IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 29, 2021, accepted June 3, 2021, date of publication June 16, 2021, date of current version June 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089946

Systematic Review on Realism Research Methodologies on Immersive Virtual, Augmented and Mixed Realities

GUILHERME GONÇALVES[1,2], PEDRO MONTEIRO[1,2], HUGO COELHO[1,2], MIGUEL MELO[2], AND MAXIMINO BESSA[1,2]

[1]Department of Engineering, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal
[2]Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal

Corresponding author: Guilherme Gonçalves (guilhermeg@utad.pt)

This work was supported in part by the National Funds through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT) under Project UIDB/50014/2020 and in part by FCT through the Project entitled Automated Creation of Authentic Virtual Experiences under Grant SFRH/BD/148714/2019.

ABSTRACT Proper evaluation of realism in immersive virtual experiences is crucial to ensure optimisation of resources. This way, we can take better decisions while designing realistic immersive experiences, prioritising factors that have a higher impact on the perceived realism of the virtual experience. This systematic review aims to provide readers with an overview of methodologies used throughout the literature to evaluate realism in immersive virtual, augmented and mixed reality. A total of 79 from 1300 gathered articles met the eligibility criteria and were analysed. Results have shown that virtual reality is by far the platform where realism studies were performed. Head-mounted displays are by far the preferred equipment for such studies. Visual realism is the most researched, followed by audiovisual. The majority of methodologies consisted of subjective, as well as a combination of objective and subjective measures. The most used evaluation instrument is questionnaires where many of which are custom and non-validated. Presence questionnaires are the most used ones and are often used to evaluate the presence, perceived realism and involvement. Cybersickness evaluation is consistently assessed by one self-report questionnaire.

INDEX TERMS Virtual reality, augmented reality, mixed reality, methodologies, realism.

I. INTRODUCTION

The possibilities of realistic virtual experiences are vast and still being explored in several application fields as technology advances, allowing higher realism. Because realism enables a better illusion of the virtual environment as being real, such can lead users to act the same way they would in an analogous situation. In terms of research, previously inconceivable experiments due to constraints in costs and logistics and ethics [1] can now be considered to be done using realistic, immersive experiences (assuming the results are replicable in real-world conditions). It also opens doors for training professionals by giving them real-world experience [2]. The knowledge they gathered in those experiences can be translated to real-world situations, bypassing costs, time, and safety constraints [3]–[6]. Immersive experiences can also help users understand complex data-sets and structures, as proven by C4X Discovery laboratory where new drugs were developed with the help of VR to visualise structures of complex molecules [7]. Several companies are already adopting virtual experiences in the marketing department to showcase the product or service to clients [8] by creating a virtual experience very close to the real one. Tourism is also being targeted in studies to research if users intent on experiencing tourism is increased by previewing those experiences firstly in VR [9]–[11]. Other studies address the potential of this technology to protect cultural heritage by giving users a similar realistic experience as if they were there [12], [13].

There is an evident pursuit for realistic virtual experiences through the years. Since the earlier days, we have dreamt of creating such realistic experiences both in entertainment (Morton Heilig's Sensorama [14]), but also in more serious applications such as training professionals (Edwin Link's trainer [15]). Ivan Sutherland's Ultimate display concept [16] is an example of that ambition. Presented in 1965, it consisted of an experience so real that users could not distinguish it

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

from reality in every aspect. Such experience would be so real that a virtual bullet could harm the human body similarly to reality. Although being impossible today, as it would require a computer able to control matter, we should note that with technological advancements, we can now render much more detailed virtual worlds and stimulate other senses, slowly approaching the extreme of the Ultimate Display concept. These advancements allow us to isolate users from real-world stimuli better while providing coherent and detailed synthesised stimuli, something considered science fiction several years ago. Nevertheless, technological barriers still exist, and it is important to wisely manage available computational resources and human resources to provide the best realistic experience within the given constraints. We need to prioritise certain factors over others so we can better optimise the use of resources depending on the virtual experience context. However, to do such, we first need to understand what can improve the perceived realism under several conditions and that researchers need proper methodologies.

Realism is defined differently between authors, but ultimately its meaning is closely the same across studies - "A virtual experience that meets our expectations of what reality is". Hoorn *et al.* [17] considered that a simulation is realistic enough when it can achieve its goal. Slater *et al.* [18] explore the visual factor of realism, segmenting it into geometric (similarity of the virtual object against its real counterpart) and illumination realism (how close to reality the lighting simulation is). Perroud *et al.* [19] further explored the realism concept in virtual experiences, presenting five realism acceptations:

1) "**Realistic looking:** very detailed shaders and materials, hard work on lights in the scene and other artistic tricks."
2) "**Realistic construction of the virtual world:** what is implemented is based on scientifically proven models (e.g. gravity, dynamics)"
3) "**Physiologic realism:** the inputs received by the body are the same that those it would receive in a real situation, even if overall it seems odd to the observer."
4) "**Psychological realism:** the implementation seems realistic to the observer, even if it is, in fact, over or under-powered (e. g. walking speed, the field of view)"
5) "**Presence:** even if the scene is only made of non-textured polygons, the maximum the presence, the better."

There are also terms closely related to realism, where authors definitions closely resemble the realism ones. Fidelity is defined by Alexander *et al.* [20] as "the extent to which the virtual environment emulates the real world", Franzluebbers and Johnsen [21] considers that "a high-fidelity approach would attempt to replicate most, if not all, of an object's physical properties" and Meyer *et al.* [22] defines it as "a measure of the degree to which a simulation system represents a real-world system". Similarly, authenticity is defined by Gilbert [23] as "whether the virtual environment provides the experience expected by the user, both consciously and unconsciously".

Presence (the subjective feeling of being in the virtual environment [24]) is often found in realism related studies [18], [25]–[27], where higher realism results in a more heightened sense of presence. However, as Bowman and McMahan [27] state, this is not always the case. In specific fields such as the oil and gas industry, a higher level of realism is necessary to visualise complex 3D models correctly, where a sense of presence is not needed.

We can verify that realism, fidelity and authenticity have very similar meanings. In this paper, we consider realism as everything that represents the real-world experience, encompassing several author definitions of realism and related terms.

This systematic review aims to give researchers an overview of used methodologies to evaluate realism related factors in immersive virtual environments (IVEs) using Virtual, Augmented and Mixed Reality. The research question (RQ) considered for this work is: "What are the methodologies being used to study the realism in IVEs"). To the best of our knowledge, no other work reviewed methodologies for realism evaluation in all stimuli in IVEs. It will identify research gaps and limitations and suggest new approaches that can be considered in future works. Readers will gain a better insight into evaluating realism while avoiding known limitations on previous studies or even test new methodologies, improving upon the previous works.

II. METHODOLOGY

We used PRISMA methodology [28] to perform this systematic review, which provides transparency in the document selection phase. By following such a methodology, the study's replicability is ensured. To meet our goals and answer the RQ we have defined inclusion and exclusion criteria which can be found in Table 1.

TABLE 1. Inclusion (IC) and exclusion criteria (EC) considered in document selection.

Criteria	Description
IC0	The title, abstract or keywords match the search query.
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 immersive setup
EC7	Out of scope

The search query considered keywords related to immersive Virtual, Augmented, and Mixed reality, as well as, terms related to realism and methodologies. Papers had to contain one of the following keyword: "virtual reality" OR "augmented reality" OR "mixed reality" OR "VR" OR "AR" OR "MR". They also had to contain one of the keywords related to immersive virtual experiences: "immersive" OR "cave" OR "hmd" OR "head mounted" OR "head mount" OR "head-mounted" OR "head-mount" OR "headset". To filter papers that do not consider realism, one of the following keywords had to be present: "authentic" OR "authenticity" OR "credibility" OR "coherent" OR "credible" OR "believable" OR "coherence" OR "fidelity" OR "aesthetics" OR "realism" OR "realness". Finally, to ensure papers included methodologies to evaluate realism, one of the following keywords was obligatory: "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".

Document search was performed in four well-known electronic databases, considered to be representative of the scope of this paper: IEEE Xplore, Elsevier Scopus, Clarivate World of Science, and ACM Digital Library. Additionally, a manual search was also performed on Google Scholar. The search was performed for documents up until the first trimester of 2020 (April 6th). A total of 1300 papers were recorded (1287 from databases searches and 13 from manual search). After duplicates removal, 810 papers were considered for abstract screening according to the inclusion and exclusion criteria, being that 425 documents were discarded. From the resulting 385 documents eligible for full-text screening, 79 were considered for analysis and data retrieval for this systematic review. A summary of the selection process is illustrated in Figure 1.

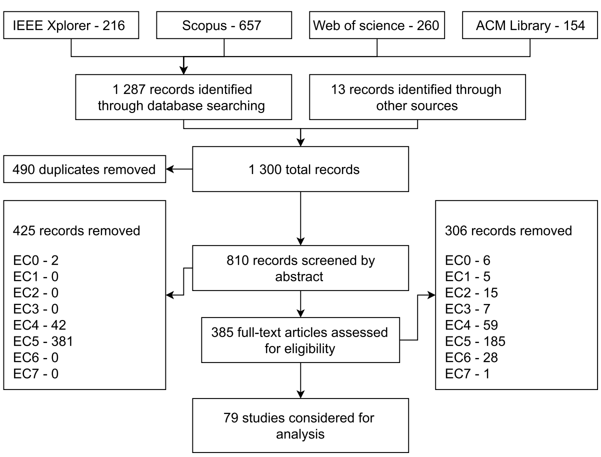


FIGURE 1. Prisma Methodology flow diagram.

A. DATA COLLECTION

From the 79 documents, we have retrieved all data needed to answer the RQ, namely:

- **Immersive Technology:** What type of immersive technology was used (i.e., Virtual Reality, Augmented Reality or Mixed Reality).
- **Immersive System:** The type of equipment used to provide the immersive experience (HMD or CAVEs).
- **Stimuli:** What stimuli were used during the experience.
- **Type of evaluation:** Whether objective or subjective methods were used to gather data.
- **Evaluation Instruments:** What type of instruments were used to gather the data.
- **Variable Group:** What dependent variables were being studied.

Regarding the categorisation of variables, no taxonomies were adequate to be used in this systematic review. To organise data, we have proposed a new taxonomy to categorise every dependent variable found from the data gathering process (table 2).

Based on the virtuality continuum [29], we segmented VR, AR and MR in the following way: VR considers experiences where everything is virtual, AR considers experiences where virtual elements are overlapped on top of the real-world, and everything in-between VR and AR extremities of the virtuality continuum is considered to be MR in this study.

A document quality analysis was performed to provide an overview of the studies methodology and sample size. The analysis was done in the course of the full-text analysis. The scoring system was based on Connoly *et al.* [30], Feng *et al.* [31], and Melo *et al.* [32] works and was performed by two researchers to avoid bias (if there were no consensus, a third reviewer would decide). The quality evaluation consisted of two items rated from 1 (lowest) to 3 (highest), with the final score being the sum of both scores. The scoring was given based on the following metrics:

Sample size: lower than the recommended number get one point, recommended sample sizes get two points, higher than recommended get three points (sample size recommendation based on Macefield *et al.* [33]).

Methodology: based on the instruments, materials, procedures, limitations and reproducibility. In the presence of severe issues, papers get one point. Papers with non-critical issues which do not affect the trustworthiness study get two points. Papers with minor issues get three points.

III. RESULTS AND DISCUSSION

The RQ seeks to identify which methodologies are being used study realism in IVEs. We start with the document quality synthesis followed by a discussion of the results of: Immersive Technology, Immersive System, Stimuli, Type of Evaluation and Evaluation Instruments.

A. DOCUMENT QUALITY SYNTHESIS

The mean quality score from all 79 documents considered for full-text was 4.32 points, with a standard deviation of 1 (Figure 2). Documents with a score equal to or higher than 4 are considered high quality. Thus, 65 documents (82.3%) are considered high-quality and provide an overall good

TABLE 2. Dependent variable Categorisation.

Dependent Categorisation	Description
User Experience (Embodiment)	How user evaluate their self-avatar
User Experience (Perceived Environment Realism)	The users subjective perception of virtual environment realism
User Experience (Task Satisfaction)	The user's subjective task satisfaction.
User Experience (Virtual Agents)	How users perceive other virtual entities
User Experience (Involvement)	User involvement with the virtual experience (e.g. pleasure, stress, engagement, boredom, etc.)
User Experience (User preference)	The subjective preference of users (e.g. prefer better lighting than better physics simulation)
User Experience (Presence)	The sense of presence felt by users [24]
User Performance (Effectiveness/Efficiency)	The effectiveness and efficiency of users when performing tasks [34].
User Behaviour	User behaviour during the virtual experience
Physiologic Responses	The user's physiological response (e.g. heart rate, skin conductance, etc.)

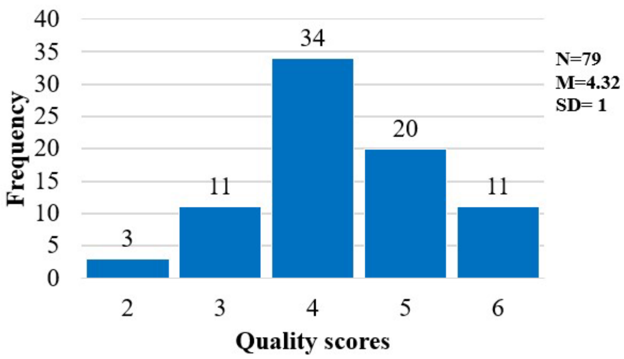


FIGURE 2. Quality assessment scores histogram.

sample count and robust methodology. We conclude that research in this field is in a mature phase where methodologies are well supported by previous studies and sample sizes high, which can be justified by the current growth potential of IVEs in several fields.

B. IMMERSIVE TECHNOLOGY

Immersive Technology is classified as VR, AR and MR. The vast majority of the documents studied realism in VR (74 occurrences - 94%), with the rest being distributed in MR (3 occurrences - 4%) and AR (2 occurrences - 2%) (Fig. 5a). Although each document focused on only one immersive technology (VR, AR or MR), two compared a fully VR experience against an MR one [43], [69]. As such, both documents counted for both VR and MR.

As the data indicates, VR is the main focus of the study of realism. The closer to the real environment in the virtuality continuum, the less control over the virtual experience. In other words, what is already real cannot be further improved as it is already fully representing reality. However, it can be manipulated and downgraded through image processing [72].

Evaluation of realism in AR and MR might work differently than in VR. Users are not totally isolated from real stimuli in AR and MR as it happens with VR. Because the reality is mixed with virtual elements, users might use it as a baseline to compare the virtual stimuli. What could be considered real in VR might be perceived as less real in AR and MR due to this comparison, especially in AR, as most of the stimuli are real. However, the same can happen the other way around as users might rate the experience more real just because they see the real stimuli, even if the virtual stimulus is rated as non-realistic. This raises the complexity of realism evaluation in AR and MR. As such, we suggest the evaluation methodology to be centred around the virtual elements and how they blend in between the real stimuli instead of only evaluating the overall experience realism.

Other reasons that can justify this result are: (a) availability, (b) popularity, (c) price, (d) setup complexity, and (e) the possibility of total control over what to simulate, totally abstracting the users from the real surroundings. The latter is of particular use when one wants to fully recreate an environment that is impossible, dangerous or expensive to replicate *in situ*. For example, a simulator to train firefighters in a building fire is a situation that may be better replicated in VR than MR or AR. MR presents the challenge of introducing seamlessly real objects in an entirely virtual environment, which often require more equipment and video segmentation techniques that may not work well in harsh or otherwise uncontrolled conditions. AR works well under the assumption that the real environment is controlled, and extra virtual elements could be overlaid on top of it. Again, considering this example, AR would not be the best platform, as it would require the environment that exists in the real world. Also, immersive AR equipment, such as HoloLens, have a narrow field of view. Due to how the overlay works, virtual objects cannot wholly occlude the real environment as they will always be transparent. Each platform has specific advantages and disadvantages. However, VR, for its ability to fully replace the real world with a virtual one, is the most flexible platform to work, contributing to its wide adoption.

C. IMMERSIVE SYSTEM

Immersive System is divided into three major groups (Fig 5b): HMD (85%), CAVE (13%) and Stereoscopic Display (2%). HMD is by far the most used immersive system,

TABLE 3. List of questionnaires used and the variable groups they evaluated.

Questionnaire	Variable Group	Freq.	Documents
Standardised Embodiment Questionnaire [35]	Embodiment	1	[36]
Botvinick and Cohen (1998) [37]	Embodiment	1	[38]
PANAS [39]	Embodiment	1	[40]
Partner Perception scale [41]	Embodiment	1	[40]
Virtual Body Ownership Illusion (VBOI) [42]	Embodiment	1	[43]
Illusion of Virtual Body Ownership (IVBO) [44]	Embodiment	1	[45]
Custom	Embodiment	3 (33.3%)	[46]–[49]
Semantic Differential (SD) rating scales [50], [51]	Per. Env. Real.	1	[52]
Presence Questionnaire (PQ) Ver. 3.0 [53]	Per. Env. Real.	1	[54]
SUS questionnaire [55]	Per. Env. Real.	1	[56]
Presence Questionnaire (PQ) [57]	Per. Env. Real.	4	[58]–[61]
Hendrix and Barfield (1996) [62]	Per. Env. Real.	1	[59]
Igroup Presence Questionnaire (IPQ) [63]	Per. Env. Real.	1	[59]
Custom	Per. Env. Real.	15 (62.5%)	[64]–[78]
Self-Assessment Manikin (SAM) [79]	Task Satisfaction	1	[80]
PSSUQ [81]	Task Satisfaction	1	[54]
NASA-TLX [82]	Task Satisfaction	1	[54]
The Borg rating of perceived exertion (RPE) scale [83]	Task Satisfaction	1	[84]
INTUI [85]	Task Satisfaction	1	[86]
Custom	Task Satisfaction	4 (44.4%)	[47], [49], [60], [74]
Rosenberg self-esteem scale [87]	Virtual Agents	1	[88]
Trust [89]	Virtual Agents	1	[45]
Self-Assessment Manikin (SAM) [79]	Virtual Agents	3	[45], [86], [90]
Humanness and Eeriness [91]	Virtual Agents	3	[45]
Custom	Virtual Agents	9 (52.9%)	[40], [74], [88], [92]–[97]
Presence Questionnaire (PQ) [57]	Involvement	1	[58]–[60]
Igroup Presence Questionnaire (IPQ) [63]	Involvement	1	[59]
PANAS [39]	Involvement	1	[98]
STICSA [99]	Involvement	1	[100]
CSAI-2R [101]	Involvement	1	[100]
Immersion IEQ [102]	Involvement	1	[86]
Engagement GEQ [103]	Involvement	1	[86]
E2I's enjoyment scale [104]	Involvement	1	[86]
Custom	Involvement	1	[48], [49], [75], [76], [78]
Custom	User Preference	10 (11.1%)	[36], [60], [67]–[69], [73], [86], [95], [97], [105]
Presence Questionnaire (PQ) [57]	Presence	7	[59], [60], [66], [74], [84], [106], [107]
Presence Questionnaire (PQ) Ver. 3.0 [53]	Presence	1	[54]
Igroup Presence Questionnaire (IPQ) [63]	Presence	10	[59], [108]–[110]
SUS questionnaire [55]	Presence	10	[36], [52], [84], [90], [95], [111]–[113]
Hendrix and Barfield (1996) [62]	Presence	1	[59]
MEC-Spatial Presence Questionnaire (MEC-SPQ) [114]	Presence	1	[108]
ITC-SOPI [115]	Presence	1	[111]
E2I's presence scale [104]	Presence	1	[86]
Social Presence Scale [116]	Presence	1	[40], [45]
Social Presence Survey [117]	Presence	2	[73], [95]
Custom	Presence	9 (17%)	[43], [43], [56], [72], [96], [108], [113], [118], [119]
Custom	User Performance	7 (100%)	[106], [112], [120]–[124]
Simulator Sickness Questionnaire (SSQ) [125]	Physiologic Responses	8	[56], [61], [66], [106], [110], [112], [126]
Motion Sickness Assessment Questionnaire [127]	Physiologic Responses	1	[109]
Custom	Physiologic Responses	1 (10%)	[72]

which can be explained by: (a) VR, AR and MR can be used with HMDs, (b) HMDs are less complex and expensive than CAVEs, (c) HMD are mobile (not requiring a fixed installation).

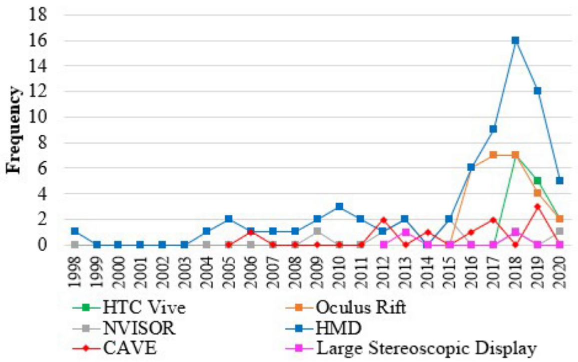


FIGURE 3. Type of immersive system used over the years. The "HMD" line considers both HTC Vive, Oculus Rift, nvisor and other disclosed and non-disclosed HMD.

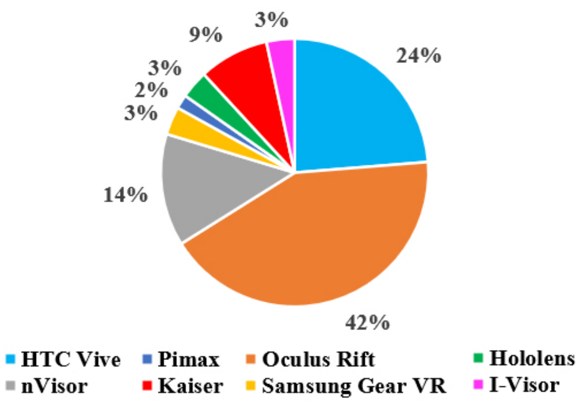


FIGURE 4. Type of HMD used.

Analysing how the use of the immersive systems evolved over the years, one can see that the use of HMD was a little more evident than CAVEs. However, after the year 2016, the use of HMDs over CAVEs highly increased. Market-share data from 2016 [128] shows a huge percentage of HMD's unit shipment was from Google Cardboard (69%) followed by Samsung Gear (17%) with a very small margin for Oculus Rift and HTC Vive (1% each). However this is not reflected in the results, as many studies used HMD from HTC (around 13 documents), Oculus (around 25 documents), nvisor (around 8 documents), or Kaiser (around 5 documents) (Figure 4). However, two documents were found to use Samsung Gear, a smartphone based mobile VR platform. Note that not every study stipulated which HMD was being used. Based on this data, we speculate that mobile solutions might not be sufficiently robust to perform this type of research due to their limited computational capabilities (e.g, lack of performance, lack of screen quality, or possible tracking issues), making the newer Oculus Rift and HTC Vive preferred platforms. The drop seen in 2020 (Fig. 3) is explained by the fact that the search was performed at the end of the first trimester of 2020.

D. STIMULI CONSIDERED

Figure 5c depicts how frequently each combination of stimuli was present between documents. The wide majority (53%) featured only visual stimuli. Around 35% used a bi-modal approach where 23% added audio, and 12% added haptic in conjunction with visual. Only 11% used a combination of three stimuli (visual, audio and haptic). Finally, only 1 document (< 1%) considered four modalities (visual, audio, haptic, and scent). Curiously, no documents were found featuring Visual+Scent. Our results corroborate the review of Melo *et al.* [32] on multisensory VR, where haptic seems to be the most used stimulus beyond visual and audio and where scent and particularly taste are a lot less researched due to being more complex to implement. Please refer to his review for a more focused and in-depth discussion about multisensory stimuli in both immersive and non-immersive VR setups.

We consider the visual stimulus the obligatory stimuli for a given system to be considered immersive. HMD, CAVE, and large stereoscopic displays all work by deceiving the visual sense. The bi-modal approach of visual and audio is present in less than a quarter of the documents. The research team expected a higher usage percentage of this particular bi-modal approach due to being easy to implement due to the wide availability of technology. Although there are complex and expensive audio setups (such as using a speaker array in an anechoic chamber [65]), headphones are relatively easy and straightforward to use. Also, some studies required users to speak out loud or talk with researchers, being that in these cases the use of headphones or audio could interfere with the communication [129], [130]. Another reason could be that researchers focused more on studying visual realism than audio, resulting in more documents considering only visual stimuli. Such could be because vision is the most important of all our senses and its importance and complexity create a bias towards vision stimulus research. Hutmacher [131] argues that this is debatable and proposes two additional explanations to why vision is the most researched sense: (a) today's technology is better suited for studying vision than other senses and (b) due to cultural reasons as we live in a visual society which places a higher value on vision than other senses. This is reflected in the data, as there were many different variables tested regarding the visual stimulus (discussed ahead). Then again, the fact that visual alone is the most explored sense could also be because IVEs rely heavily on vision.

Haptic stimuli include both passive and active. The use of haptic may not be as easy to implement as audio, sometimes requiring specific equipment and a precise tracking system (particularly for active haptics) that may not be feasible in specific contexts. Such may explain in part why it is present in only 23% of the documents. Passive haptics consists of real objects with the same shape and placed in the same place as a virtual object. When users interact with virtual objects and feel that something is really there, then the illusion that the virtual experience is real might be higher than when users touch virtual object to see their body part go right through them. Regarding active haptics, the most used ones consisted of vibration or force-feedback. Stimuli such as Wind and Thermal were only found in 2 and 1 studies, respectively. Nevertheless, active haptics setups might be highly complex as they ranged from small vibrotactile actuators [132] to torque-controlled robot arms [75], motion platform simulating the motion of vehicles [66], to prototypes synthesising texture haptic feeling [133] and mid-air haptic devices [38].

The Scent is by far the least studied stimulus (in the context of realism studies in IVE), being found in one document together with visual, audio and haptic [110]. No studies were found that focused on how scent alone influences the realism of an IVE. Some reasons may be that (a) scent is not a critical stimulus in several contexts. In experiences/simulations such as production line assembly or teleconferences meetings, Scent may not play a critical role in the experience and the user to be successful. Therefore, Scent may be discarded as its presence is not worth the added apparatus complexity. However, in some contexts such as firefighter training or gas leak detection, Scent may be such a critical stimulus that its absence may render the simulation incomplete. Another reason may be due to (b) the increase of complexity and expenses through specialised equipment to capture and/or deliver Scent may not justify it in contexts where scent is not required to fulfil the experience's purpose. Nevertheless a clear gap exists here, as scent should be more researched in IVE to fully understand its impact on realism in both contexts where it is critical and non-critical.

E. TYPE OF EVALUATION

Overall, 44% of the documents considered both Subjective and Objective metrics, 41% considered only Subjective metrics and the 15% left only considered objective ones (Figure 5d). The analysed papers show a clear preference for subjective metrics, being present in 85% of works. We could speculate that some of the reasons could be due to (a) how the user perceives realism and how it impacts their experience is something that is intrinsic to the mind and how it understands the world. IVEs are all about deceiving the brain by isolating the user from the real world stimuli and providing synthesised ones. Therefore the subjective evaluation is expected to be present in this type of studies. Another reason may be due to (b) researchers trying to find correlations between subjective metrics and objective metrics and, for example, assessing one's notion of how realistic was the experience and then understanding how it affected their performance (objective). If replicable, one could then use objective metrics to assess the level of realism users felt. It may also be due to (c) producing more reliable results by using subjective and objective metrics to back up each other and identify possible subjective bias or limitations in the methodology. Such cases could justify the majority of studies applying both subjective and objective metrics. Finally, it could also be due to (d) the existence of more validated subjective metrics than objective, leading researchers to use them over the objective ones, or use both and try to close such gap.

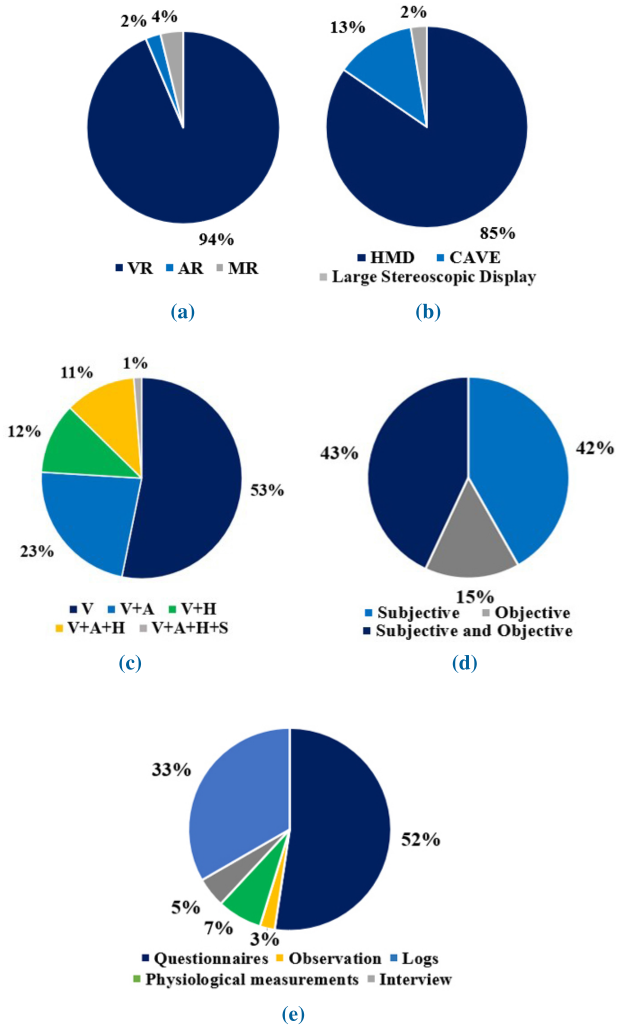


FIGURE 5. **(a) Immersive Technology. (b) Immersive System. (c) Stimuli (V-Visual, A-Audio, H-Haptic, S-Scent). (d) Type of evaluation. (e) Evaluation Instruments.**

F. EVALUATION INSTRUMENTS

The evaluation instruments were categorised in Questionnaires (55%), Logs (32%) - data automatically logged by the system or manually noted by the researcher, Interview (5%), Observation (2%) - the researcher observing the participants, and Physiological measurements (6%) - data gathered through physiologic equipment.

As figure 5e shows, Questionnaires are the most used instrument to gather data, which could result in both subjective and objective data (e.g. measuring participant's performance through a test). Table 3 shows the most used questionnaires and what dependent factors they evaluated.

We should also note that many authors used custom or adapted versions of validated questionnaires. The high number of custom questionnaires addressing the perceived environmental realism and the need to use subscales of validated presence questionnaires might suggest the lack of a validated realism focused questionnaire. We also note that custom non-validated questionnaires should be used with caution, as they might not evaluate what researchers expect them to evaluate. We can verify in Table 3 that custom questionnaires are more frequent in some variable groups. This may suggest that (a) there is no validated questionnaire capable of evaluating what researchers are trying to study, (b) the questionnaire is not fit to be used within the chosen methodology (e.g. too long to be used between or even within the virtual experience itself) or (c) what is being evaluated is too specific to be covered by a validated questionnaire (such as user performance and user preference questionnaires). The data also shows that some variable groups present a more consistent use of validated questionnaires. On perceived realism, Presence Questionaire (PQ) was used consistently. On virtual agents evaluation, both the Self-Assessment Manikin (SAM) and Human and Eeriness questionnaires were persistently used. Presence evaluation was frequently evaluated by essentially three questionnaires: Presence Questionaire (PQ), Igroup Presence Questionnaires (IPQ) and SUS Questionnaire. However, there is an ongoing discussion about how presence questionnaires are not robust enough to evaluate such an subjective metric [134], [135], suggesting that we need to move away from questionnaires and adopt other methodologies. Regarding cybersickness, Simulator Sickness Questionnaire (SSQ) was shown to be the most consistently used questionnaire to address this physiologic response. The fact that this questionnaire was published in 1993 and is still vastly used today indicates it's robustness to evaluate such symptoms.

As these type of studies revolve around computers, logs also can be easily setup, which may justify why it is the second most used type of evaluation instrument. They allow recording multiple variables in a precise, unbiased and objective way.

Interviews are less used in comparison with the former instruments. They allow open-ended questions with the possibility of adaptation in real-time. Because there may be several variables that can change how users perceive the IVE as real, little details might be missed through other evaluation instruments, where interviews can provide valuable insight.

Physiological measurements are present in 6% of the works. Overall, physiological measurements consisted of heart rate, skin conductance, and electrocardiograms. The research team was expecting a higher percentage because physiologic responses are 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 it may not be easy to set some physiological measurements, as they require proper equipment and proper analysis and may be intrusive during the IVE. The fact that users might 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. Also, due to the HMD apparatus covering the head, an electroencephalogram, for example, might prove difficult to correctly setup. However, it may present a powerful instrument to evaluate realism, and it is unbiased. Although objective, it is open to several interpretations of what the results mean. Lastly, observation is the least used. It may be because it can be highly subjective and prone to bias. However, it can still provide helpful anecdotal data regarding users behaviour or other events.

IV. CONCLUSION

Results indicated that the vast majority of studies are being conducted using VR in detriment to AR and MR, which leads us to suggest further work on realism on AR and MR based systems. More than 4/5 of the studies used HMD (85%) instead of CAVE or Stereoscopic displays. The most used stimuli were visual alone, followed by bi-modal approaches such as visual+audio and to a minor degree visual+haptic and tri-modal visual+audio+haptic. Only one document explored the effects of 4 simultaneous stimuli on realism, and none studied scent individually. We highly suggest more research on how the increase of realism provoked by multisensory stimulation, and individual stimulus (especially scent) in IVEs can affect the user experience.

The majority of studies used subjective metrics or a combination of subjective and objective. Because the perception of realism can be highly subjective and therefore biased, we also recommend using objective metrics (such as physiological data) to support the subjective data.

More than half of the evaluations consisted of questionnaires, followed by logs, interviews, observation and physiologic measurement equipment. The questionnaires used were synthesised in one table, providing researchers information about which questionnaires might be more appropriate for their future studies. The most frequent ones are presence questionnaires, which can also address realism and involvement metrics.To address cybersickness, authors usually use one questionnaire, the simulator sickness questionnaires (SSQ). It was also identified the non existence of a standard questionnaire to evaluate realism, as several authors opted to use subscales of presence questionnaires or custom questionnaires to address it. We take note and advise the use of validated questionnaires instead of custom non-validated ones. Non-validated questionnaires might not evaluate what researchers expect them to evaluate. We also suggest more research on physiologic measures as they are objective/non-biased and were one of the least used metrics.

This work provides researchers in the field with insightful knowledge about how realism is measured. Such will help researchers better select their methodologies, considering their advantages and disadvantages and best practices depending on their study design, thus increasing the robustness and validity of their studies. The gaps found will also help researchers to explore new methodologies to evaluate realism. Researchers will also be aware of which immersive setups realism studies are taking place and help the body of knowledge in the lesser explored ones.

V. LIMITATIONS

We wanted the immersive visual system to be consistent, and therefore studies that only performed comparisons between different systems, such as HMD and CAVEs, were not included.

The search query aimed at including realism related studies in IVEs. Still, some documents might have been left out due to authors testing variables indirectly associated with the IVE realism, or even directly but without mentioning realism as the study scope.

Some variables were ambiguous, as in, they could be categorised in more than one category (e.g. immersion could be linked to both User Experience (Involvement) and User Experience (Presence)). Depending on the study context and scope, and author definition (when given), the research team categorised these variables where they were best fit.

REFERENCES

[1] M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, N. Pistrang, and M. V. Sanchez-Vives, "A virtual reprise of the Stanley Milgram obedience experiments," *PLoS ONE*, vol. 1, no. 1, p. e39, Dec. 2006.

[2] (Aug. 2020). *PIXO VR Home*. [Online]. Available: https://pixovr.com/

[3] P. Monteiro, M. Melo, A. Valente, J. Vasconcelos-Raposo, and M. Bessa, "Delivering critical stimuli for decision making in VR training: Evaluation study of a firefighter training scenario," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 2, pp. 65–74, Apr. 2021.

[4] D. Narciso, M. Melo, S. Rodrigues, J. P. S. Cunha, and M. Bessa, "Impact of different stimuli on user stress during a virtual firefighting training exercise," in *Proc. IEEE 20th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2020, pp. 813–818.

[5] R. Eiris, M. Gheisari, and B. Esmaeili, "Desktop-based safety training using 360-degree panorama and static virtual reality techniques: A comparative experimental study," *Autom. Construct.*, vol. 109, Jan. 2020, Art. no. 102969.

[6] S. Joshi, M. Hamilton, R. Warren, D. Faucett, W. Tian, Y. Wang, and J. Ma, "Implementing virtual reality technology for safety training in the precast/prestressed concrete industry," *Appl. Ergonom.*, vol. 90, Jan. 2021, Art. no. 103286.

[7] D. Sage. (Jun. 2018). *Virtual Reality is Helping Chemists Discover New Life-Enhancing Drugs*. [Online]. Available: https://www.c4xdiscovery.com/news/in-the-news/virtual-reality-is-helping-chemists-discover-new-life-enhancing-drugs.html

[8] K. Sloan. (Dec. 2019). *Air Canada Uses VR to Showcase Its Top-Flight Business Class Service*. [Online]. Available: https://www.unrealengine.com/en-US/spotlights/air-canada-uses-vr-to-showcase-its-top-flight-business-class-service

[9] C. Flavián, S. Ibáñez-Sánchez, and C. Orús, "Impacts of technological embodiment through virtual reality on potential guests' emotions and engagement," *J. Hospitality Marketing Manage.*, vol. 30, no. 1, pp. 1–20, Jan. 2021.

[10] W. H. Lo and K. L. B. Cheng, "Does virtual reality attract visitors? The mediating effect of presence on consumer response in virtual reality tourism advertising," *Inf. Technol. Tourism*, vol. 22, no. 4, pp. 537–562, Dec. 2020.

[11] Y. C. Huang, K. F. Backman, S. J. Backman, and L. L. Chang, "Exploring the implications of virtual reality technology in tourism marketing: An integrated research framework," *Int. J. Tourism Res.*, vol. 18, no. 2, pp. 116–128, Mar. 2016.

[12] N. Losada, F. Jorge, M. S. Teixeira, M. Melo, and M. Bessa, "Could virtual reality substitute the 'real' experience? Evidence from a unesco world heritage site in northern Portugal," in *Proc. Int. Conf. Tourism, Technol. Syst.* Singapore: Springer, 2020, pp. 153–161.

[13] G. Gonçalves, M. Melo, J. Vasconcelos-Raposo, and M. Bessa, "A novel method to enhance the touristic 360° promotional video experience," *Multimedia Tools Appl.*, vol. 79, nos. 31–32, pp. 22905–22927, Aug. 2020.

[14] M. L. Heilig, "Sensorama simulator," U.S. Patent 3 050 870 A, Aug. 28, 1962.

[15] J. E. A. Link, "Trainer for aviators," U.S. Patent 2 099 857 A, Nov. 23, 1937.

[16] I. E. Sutherland and I. E. Sutherland, "The ultimate display," in *Proc. IFIP Congr.*, vol. 2, 1965, pp. 506–508.

[17] J. F. Hoorn, E. A. Konijn, and G. C. Van der Veer, "Virtual reality: Do not augment realism, augment relevance," *Hum.-Comput. Interact., Overcoming Barriers*, vol. 4, no. 1, pp. 18–26, 2003.

[18] M. Slater, P. Khanna, J. Mortensen, and I. Yu, "Visual realism enhances realistic response in an immersive virtual environment," *IEEE Comput. Graph. Appl.*, vol. 29, no. 3, pp. 76–84, May/Jun. 2009.

[19] B. Perroud, S. Régnier, A. Kemeny, and F. Mérienne, "Model of realism score for immersive VR systems," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 61, pp. 238–251, Feb. 2019.

[20] A. L. Alexander, T. Brunyé, J. Sidman, and S. A. Weil, "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," Group, Sci. Article, Aptima, Woburn, MA, USA, 2005.

[21] A. Franzluebbers and K. Johnsen, "Performance benefits of high-fidelity passive haptic feedback in virtual reality training," in *Proc. Symp. Spatial User Interact. (SUI)*, Oct. 2018, pp. 16–24.

[22] G. F. Meyer, L. T. Wong, E. Timson, P. Perfect, and M. D. White, "Objective fidelity evaluation in multisensory virtual environments: Auditory cue fidelity in flight simulation," *PLoS ONE*, vol. 7, no. 9, Sep. 2012, Art. no. e44381.

[23] S. B. Gilbert, "Perceived realism of virtual environments depends on authenticity," *Presence, Teleoperators Virtual Environ.*, vol. 24, pp. 322–324, Dec. 2016.

[24] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 6, no. 6, pp. 603–616, Dec. 1997.

[25] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater, "Visual realism enhances realistic response in an immersive virtual environment—Part 2," *IEEE Comput. Graph. Appl.*, vol. 32, no. 6, pp. 36–45, Nov. 2012.

[26] J. Hvass, O. Larsen, K. Vendelbo, N. Nilsson, R. Nordahl, and S. Serafin, "Visual realism and presence in a virtual reality game," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jun. 2017, pp. 1–4.

[27] D. A. Bowman and R. P. Mcmahan, "Virtual reality: How much immersion is enough?" *Computer*, vol. 40, no. 7, pp. 36–43, Jul. 2007.

[28] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Med.*, vol. 6, pp. 1–6, Jul. 2009.

[29] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Trans. Inf. Syst.*, vol. 77, no. 12, pp. 1321–1329, 1994.

[30] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Comput. Educ.*, vol. 59, no. 2, pp. 661–686, Sep. 2012.

[31] Z. Feng, V. A. González, R. Amor, R. Lovreglio, and G. Cabrera-Guerrero, "Immersive virtual reality serious games for evacuation training and research: A systematic literature review," *Comput. Educ.*, vol. 127, pp. 252–266, Dec. 2018.

[32] M. Melo, G. Goncalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa, "Do multisensory stimuli benefit the virtual reality experience? A systematic review," *IEEE Trans. Vis. Comput. Graphics*, early access, Jul. 17, 2020, doi: 10.1109/TVCG.2020.3010088.

[33] R. Macefield, C. Close, and W. W. Nu, "How to specify the participant group size for usability studies: A practitioner's guide," *J. Usability Stud.*, vol. 5, no. 1, pp. 34–45, 2009.

[34] *Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts*, ISO 9241-11:2018(en), 2018.

[35] M. Gonzalez-Franco and T. C. Peck, "Avatar embodiment. Towards a standardized questionnaire," *Frontiers Robot. AI*, vol. 5, p. 74, Jun. 2018.

[36] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet, "Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 2062–2072, May 2020.

[37] M. Botvinick and J. Cohen, "Rubber hands 'feel' touch that eyes see," *Nature*, vol. 391, no. 6669, p. 756, 1998.

[38] D. Pittera, E. Gatti, and M. Obrist, "I'm sensing in the rain: Spatial incongruity in visual-tactile mid-air stimulation can elicit ownership in VR users," in *Proc. CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–15.

[39] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Pers. Social Psychol.*, vol. 54, no. 6, p. 1063, 1988.

[40] D. Krum, S.-H. Kang, and T. Phan, "Influences on the elicitation of interpersonal space with virtual humans," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, Mar. 2018, pp. 223–229.

[41] S.-H. Kang and J. Gratch, "Exploring users' social responses to computer counseling interviewers' behavior," *Comput. Hum. Behav.*, vol. 34, pp. 120–130, May 2014.

[42] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, J. T. McConville, and I. Tebbets, "Anthropometric survey of us personnel: Summary statistics interim report," US Army Natick Res., Natick, MA, USA, Tech. Rep. NATICK/TR-89/027, 1989.

[43] S. Jung, G. Bruder, P. J. Wisniewski, C. Sandor, and C. E. Hughes, "Over my hand: Using a personalized hand in VR to improve object size estimation, body ownership, and presence," in *Proc. Symp. Spatial User Interact.*, Oct. 2018, pp. 60–68.

[44] D. Roth, J.-L. Lugrin, M. E. Latoschik, and S. Huber, "Alpha IVBO–construction of a scale to measure the illusion of virtual body ownership," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, May 2017, pp. 2875–2883.

[45] M. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch, "The effect of avatar realism in immersive social virtual realities," in *Proc. 23rd ACM Symp. Virtual Reality Softw. Technol.* New York, NY, USA: Association for Computing Machinery, 2017, pp. 1–10.

[46] E. Ebrahimi, L. S. Hartman, A. Robb, C. C. Pagano, and S. V. Babu, "Investigating the effects of anthropomorphic fidelity of self-avatars on near field depth perception in immersive virtual environments," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2018, pp. 1–8.

[47] D. Medeiros, J. Pereira, R. Dos Anjos, A. Raposo, D. Mendes, and J. Jorge, "Keep my head on my shoulders! Why third-person is bad for navigation in VR," in *Proc. 24th ACM Symp. Virtual Reality Softw. Technol.* New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–10.

[48] A. Maselli and M. Slater, "The building blocks of the full body ownership illusion," *Frontiers Hum. Neurosci.*, vol. 7, 2013, doi: 10.3389/fnhum.2013.00083.

[49] A. Pusch, O. Martin, and S. Coquillart, "Effects of hand feedback fidelity on near space pointing performance and user acceptance," in *Proc. IEEE Int. Symp. VR Innov.*, Singapore, Mar. 2011, pp. 97–102.

[50] J. Flynn, "A study of lighting as a system of spatial cues," in *Proc. EDRA Workshop Psychol. Potential Illumination*. Lawrence, KS, USA: Univ. Kansas, 1975.

[51] J. E. Flynn, "A study of subjective responses to low energy and nonuniform lighting systems," 1977.

[52] K. Mania and A. Robinson, "An experimental exploration of the relationship between subjective impressions of illumination and physical fidelity," *Comput. Graph.*, vol. 29, no. 1, pp. 49–56, Feb. 2005.

[53] B. G. Witmer, C. J. Jerome, and M. J. Singer, "The factor structure of the presence questionnaire," *Presence, Teleoperators Virtual Environ.*, vol. 14, no. 3, pp. 298–312, Jun. 2005.

[54] J. Bertrand, A. Bhargava, K. C. Madathil, A. Gramopadhye, and S. V. Babu, "The effects of presentation method and simulation fidelity on psychomotor education in a bimanual metrology training simulation," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2017, pp. 59–68.

[55] M. Slater, J. McCarthy, and F. Maringelli, "The influence of body movement on subjective presence in virtual environments," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 40, no. 3, pp. 469–477, Sep. 1998.

[56] K. J. Blom and S. Beckhaus, "Virtual travel collisions: Response method influences perceived realism of virtual environments," *ACM Trans. Appl. Perception*, vol. 10, no. 4, pp. 1–19, Oct. 2013.

[57] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence, Teleoperators Virtual Environ.*, vol. 7, no. 3, pp. 225–240, Jun. 1998.

[58] N. Ranasinghe, P. Jain, S. Karwita, D. Tolley, and E. Y.-L. Do, "Ambiotherm: Enhancing sense of presence in virtual reality by simulating real-world environmental conditions," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 1731–1742.

[59] A. C. Kern and W. Ellermeier, "Audio in VR: Effects of a soundscape and movement-triggered step sounds on presence," *Frontiers Robot. AI*, vol. 7, pp. 1–13, Feb. 2020.

[60] T. Muender, A. Reinschluessel, S. Drewes, D. Wenig, T. Döring, and R. Malaka, "Does it feel real? Using tangibles with different fidelities to build and explore scenes in virtual reality," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–12.

[61] C. Di Loreto, J.-R. Chardonnet, J. Ryard, and A. Rousseau, "WoaH: A virtual reality work-at-height simulator," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2018, pp. 281–288.

[62] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 5, no. 3, pp. 290–301, Jan. 1996.

[63] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence, Teleoperators Virtual Environ.*, vol. 10, no. 3, pp. 266–281, Jun. 2001.

[64] Z. S. See, L. Goodman, C. Hight, M. S. Sunar, A. Dey, Y. K. Ng, and M. Billinghurst, "Creating high fidelity 360° virtual reality with high dynamic range spherical panorama images," *Virtual Creativity*, vol. 9, no. 1, pp. 73–109, Dec. 2019.

[65] J. Y. Hong, B. Lam, Z.-T. Ong, K. Ooi, W.-S. Gan, J. Kang, J. Feng, and S.-T. Tan, "Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications," *Building Environ.*, vol. 149, pp. 1–14, Feb. 2019.

[66] D. Yeo, G. Kim, and S. Kim, "Toward immersive self-driving simulations: Reports from a user study across six platforms," in *Proc. CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–12.

[67] J. Thorn, R. Pizarro, B. Spanlang, P. Bermell-Garcia, and M. Gonzalez-Franco, "Assessing 3D scan quality through paired-comparisons psychophysics," in *Proc. 24th ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 147–151.

[68] S. Lee, T. Chen, J. Kim, G. Kim, S. Han, and Z.-G. Pan, "Affective property evaluation of virtual product designs," in *Proc. IEEE Virtual Reality*, C. J. Y. Ikei and M. Gobel, Ed., Chicago, IL, USA, Mar. 2004, pp. 207–214.

[69] S. Lee, T. Chen, J. Kim, G. J. Kim, S. H. Han, and Z. Pan, "Effects of tactile augmentation and self-body visualization on affective property evaluation of virtual mobile phone designs," *Presence, Teleoperators Virtual Environ.*, vol. 16, no. 1, pp. 45–64, Feb. 2007.

[70] M. Weier, T. Roth, E. Kruijff, A. Hinkenjann, A. Pérard-Gayot, P. Slusallek, and Y. Li, "Foveated real-time ray tracing for head-mounted displays," *Comput. Graph. Forum*, vol. 35, no. 7, pp. 289–298, Oct. 2016.

[71] R. Konrad, A. Angelopoulos, and G. Wetzstein, "Gaze-contingent ocular parallax rendering for virtual reality," *ACM Trans. Graph.*, vol. 39, no. 2, pp. 1–12, Apr. 2020.

[72] K. Vaziri, P. Liu, S. Aseeri, and V. Interrante, "Impact of visual and experiential realism on distance perception in VR using a custom video see-through system," in *Proc. ACM Symp. Appl. Perception*, 2017, pp. 1–8.

[73] J. Wendt, B. Weyers, J. Stienen, A. Bönsch, M. Vorländer, and T. Kuhlen, "Influence of directivity on the perception of embodied conversational agents' speech," in *Proc. 19th ACM Int. Conf. Intell. Virtual Agents*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 130–132.

[74] M. Kyriakou, X. Pan, and Y. Chrysanthou, "Interaction with virtual crowd in immersive and semi-immersive virtual reality systems," *Comput. Animation Virtual Worlds*, vol. 28, no. 5, p. e1729, Sep. 2017.

[75] M. Sagardia and T. Hulin, "Multimodal evaluation of the differences between real and virtual assemblies," *IEEE Trans. Haptics*, vol. 11, no. 1, pp. 107–118, Jan. 2018.

[76] J. Y. Jeon and H. I. Jo, "Three-dimensional virtual reality-based subjective evaluation of road traffic noise heard in urban high-rise residential buildings," *Building Environ.*, vol. 148, pp. 468–477, Jan. 2019.

[77] H. G. Hoffman, "Physically touching virtual objects using tactile augmentation enhances the realism of virtual environments," in *Proc. IEEE Virtual Reality Annu. Int. Symp.*, Atlanta, GA, USA, Mar. 1998, pp. 59–63.

[78] X. Min, W. Zhang, S. Sun, N. Zhao, S. Tang, and Y. Zhuang, "VPModel: high-fidelity product simulation in a virtual-physical environment," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 11, pp. 3083–3093, Nov. 2019.

[79] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[80] X. Zhou and P.-L.-P. Rau, "Determining fidelity of mixed prototypes: Effect of media and physical interaction," *Appl. Ergonom.*, vol. 80, pp. 111–118, Oct. 2019.

[81] J. R. Lewis, "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," *Int. J. Hum.-Comput. Interact.*, vol. 7, no. 1, pp. 57–78, Jan. 1995.

[82] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Advances in Psychology*, vol. 52. Amsterdam, The Netherlands: Elsevier, 1988, pp. 139–183.

[83] N. Williams, "The borg rating of perceived exertion (RPE) scale," *Occupational Med.*, vol. 67, no. 5, pp. 404–405, Jul. 2017.

[84] R. Joyce and S. Robinson, "Passive haptics to enhance virtual reality simulations," in *Proc. AIAA*, 2017, p. 1313.

[85] D. Ullrich and S. Diefenbach, "INTUI. Exploring the facets of intuitive interaction," in *Proc. Mensch Comput., Interaktive Kulturen*, 2010, pp. 251–260.

[86] K. Rogers, J. Funke, J. Frommel, S. Stamm, and M. Weber, "Exploring interaction fidelity in virtual reality: Object manipulation and whole-body movements," in *Proc. CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–14.

[87] M. Rosenberg, "Rosenberg self-esteem scale (RSE)," *Acceptance Commitment Therapy Measures Package*, vol. 61, no. 52, p. 18, 1965.

[88] G. Gorisse, O. Christmann, S. Houzangbe, and S. Richir, "From robot to virtual doppelganger: Impact of avatar visual fidelity and self-esteem on perceived attractiveness," in *Proc. Int. Conf. Adv. Visual Interfaces*, L. F. M. M. T. Catarci and A. Marrella, Eds. New York, NY, USA: Association for Computing Machinery, 2018.

[89] K.-T. Chun and J. B. Campbell, "Dimensionality of the rotter interpersonal trust scale," *Psychol. Rep.*, vol. 35, no. 3, pp. 1059–1070, Dec. 1974.

[90] J. Hvass, O. Larsen, K. Vendelbo, N. Nilsson, R. Nordahl, and S. Serafin, "Visual realism and presence in a virtual reality game," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jun. 2017, pp. 1–4.

[91] C.-C. Ho, K. F. MacDorman, and Z. A. D. D. Pramono, "Human emotion and the uncanny valley: A GLM, MDS, and isomap analysis of robot video ratings," in *Proc. 3rd Int. Conf. Hum. Robot Interact. (HRI)*, 2008, pp. 169–176.

[92] R. E. Guadagno, K. R. Swinth, and J. Blascovich, "Social evaluations of embodied agents and avatars," *Comput. Hum. Behav.*, vol. 27, no. 6, pp. 2380–2385, Nov. 2011.

[93] F. Boucaud, Q. Tafiani, C. Pelachaud, and I. Thouvenin, "Social touch in human-agent interactions in an immersive virtual environment," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 129–136.

[94] K. Zibrek, E. Kokkinara, and R. McDonnell, "Don't stand so close to me: Investigating the effect of control on the appeal of virtual humans using immersion and a proximity-based behavioral task," in *Proc. ACM Symp. Appl. Perception*, 2017, pp. 1–11.

[95] A. Bonsch, T. Vierjahn, and T. W. Kuhlen, "Evaluation of approaching-strategies of temporarily required virtual assistants in immersive environments," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Mar. 2017, pp. 69–72.

[96] T. Randhavane, A. Bera, and D. Manocha, "F2FCrowds: Planning agent movements to enable face-to-face interactions," *Presence, Teleoperators Virtual Environ.*, vol. 26, no. 2, pp. 228–246, May 2017.

[97] H. Salzmann and B. Froehlich, "The two-user seating buck: Enabling face-to-face discussions of novel car interface concepts," in *Proc. IEEE Virtual Reality Conf.*, Reno, NV, USA, Mar. 2008, pp. 75–82.

[98] H. Zou, N. Li, and L. Cao, "Emotional response–based approach for assessing the sense of presence of subjects in virtual building evacuation studies," *J. Comput. Civil Eng.*, vol. 31, no. 5, Sep. 2017, Art. no. 04017028.

[99] M. J. Ree, D. French, C. MacLeod, and V. Locke, "Distinguishing cognitive and somatic dimensions of state and trait anxiety: Development and validation of the state-trait inventory for cognitive and somatic anxiety (STICSA)," *Behav. Cogn. Psychotherapy*, vol. 36, no. 3, p. 313, 2008.

[100] C. Stinson and D. A. Bowman, "Feasibility of training athletes for high-pressure situations using virtual reality," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 4, pp. 606–615, Apr. 2014.

[101] R. H. Cox, M. P. Martens, and W. D. Russell, "Measuring anxiety in athletics: The revised competitive state anxiety inventory–2," *J. Sport Exercise Psychol.*, vol. 25, no. 4, pp. 519–533, Dec. 2003.

[102] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *Int. J. Hum.-Comput. Stud.*, vol. 66, no. 9, pp. 641–661, Sep. 2008.

[103] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the game engagement questionnaire: A measure of engagement in video game-playing," *J. Exp. Social Psychol.*, vol. 45, no. 4, pp. 624–634, Jul. 2009.

[104] J. J.-W. Lin, H. B. L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness, "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment," in *Proc. IEEE Virtual Reality*, Mar. 2002, pp. 164–171.

[105] H. Tian, C. Wang, D. Manocha, and X. Zhang, "Realtime hand-object interaction using learned grasp space for virtual environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 8, pp. 2623–2635, Aug. 2019.

[106] J. Harman, R. Brown, D. Johnson, and S. Turkay, "The role of visual detail during situated memory recall within a virtual reality environment," in *Proc. 31st Austral. Conf. Hum.-Comput.-Interact.*, Dec. 2019, pp. 138–148.

[107] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer, "The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment," *PLoS ONE*, vol. 13, no. 2, Feb. 2018, Art. no. e0191846.

[108] D. Gall and M. E. Latoschik, "The effect of haptic prediction accuracy on presence," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2018, pp. 73–80.

[109] M. AlQassab, A. Gomes, M. Karam, D. Wang, Z. Du, O. Bruckman, and R. Bustos, "The impact of tactile sensations on virtual reality impairment," in *Proc. Int. Conf. Universal Access Hum.-Comput. Interact.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9738, 2016, pp. 231–240.

[110] G. Gonçalves, M. Melo, J. Vasconcelos-Raposo, and M. Bessa, "Impact of different sensory stimuli on presence in credible virtual environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 11, pp. 3231–3240, Nov. 2020.

[111] A. Brogni, V. Vinayagamoorthy, A. Steed, and M. Slater, "Variations in physiological responses of participants during different stages of an immersive virtual environment experiment," in *Proc. ACM Symp. Virtual Reality Softw. Technol. (VRST)*, Limassol, Cyprus, 2006, pp. 376–382.

[112] K. Mania, D. Wooldridge, M. Coxon, and A. Robinson, "The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 3, pp. 396–404, May 2006.

[113] C. Kyrlitsias and D. Michael-Grigoriou, "Asch conformity experiment using immersive virtual reality," *Comput. Animation Virtual Worlds*, vol. 29, no. 5, p. e1804, 2018.

[114] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, F. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, C. Klimmt, J. Laarni, N. Ravaja, A. Sacau, T. Baumgartner, and P. Jäncke, "MEC spatial presence questionnaire (MEC-SPQ): Short documentation and instructions for application," Rep. Eur. Community, Project Presence MEC (IST-2001-37661), 2004.

[115] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, "A cross-media presence questionnaire: The ITC-sense of presence inventory," *Presence, Teleoperators Virtual Environ.*, vol. 10, no. 3, pp. 282–297, Jun. 2001.

[116] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 12, no. 5, pp. 481–494, Oct. 2003.

[117] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Equilibrium theory revisited: Mutual gaze and personal space in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 10, no. 6, pp. 583–598, Dec. 2001.

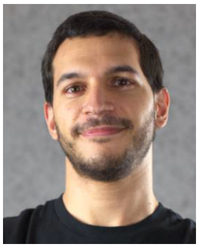
[118] M. Slater, P. Khanna, J. Mortensen, and I. Yu, "Visual realism enhances realistic response in an immersive virtual environment," *IEEE Comput. Graph. Appl.*, vol. 29, no. 3, pp. 76–84, May 2009.

[119] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater, "Visual realism enhances realistic response in an immersive virtual environment—Part 2," *IEEE Comput. Graph. Appl.*, vol. 32, no. 6, pp. 36–45, Nov. 2012.

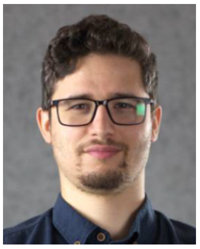
[120] A. Petti, W. Hutabarat, J. Oyekan, C. Turner, A. Tiwari, N. Prajapat, X.-P. Gan, and N. Ince, "Impact of model fidelity in factory layout assessment using immersive discrete event simulation," in *Proc. 8th Oper. Res. Soc. Simulation Workshop*, A. A. K. Hoad and M. Kunc, Eds. Birmingham, U.K.: OR Society, 2016, pp. 124–134.

[121] N. Mourkoussis, F. M. Rivera, T. Troscianko, T. Dixon, R. Hawkes, and K. Mania, "Quantifying fidelity for virtual environment simulations employing memory schema assumptions," *ACM Trans. Appl. Perception*, vol. 8, no. 1, pp. 1–21, Oct. 2010.
[122] K. Mania, A. Robinson, and K. R. Brandt, "The effect of memory schemas on object recognition in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 14, no. 5, pp. 606–615, Oct. 2005.
[123] K. Mania, S. Badariah, M. Coxon, and P. Watten, "Cognitive transfer of spatial awareness states from immersive virtual environments to reality," *ACM Trans. Appl. Perception*, vol. 7, no. 2, pp. 1–14, Feb. 2010.
[124] J. Jung and Y. J. Ahn, "Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study," *Comput. Animation Virtual Worlds*, vol. 29, nos. 3–4, p. e1812, May 2018.
[125] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *Int. J. Aviation Psychol.*, vol. 3, no. 3, pp. 203–220, Jul. 1993.
[126] J. Thomas, M. Azmandian, S. Grunwald, D. Le, D. Krum, S.-H. Kang, and E. Rosenberg, "Effects of personalized avatar texture fidelity on identity recognition in virtual reality," in *Proc. ICAT-EGVE*, 2017, pp. 97–100.
[127] P. J. Gianaros, E. R. Muth, J. T. Mordkoff, M. E. Levine, and R. M. Stern, "A questionnaire for the assessment of the multiple dimensions of motion sickness," *Aviation, Space, Environ. Med.*, vol. 72, no. 2, p. 115, 2001.
[128] L. S. Vailshery. (2021). *Global VR Headset Vendor Share 2016 | Statista*. [Online]. Available: https://www.statista.com/statistics/755527/global-vr-headset-market-share-by-vendor/
[129] B. Laha, D. A. Bowman, and J. D. Schiffbauer, "Validation of the MR simulation approach for evaluating the effects of immersion on visual analysis of volume data," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 4, pp. 529–538, Apr. 2013.
[130] Q. Lin, J. Rieser, and B. Bodenheimer, "Affordance judgments in HMD-based virtual environments: Stepping over a pole and stepping off a ledge," *ACM Trans. Appl. Perception*, vol. 12, no. 2, pp. 1–21, Apr. 2015.
[131] F. Hutmacher, "Why is there so much more research on vision than on any other sensory modality?" *Frontiers Psychol.*, vol. 10, p. 2246, Oct. 2019.
[132] R. Nordahl, S. Serafin, N. Nilsson, and L. Turchet, "Enhancing realism in virtual environments by simulating the audio-haptic sensation of walking on ground surfaces," in *Proc. IEEE Virtual Reality Workshops*, Costa Mesa, CA, USA, Mar. 2012, pp. 73–74.
[133] H. Benko, C. Holz, M. Sinclair, and E. Ofek, "Normaltouch and texturetouch: High-fidelity 3D haptic shape rendering on handheld virtual reality controllers," in *Proc. 29th Annu. Symp. User Interface Softw. Technol.* New York, NY, USA: Association for Computing Machinery, 2016, pp. 717–728.
[134] M. Slater, "How colorful was your day? Why questionnaires cannot assess presence in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 13, no. 4, pp. 484–493, Aug. 2004.
[135] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence, Teleoperators Virtual Environ.*, vol. 9, no. 5, pp. 497–503, Oct. 2000.

GUILHERME GONÇALVES received the bachelor's degree from the University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, in 2015, where he is currently pursuing the M.Sc. degree in multimedia. Since 2018, he has been a Research Fellow with the INESC TEC, Porto, Portugal. His research interest includes multisensory virtual reality.



PEDRO MONTEIRO received the B.S. and M.S. degrees in computer science from the University of Trás-os-Montes e Alto Douro, Vila Real, Portugal, in 2015 and 2018, respectively. Since 2016, he has been a Research Fellow with the INESC TEC, Porto, Portugal. His research interests include virtual reality, the use of multisensory stimuli in virtual reality applications, virtual reality interaction, and virtual reality user interfaces.



HUGO COELHO received the B.S. and M.S degrees in computer science from the University of Trás-os-montes e Alto Douro, Vila Real, Portugal, in 2015 and 2018, respectively. From 2017 to 2018, he was a Research Fellow the University of Trás-os-montes e Alto Douro. Since 2018, he has been a Research Fellow with the INESC TEC, Porto, Portugal. His research interests include multisensory virtual reality, virtual reality user interfaces, and mixed reality applications.

MIGUEL MELO is currently an Assistant Researcher with the INESC TEC, specialized in computer graphics. He is a Manager of the Multisensory Virtual Reality Laboratory. His research interests include computer graphics, HDR, and multisensory virtual reality. He is also a member of the Executive Committee of Eurographics.

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