics. In *Proceedings of the 24th ACM International Conference on Multimedia*. ACM. DOI : <https://doi.org/10.1145/2964284.2967200>
- [119] H. Tian, C. Wang, D. Manocha, and X. Zhang. 2018. Realtime hand-object interaction using learned grasp space for virtual environments. *IEEE Trans. Visualiz. Comput. Graph.* 25, 8 (2018), 2623–2635. DOI : <https://doi.org/10.1109/TVCG.2018.2849381>
- [120] Dimitar Valkov, John Martens, and Klaus Hinrichs. 2016. Evaluation of the effect of a virtual avatar's representation on distance perception in immersive virtual environments. In *Proceedings of the IEEE Virtual Reality (VR)*. IEEE. DOI : <https://doi.org/10.1109/vr.2016.7504775>
- [121] Koorosh Vaziri, Peng Liu, Sahar Aseeri, and Victoria Interrante. 2017. Impact of visual and experiential realism on distance perception in VR using a custom video see-through system. In *Proceedings of the ACM Symposium on Applied Perception*. ACM. DOI : <https://doi.org/10.1145/3119881.3119892>
- [122] M. Weier, T. Roth, E. Kruijff, A. Hinkenjann, A. Pérard-Gayot, P. Slusallek, and Y. Li. 2016. Foveated real-time ray tracing for head-mounted displays. *Comput. Graph. Forum* 35, 7 (2016), 289–298. DOI : <https://doi.org/10.1111/cgf.13026>
- [123] Jonathan Wendt, Benjamin Weyers, Jonas Stienen, Andrea Bönsch, Michael Vorländer, and Torsten W. Kuhlen. 2019. Influence of directivity on the perception of embodied conversational agents' speech. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. ACM. DOI : <https://doi.org/10.1145/3308532.3329434>
- [124] Bob G. Witmer and Michael J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Pres.: Teleop. Virt. Environ.* 7, 3 (1998), 225–240. DOI : <https://doi.org/10.1162/105474698565686>
- [125] B. G. Witmer and M. J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240. DOI : <https://doi.org/10.1162/105474698565686>
- [126] Nick Yee, Jeremy N. Bailenson, and Nicolas Ducheneaut. 2009. The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Commun. Res.* 36, 2 (2009), 285–312.
- [127] Dohyeon Yeo, Gwangbin Kim, and Seungjun Kim. 2020. Toward immersive self-driving simulations: Reports from a user study across six platforms. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM. DOI : <https://doi.org/10.1145/3313831.3376787>
- [128] E. Yildiz, M. Melo, C. Möller, and M. Bessa. 2019. Designing collaborative and coordinated virtual reality training integrated with virtual and physical factories. In *Proceedings of the International Conference on Graphics and Interaction (ICGI)*. 48–55. DOI : <https://doi.org/10.1109/ICGI47575.2019.8955033>
- [129] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater. 2012. Visual realism enhances realistic response in an immersive virtual environment - Part 2. *IEEE Comput. Graph. Applic.* 32, 6 (2012), 36–45. DOI : <https://doi.org/10.1109/MCG.2012.121>
- [130] David Zeltzer. 1992. Autonomy, interaction, and presence. *Pres.: Teleop. Virt. Environ.* 1, 1 (1992), 127–132. DOI : <https://doi.org/10.1162/pres.1992.1.1.127>
- [131] X. Zhou and P.-L. P. Rau. 2019. Determining fidelity of mixed prototypes: Effect of media and physical interaction. *Appl. Ergon.* 80 (2019), 111–118. DOI : <https://doi.org/10.1016/j.apergo.2019.05.007>

Received 2 March 2021; revised 13 January 2022; accepted 23 April 2022