IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 February 2023, accepted 6 March 2023, date of publication 9 March 2023, date of current version 15 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3254892

RESEARCH ARTICLE

How Much Presence is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score

MIGUEL MELO [1,2], GUILHERME GONÇALVES [1,2], JOSÉ VASCONCELOS-RAPOSO [1,3], AND MAXIMINO BESSA [1,2]

[1]Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal
[2]Department of Engineering, University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal
[3]Department of Education and Psychology, University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal

Corresponding author: Miguel Melo (mcmelo@inesctec.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 LA/P/0063/2020; and in part by the European Union's Horizon 2020–The EU Framework Program for Research and Innovation 2014–2020 under Grant 833573.

ABSTRACT Presence is often used to evaluate Virtual Reality (VR) applications. However, the raw scores are hard to interpret and need to be compared to other data to be meaningful. This paper leverages a database of 1909 responses to the Igroup Presence Questionnaire (IPQ) in different contexts to put forward a scale that qualitatively interprets raw Presence scores for VR experiences. The qualitative grading encompasses the acceptability dimension and analogous academic grading scales ranging from A to F and the adjective of such scores in a scale from Excellent to Unacceptable. Furthermore, the qualitative grading system encompasses Presence and its subscales Spatial Presence, Involvement, and Experienced Realism as defined by the IPQ. Adopting this grading system, supported by a robust dataset of Presence scores, enables practitioners to evaluate and interpret individual IPQ scores, allowing them to gain insights regarding the evaluated applications' effectiveness.

INDEX TERMS Virtual reality, evaluation metrics, presence.

I. INTRODUCTION

Virtual Reality (VR) technologies can transport users to Virtual Environments (VE). Its nature allows the development of virtual experiences targetting any application field, bringing significant benefits. As examples, we can quickly identify the advantages of adopting VR solutions in different application fields, such as entertainment, training, or education. For entertainment, richer content can produce a better quality of experience for users [1]. In training, VR solutions allow for more resilient training programs where trainers have more control under the training scenarios and ensure the trainee's security even under dangerous environments [2]. For education, VR systems allow illustrating abstract concepts to explain them better or even transport users to historical sites/moments so they can better understand the topics being taught [3].

To ensure that VR applications match their purpose, i.e., deliver an experience that delivers a good user experience and enables the fulfilment of the goals for which it was designed, it is of utmost importance to evaluate them properly. To this end, the body of knowledge is consensual that the sense of presence is a key factor in evaluating a VR application's effectiveness [4], [5], [6], which can be further complemented with other variables such as usability, satisfaction, cognitive load, or cybersickness, depending on the purpose of the VR application being evaluated. However, despite the common understanding that presence can be defined as the sense of "being there" (i. e., in the VE), different definitions were proposed for this phenomenon and how it develops.

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

A. THE CONCEPT OF PRESENCE

Skarbez et al. [7] categorise presence definitions according to three dimensions: ''being there'', ''non-mediation'', and ''other''. One of the first references to a state of ''being there'' is attributed to Minsky [8] and evolved to an explicit VR context by Steuer [9] that define presence as ''*the sense of being in a mediated environment*''. In this context, some authors define presence with a focus on a spatial domain, defining presence as the subjective feeling of being physically present in a mediated environment [10], [11]. Other authors focus on the psychological state of the illusion of an actor of an alternative world [12], [13]. Biocca [14] extends the definition of presence to the sense of ''being there'' combined with the capability of being *active* on the VE, i.e. being capable of initiating actions in the VE and having the VE responding accordingly. The ''non-mediation'' dimension aggregates Presence definitions that frame presence as the level of abstraction in a technology-mediated environment. More precisely, authors define presence as the perceptual illusion of non-mediation [15] or as the suspension of disbelief where users forget about the real world and believe that they are actually in another world [16]. In its turn, the Presence definitions categorised as ''other'' define presence as the perceptual processing of the virtual elements as real dimensions. In this context, Parola et al. [17] coined presence as the sense of feeling the VE as being real. Waterworth and Waterworth proposed another conceptualisation, which divided presence into three components: focus (perceptual processing of the stimulus vs conceptual processing), locus (the degree of attention devoted to the virtual or the real world), and sensus (level of conscious arousal when interacting with the VE) [18]. Note that despite these three different categorisations considering that presence is a media-related phenomenon, i. e. they are linked to a perceptual illusion, some authors theorise that presence is a construct of both the perceptual illusion of being in the VE (media presence) and the psychological/ecological phenomena that virtual medium elicits that can impact the sense of presence even if the mediated environment can create an optimal non-mediated illusion (inner presence) [19].

Sanchez-Vives and Slater [20] linked the concept of presence to the neuroscience field, relating the state of consciousness of being attached to a virtual place and situation to the extent to which the users would respond realistically to the virtual stimuli at physiological, emotional and behavioural levels. Subsequently, Slater [21] proposed a theoretical framework for presence which postulates that presence is composed of two components: Place Illusion (PI) and Plausibility illusion (Psi). The PI component embraces the ''being there'' definition of presence, defined as the illusion of being in a place (e.g., a virtual environment) even knowing that one is not there. The Psi component refers to the capability of the VR system to mimic real-world events in such a credible way that users are illuded that they are happening in reality. The fulfilment of PI and Psi by a VR system will lead users to act in the virtual world as if they were in a real scenario [21]. VR is not limited to a single user and, consequently, in addition to the considerations of the different visions on presence, some authors embrace the concepts of Copresence (the sense of sharing the virtual space with another persona) [22] or Social Presence (extends the awareness of other personas in the virtual space with the possibility of developing interpersonal interaction) [23]. From a neuroscience point of view, Riva et al. [24] have studied the use of VR in behavioural health and suggested a strong interconnection exists between how the brain processes reality and the features of VR technology that allows the developing of a sense of presence. In the real world, the brain builds an embodied simulation of the real world surrounding the users to regulate and control the body effectively. Taking this into account, if VR technology can predict the sensory consequences of their actions and respond to the users' actions as expected, VR will be able to deceive the human brain and successfully elicit a sense of presence [24], [25].

As a result of a literature survey on Presence, Skarbez et al. [6] have proposed a Presence model that considers these different views on presence, where they describe presence as a function of the subjective feeling of PI, PsI, and Social Presence Illusion. They further theorise that PI is elicited by the immersion provided by the VR system, that PsI is a consequence of the coherence of the scenario, and that the Social Presence Illusion is supported by the Copresense Illusion, which is given by the capability of the system of supporting other personas alongside the virtual experience.

Presence is closely related to key concepts such as immersion, involvement, and realism. The concept of immersion is often blended with the concept of presence. In the literature, immersion has been defined from different angles. For instance, Slater [21] defines Immersion as an objective variable that consists of the capability of the VR system to isolate the users from the real environment. On the other hand, Witmer and Singer [26] define presence as the subjective feeling that a user develops for being continuously enveloped, included in, and interacting with the VE. Involvement refers to the degree of attention devoted to the VE and degree of engagement with the stimuli, activities, and events of the VE [26], [27].

According to Perroud et al. [28], realism can be divided into five components: realistic-looking (quality of shaders, lights and overall artistic look), realistic construction of the virtual world (the use of scientific models such as laws of physics), physiologic realism (objective realism of a given virtual environment where the stimuli users receive are the same as in the real world), psychological realism (the user subjective perception of realism) and presence. These different dimensions allow us to categorise realism into two categories: subjective (user perception) and objective (how close to the virtual experience's real-world counterpart) [5]. Therefore, regarding subjective user perception, one can sense that the virtual environment is realistic even if the environment does not replicate the real-world stimuli accurately. However, on the other hand, a user can also feel that a virtual experience is unrealistic, even though it replicates real-world stimuli accurately. One can only objectively compare the realism of a virtual experience if it can also be replicated in the real world. However, some experiences do not aim to simulate real-world conditions but can still be considered authentic by users, developing a sense of presence. The concept of coherence helps to explain this. Skarbez et al. [29] define coherence as ''the set of objectively reasonable circumstances that the scenario can demonstrate without introducing objectively unreasonable circumstances''. A coherent virtual experience does not require it to be replicable in the real world. It depends on the user's prior knowledge, experiences, and expectations; therefore, it cannot be entirely objective. However, the author further affirms that it should still be considered objective; despite the inability to control the users' prior knowledge and experiences, they can control whether the virtual experience events are internally consistent. So, a virtual experience based on fantasy (i.e., impossible to happen in the real world) can still be coherent as long as it follows the internal rules of that VE.

B. EVALUATING PRESENCE

As a result of the different views on presence and its associated concepts, there are multiple approaches to evaluating presence. The Presence evaluation can be carried out using objective, subjective, or complementary metrics. Objective metrics are based on evaluation instruments that record data regarding the virtual experience of the participant and do not depend on their opinion. Examples of objective metrics are physiological responses such as cardiac measures or skin conductance and behavioural analysis. For instance, cardiac measures are used as an indicator of presence based on the fact that the heart rate and skin conductance are significantly correlated with the reported sense of presence via questionnaires [30], [31]. The objective metrics based on behavioural analysis refer to metrics such as facial expression analysis that can be used to detect emotions, being suggested that more emotions can suggest more presence [32] or postural and reflex responses where it is expected that when presence is developed, the users will respond to the virtual stimuli as they would to the real stimuli [33], [34]. The subjective metrics essentially consist of reports given by users after exposure to VR applications where questionnaires or interviews are the most used instruments [35]. Most of the evaluations of VR applications rely on subjective evaluation, mainly obtained via questionnaires, as they are easy to adopt and administer. In contrast, objective methodologies such as analysing physiological signs are often invasive and require complex and often expensive setups, which are not as widely available as a simple questionnaire [36].

Among the most used questionnaires, there are the Slater-Usoh-Steed (SUS) questionnaire [37], the Presence Questionnaire (PQ) [13], [26], and the Igroup Presence Questionnaire (IPQ) [38], [39]. The SUS questionnaire was one of the first proposals to measure Presence, and it is composed of seven items that measure the overall sense of presence that, according to the authors, refers to the ''subjective experience of being in one environment (there) when physically in another environment (here)'' [26]. As for the PQ, it was first proposed in 1998, and it was composed of 19 items that evaluate three dimensions: involved/control (perceived control of the events of the VE, how involving the visual aspects are and how involved the participant becomes in the experience), natural (to which extent interactions are felt natural and the extent of consistency between the VE and the reality), and interface quality (control/display devices interfere or distract the users and the degree of concentration that users can devote to the tasks in the VE) [26]. This questionnaire was later revised, and its last version is composed of 29 items that measure four dimensions: involvement, sensory fidelity (how users can feel the various stimuli of the VE), adaptation/immersion (capability of adapting to the VE and feel involved by it), and interface quality (the extent to which the system can deliver the virtual content without interfering with the user's virtual experience) [13]. The IPQ questionnaire encompasses four subscales: Spatial Presence, Involvement, Experienced Realism and Overall Sense of Presence. The Spatial Presence subscale refers to users' sense of being physically present in the VE. Involvement measures the attention devoted to the VE and the involvement experienced, i. e. the sense of taking part in the VE and being engaged with it. Experienced Realism assesses the subjective experience of realism in the VE. At last, the Overall Sense of Presence (hereinafter referred to as Presence) determines the general subjective sense of being in the VE, which is obtained by averaging the mean scores of the Spatial Presence, Involvement, and Experienced Realism subscales.

C. MOTIVATION AND CONTRIBUTIONS

Presence questionnaires are valuable in assessing the effectiveness of VR applications. However, the evaluation results of a particular application are hard to interpret as the administration of these questionnaires alone only provides a score, which is the major shortcoming that they do not qualify the degree of effectiveness of a system. For instance, how does one interpret the average score of four points? It is imprecise to conclude that a VR application successfully creates this sense of presence if the achieved Presence score is equal to or higher than 4 or if a score of 2.5 means that a given VR application is average or poor. This work provides a new instrument for evaluating VR applications: a set of qualitative scales that allows researchers to evaluate VR applications based on the Presence scores obtained qualitatively, allowing meaningful insights into the VR applications evaluated. This work focuses on using the IPQp as an evaluation instrument, and the robustness of the proposed qualitative scales is given by a database of more than 1900 evaluations across 21 different studies. Moreover, the proposed qualitative scale is envisaged to follow a collaborative approach where everyone can contribute to the robustness of the database.

II. METHODOLOGY

The main goal of this paper is to propose a qualitative scale for the IPQ so it is possible to add meaning to the Presence scores obtained when evaluating a VR application. Information regarding how this process was carried out is described next.

A. IPQ ITEMS

The IPQ questionnaire comprises 14 items that measure four factors: Presence, Spatial Presence, Involvement, and Experienced Realism. Table 1 shows the questionnaire items grouped by factor. A complete description of each item, the anchors for classifying them and the calculation of the different factors can be found in the corresponding validation papers of the English version [40] and Portuguese version [4].[1]

TABLE 1. Items of the IPQ divided by subscales.

Factor	Item
Presence	In the computer generated world I had a sense of “being there”
Spatial Presence	Somehow I felt that the virtual world surrounded me.
	I felt like I was just perceiving pictures.
	I did not feel present in the virtual space.
	I had a sense of acting in the virtual space, rather than operating something from outside.
	I felt present in the virtual space.
Involvement	How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?
	I was not aware of my real environment.
	I still paid attention to the real environment.
	I was completely captivated by the virtual world.
Experienced Realism	How real did the virtual world seem to you?
	How much did your experience in the virtual environment seem consistent with your real world experience?
	How real did the virtual world seem to you?
	The virtual world seemed more realistic than the real world.

B. IPQ DATA

The data was obtained by contacting several authors that carried out VR evaluations using the IPQ [38]. The data collected comes from 21 different studies and consists of a total of 1909 responses to the IPQ, being that 1367 of the responses were obtained with the translated and validated Portuguese version of the questionnaire [39] (hereinafter referred to as IPQp) provided by the studies’ authors and 542 of the responses were obtained through the original German version of the questionnaire.[2] The data sources are overviewed in Table 2. It is important to refer that the German and Portuguese versions using different Likert scales. The German IPQ uses a 7-point Likert scale, while the Portuguese IPQ uses a 5-point Likert scale. Thus the database creation included converting the 5-point Likert scores to 7-point Likert scores through a linear interpolation with arithmetic adjustment. For this, the scale was adjusted to start at 0 by subtracting 1 from the 5-point rating, followed by a multiplication by $\frac{6}{4}$ (maximum value of the 0-6 point scale by the maximum value of the 0-4 scale), and finally by adding 1 to adjust the scale again. Please note that this rescaling does not affect the data characteristics required for this study [41].

[1] The exact calculation formulas can also be found at http://www.igroup.org/pq/ipq/data.php (for the English version) and https://massive.inesctec.pt/wp-content/uploads/2016/11/IPQ-Calculo-1.pdf (for the Portuguese version)

[2] The German database is publicly available at Igroup website at http://www.igroup.org/pq/ipq/data.php

TABLE 2. Overview of the data sources of the IPQ scores used.

Study	Sample	Age Avg.	Presence Mdn.	Spatial Presence Mdn.	Inv. Mdn.	Exp. Realism Mdn.
[42]	20	23.70	2.54	2.51	3.00	1.88
[43]	104	21.06	3.71	4.50	3.38	3.38
[44]	48	23.42	3.65	4.38	3.00	2.40
[39]	478	24.54	3.50	4.50	3.38	2.63
[45]	50	22.80	3.54	3.75	2.81	3.38
[46]	48	21.38	3.75	4.63	3.75	3.00
[47]	59	22.73	3.32	4.00	3.00	1.88
[48]	99	23.14	4.04	5.00	4.13	3.00
[49]	10	N.A.	3.48	4.13	3.56	2.06
[50]	48	24.50	3.38	3.75	3.38	3.00
[51]	68	21.43	3.96	4.25	4.13	3.57
[52]	78	22.71	3.86	4.25	4.13	3.38
[53]	37	22.35	3.96	4.01	4.50	3.75
[54]	7	21.14	3.54	4.25	2.63	2.25
[55]	24	25.38	3.49	4.01	3.75	3.00
[56]	32	21.97	3.32	3.75	2.82	2.63
[57]	32	22.50	3.86	4.50	3.57	3.00
[58]	35	23.38	3.42	3.75	3.75	3.00
[59]	90	42.27	3.49	4.01	2.63	3.94
[38] (I)	246	24.54	3.06	3.60	3.00	1.75
[38](II)	296	24.66	2.99	3.20	3.25	2.99

C. THE QUALITATIVE SCALES

Presence scores are typically used to evaluate VR applications as a comparison metric, i. e. different versions of a VR application are evaluated, and the scores are compared to each other to understand which version had the best scoring. Despite being valuable for relative comparisons, the scores alone do not indicate if the applications are acceptable (or not) or if they are already at a level of excellence. The only reference is the scale (0 to 6), where the evaluators can assume that 0 classifies the VR application as the worst possible application and 6 as the best one imaginable. Although this interpretation is valid for extreme values, it is incorrect to directly translate intermediate values since this type of score follows a curve grading, implying that a score of 3.50 can have different meanings depending on the subscale. As can be verified in Table 2, the Presence scores tend to diverge from the mid-value of the scale (3 on a scale from 0 to 6), meaning that using simple averages to convert the raw Presence scores into percentages and/or qualitative grades would be misleading. Thus, to properly compare a raw Presence score and the existing database, percentile ranks were adopted. This approach also allows an understanding of how well the VR application performed against others.

The conversion to percentile ranks allows a more accurate interpretation of the application's overall performance and its level of maturity. Nevertheless, we propose to associate a qualitative adjective to these scores for a more meaningful result regarding the Presence subscale scores that is easier to interpret by the general public. To define the ranges of each classification and attribute it to a qualitative evaluation, we adopted the approach that Bangor et al. [60] used when proposing a qualitative scale for the System Usability Scale (SUS) [61], which is based on the well-known academic grading system where grades can range from A (excellent) to F (unacceptable) and adapted inspired by the benchmarks by Sauro [62] as follows:

- A: outstanding performance, users enjoyed it and might recommend it;
- B: solid performance, minor improvements can still be made;
- C: satisfactory performance but with a margin for improvement;
- D: marginally acceptable, has the minimum performance for being acceptable, but it is advisable to be reviewed for improvement;
- E: unsatisfactory performance, could be used in particular scenarios cases but shall not be distributed to the public;
- F: unacceptable performance with critical aspects that shall be fixed as soon as possible before further usage.

Following the previous line of thought and inspired by Bangor et al. [60], it is important also to provide a grading scale that allows evaluating the application from a usability point of view. This categorization will consider the Presence scores higher than the percentile 70 acceptable, between the 50 and 70 percentiles considered marginally acceptable, and below the 50 percentile considered not acceptable. Table 3 summarizes the grading systems used as qualitative reference scales for raw Presence scores.

TABLE 3. Qualitative grading description.

Percentile	Grade	Adjective	Acceptability
≥ 90	A	Excellent	Acceptable
≥ 80	B	Very Good	
≥ 70	C	Satisfactory	
≥ 60	D	Marginal	Marginally acceptable
≥ 50	E	Unsatisfactory	
< 50	F	Unacceptable	Not Acceptable

III. RESULTS

As mentioned in section II-B, the data was processed using a 7-point Likert score as a basis. To allow a better data characterisation, we first present the results regarding the number of participants per study, the samples' ages, the distribution of the studies comprised in the database by application field, and

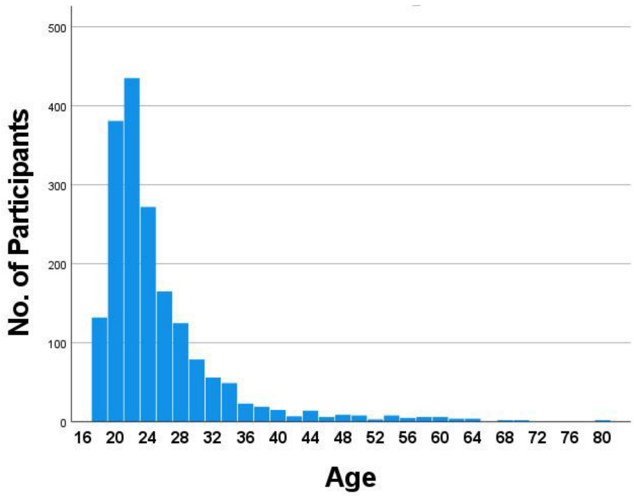

FIGURE 1. Age of the study's participants.

the distribution of the studies by type of VR setup used. Then, the data distribution regarding each subscale is presented, followed by the percentile values divided by subscale.

A. DATA CHARACTERIZATION

A per study-based analysis reveals a median sample of 48 participants (*Std. Dev.* = 7.85) with a minimum of 7 participants and a maximum of 478 participants per study.

The average age of the participants is 24.58 (*Std. Dev.* = 7.85). Figure 1 shows the distribution of ages of the participants of the analysed studies.

The majority of the studies (15 out of 21 - approximately 71%) comprised in the database are general-purpose, i.e. they are not focused on a specific application field. From the remaining studies, four studies focus on using VR for training applications; one study is devoted to the tourism sector, and the other to the entertainment field.

A data analysis taking into account the experimental scenarios of each study has revealed that the majority of the answers in the database (1317 answers, corresponding to 69%) were given after the exposure to an immersive setup (i. e., head-mounted display), being that non-immersive desktop-based setups were used in 29% of the cases. There were 2% of the participants whose VR setup information was not reported.

B. DATA DISTRIBUTION

The mean, median, score ranges, and skewness values of the aggregated data of all studies are shown in Table 4. The average Presence score for all the studies is 3.47 (*Std. Dev.* = 0.85) with a median of 3.47 and a range from 0.21 to 5.88 (Table 4 presents the complete data for all subscales).

C. DATA PERCENTILES

As mentioned previously (refer to section II-B), depending on the version of the questionnaire, a 5-point or a 7-point Likert scale can be used. Having as a reference the 7-point Likert

TABLE 4. Overall descriptive statistics of the IPQ data sources.

Scale	Mean	Std. Dev.	Mdn.	Min.	Max.	Skewness
Presence	3.40	0.85	3.47	0.21	5.88	-0.369
Spatial Presence	3.99	1.05	4.01	0.00	6.00	-0.680
Involvement	3.33	1.13	3.38	0.00	6.00	0.039
Experienced Realism	2.65	1.23	2.63	0.00	6.00	0.358

scale, the percentile values calculated based on the database aggregating all data sources are shown in Table 5. The conversion of the 7-point Likert scores to 5-point Likert scores was achieved through a linear interpolation with arithmetic adjustment, ensuring no effect on the data characteristics [41]. A similar process as described in section II-B was adopted: we subtracted 1 from the 7-point Likert score to adjust the scale to 0, multiplied by $\frac{4}{6}$ and then added back 1 to adjust the scale again. Table 6 shows the scores converted from a 7-point scale to a 5-point scale.

IV. DISCUSSION

This work focused on providing researchers with a tool to evaluate VR applications independently and thoughtfully. Current methods only provide a raw score that is not very informative unless we have a comparison term. For this purpose, we considered data from over 1900 responses to the IPQ questionnaire to propose a qualitative scale that gives meaning to the raw evaluation scores.

To better understand the source of the data, a data characterisation analysis was carried out regarding the sample size of the studies, their ages, the application field of each study, the number of studies per rank, and the average scores per study over the years. Regarding the sample sizes, there is a median sample of 48 participants per study, with three studies standing out: two published in the same paper with 246 and 296 participants [38], and one other study with 478 participants [39]. These works refer to the original proposal and validation of the IPQ questionnaire and the Portuguese translation and validation of the questionnaire. Thus, such a large number of participants is justified because the questionnaire validation was conducted through factor analysis. On the opposite end, there are two studies with a low number of participants: one pilot study with ten participants that assesses correlations between subjective and objective Presence metrics [49] and work by Narciso et al. [54] that focus on experimental studies to assess the effectiveness of VR in the training of firefighters. Despite this variance in the samples across studies, one shall consider that the database used is composed of studies that adopted validated versions of the IPQ questionnaire and that these studies were duly published in recognised peer-reviewed venues. This is a criterion of the quality of the database as the peers' revision ensures a formal verification of the samples' adequacy and the validity of the experimental study itself.

The average age of the participants was 24.58; the youngest participant was 18 years old, while the older participant was 80. It is noticeable that there is a predominance of participants between 20 and 30 years. A reason that can help explain why there are no participants below the age of 18 is that in the evaluations carried out in the studies, there was a need to provide informed consent, which only people 18 or more years could do. Another reason that can also explain why participants are mostly between 18 and 34 is that, generally, the evaluations adopted a convenience sampling technique. In addition, most of the studies occurred in university-based locations. As such, the persons available in that context are typically between those ages. Although this can be identified as a limitation, this is mitigated by the fact that the age distribution of VR users is similar to the distribution of ages of the sample [63].

Results also revealed that the database mainly comprises general-purpose studies (16 out of 21) and includes four studies focused on virtual training and one applied to virtual tourism. This indicates that the data is not biased towards a particular application field. Thus, the data can be interpreted without such a restriction. As for the data distribution by the type of VR setup used, the database is composed of considerable samples of both immersive VR systems (69%) and non-immersive VR systems (29%), demonstrating the heterogeneity of the VR landscape.

The distribution of the studies per rank is illustrated in Figure 2. As can be seen, there are no studies that achieved Grade A, and only a small fraction achieved Grade B ([48] for Spatial Presence, [53] for Involvement, and [53], [59] for Experienced Realism - all general purpose studies) and the majority of the studies were graded as F, meaning that it is desirable that they are further improved. This may suggest that, currently, VR experiences can potentially deliver an experience of realism and involve users. However, the considered studies could not elicit the same levels of Presence or Spatial Presence, meaning that, regarding the considered studies, these two components deserve more focus to improve the evaluated VR solutions.

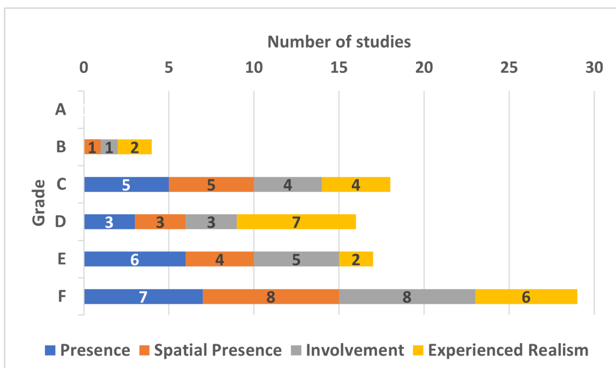

FIGURE 2. Distribution of the studies per ranks regarding the different Presence subscales.

Figure 3 was developed to allow a better understanding of the evolution of the Presence scores in the considered studies taking into account the publication year. As can be

TABLE 5. Qualitative grading description according to IPQ subscale scores (7-point Likert scales).

Presence	Spatial Presence	Involvement	Experienced Realism	Grade	Adjective	Acceptability
≥ 4.41	≥ 5.25	≥ 4.87	≥ 4.50	A	Excellent	Acceptable
≥ 4.07	≥ 4.76	≥ 4.50	≥ 3.75	B	Very Good	
≥ 3.86	≥ 4.50	≥ 4.00	≥ 3.38	C	Satisfactory	
≥ 3.65	≥ 4.25	≥ 3.75	≥ 3.00	D	Marginal	Marginally acceptable
≥ 3.47	≥ 4.01	≥ 3.38	≥ 2.63	E	Unsatisfactory	
< 3.47	< 4.01	< 3.38	< 2.63	F	Unacceptable	Not Acceptable

TABLE 6. Qualitative grading description according to IPQ subscale scores (converted to 5-point scales).

Presence	Spatial Presence	Involvement	Experienced Realism	Grade	Adjective	Acceptability
≥ 3.94	≥ 4.50	≥ 4.25	≥ 4.00	A	Excellent	Acceptable
≥ 3.71	≥ 4.17	≥ 4.00	≥ 3.50	B	Very Good	
≥ 3.57	≥ 4.00	≥ 3.67	≥ 3.25	C	Satisfactory	
≥ 3.43	≥ 3.83	≥ 3.50	≥ 3.00	D	Marginal	Marginally acceptable
≥ 3.31	≥ 3.67	≥ 3.25	≥ 2.75	E	Unsatisfactory	
< 3.31	< 3.67	< 3.25	< 2.75	F	Unacceptable	Not Acceptable

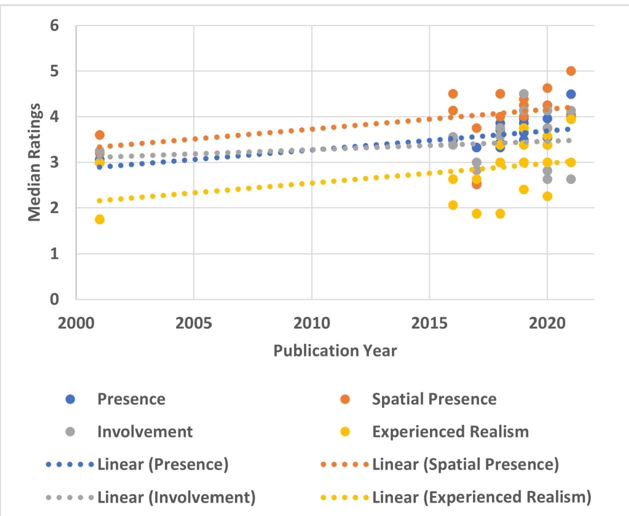

FIGURE 3. Median scores of the considered studies over the years.

verified by the dotted lines, there is an upward trend in the median scores in every subscale. This can be attributed to the fact that VR technologies are evolving and, consequently, more recent virtual experiences have more potential to be better. Nevertheless, this can be moderated because the current context also shapes the users' expectations. While it is true that a novelty factor can contribute to better scoring, familiarity with the technology tends to normalise the scoring [64]. Following that line of thought, we speculate that an application developed in the early 2000s, when compared directly to an application developed in the 2020s, is normal to obtain lower scores. Still, when evaluated in their original context (i.e., when each one was released), they will be rated with identical criteria as user expectations are levelled by the current technology expectations and novelty of their current context. This is a topic to be explored in future research. Nevertheless, it shall be continuously updated to strengthen the scale's robustness (section IV-A expands on this limitation and future work).

Overall, the data distribution corroborates the outcomes of Schubert et al. [38] and Regenbrecht and Schubert [65] that Presence and Spacial Presence have a similar distribution (Figure 4 and 5, respectively) as it was demonstrated that Spatial Presence has an extensive loading on Presence, meaning that the subscales overlap of the components evaluated. This is explained by the fact that the Spatial Presence subscale comprehends the widely accepted concept of presence as the feeling of "being there" and emphasises the sense of acting inside the virtual world.

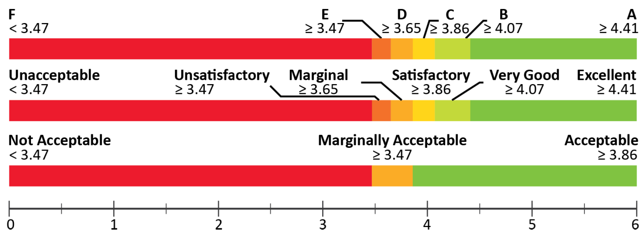

FIGURE 4. Presence score scale for 7-point Likert scales.

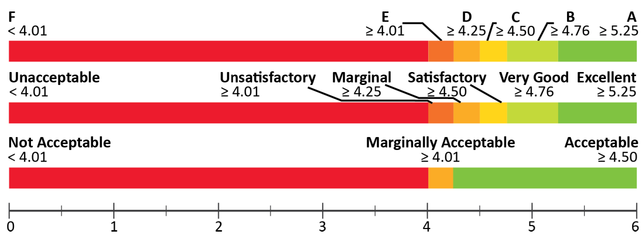

FIGURE 5. Spatial presence score scale for 7-point Likert scales.

As for Involvement, the skewness value is higher than Presence and Spatial Presence skewness values but lower than Experienced Realism's skewness. The data distribution is slightly skewed left (Figure 6, meaning that the data is roughly equally balanced around the mean but tends to be positive. Such can be attributed to the fact that Involvement measures the awareness and attention processes which means that it is a general Presence factor as framed by Witmer and Singer [26]. It also means that it loads Experienced Realism as it also considers the judgement over some items of the VE that relate to their comparability with reality, as can be verified in Table 1.

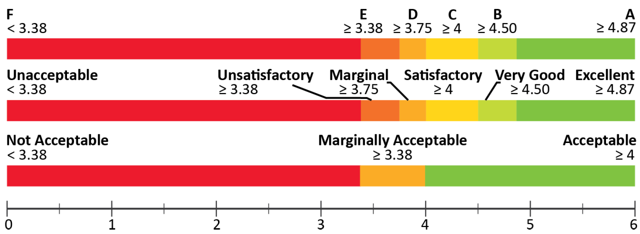

FIGURE 6. Involvement score scale for 7-point Likert scales.

Experienced Realism is the only right-skewed factor (7), indicating that its focus is on distinct components of the virtual experience despite loading the remaining factors. Namely, this factor loads on Presence by attributing realness to the virtual content, being the questions focused on comparisons between the virtual and the real world. These comparisons are affected by different factors, such as the personal background of subjects. For instance, it is expected that a gamer who attributes higher scores to Experience Realism is more prone to ignore the constraints of reality [38]. Another important factor is the application type (for instance, a fantasy game or a training application), and the interaction metaphors used (for instance, real vs illusory) can shape the judgement of realness [65]. Consequently, the scoring of this factor is lower.

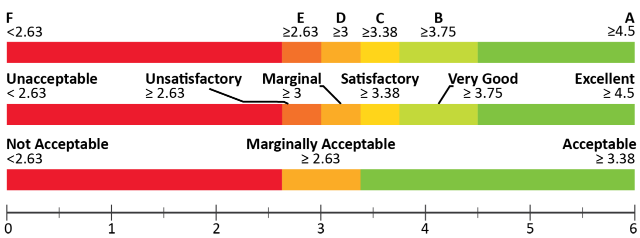

FIGURE 7. Experienced realism score scale for 7-point Likert scales.

As a contribution to the adopters of the qualitative scales proposed in this paper, a graphical scale was developed for each subscale (Figures 4, 5,6, and 7) for ease of reference and reporting of the results.

The adoption of the qualitative scales proposed, more than evaluating new VR applications, it is also possible to interpret and compare results with older studies and open new research lines by studying which factors can bring Presence scores to acceptable levels and which have little impact and might not be worth implementing considering the system performance impact or human resources needed. Also, it can bring new insights about how acceptable the previous VR applications' Presence levels (and its constructs) were and possibly justify the need to improve them further or, on the other hand, further prove their "Presence effectiveness". It can also help better explain the behaviour of other variables, directly or indirectly related to Presence, found in previous studies by knowing what an acceptable/satisfactory Presence score is. Furthermore, the evolution of the understanding of the presence concept from a neuroscience point of view can provide valuable insights to VR developers when developing the virtual experience and associated stimuli and behaviours, as they can take advantage of the knowledge regarding the brain model and how the brain generates the embodied simulation of the real world [24]. This includes also understanding from a biological point of view how the human perceives spatial presence and develops a sense of involvement or the sense of experienced realism. For instance, navigation is a major variable for spatial presence, and walking metaphors are preferable since they are more natural. Still, they are not always effective since the virtual space can be larger than the real space and requires complementary navigation techniques such as teleporting or redirect walking. As such, one may take advantage of the visual cue-related activity of cells in the medial entorhinal cortex for estimating positions through path navigation and manipulating them to induce the desired sense of location in the environment and, consequently, increase the sense of spatial presence [66]. In the involvement and experienced realism domains, the same is applicable. For instance, understanding how involvement develops taking into account the experience of flow and the associated cognitive functions [67] or how the external stimuli interact with the internal signals at a neuroscience level contribute to the sense of reality [68] can be determinant for optimal development of the sense of presence.

A. LIMITATIONS AND FUTURE WORK

The present work is not free of limitations. For example, a latent limitation is the limited sample distribution across different application domains, as the definition of the effectiveness of a VE can vary across domains. Thus, when this verifies, we recommend that the qualitative evaluation of VR applications are complemented with additional metrics.

Due to the fast pace of VR, VR technologies will allow one to put forward better virtual experiences. This fact, associated with the fact that the body of literature is expanding, can contribute to a scale bias, as the new applications would be skewed right in the scale. Conscious of such a limitation, the authors face this work as a living scale: the scale is designed to be open to the community for consultation and contributions. More specifically, the latest version of the scale will always be available as well as anyone can contribute to it by sharing databases of studies that make use of the IPQ questionnaire so it can be periodically updated[3]

[3]Contact the manuscript authors or visit the website https://massive.inesctec.pt/publications/resources/ for more information

Additionally, in future work, as new studies are carried out and new databases are made available, they will be considered to optimise the qualitative scales according to the different VR application fields. Such would be an important topic as some application fields might not require the same levels of presence to be considered acceptable, while others might require above-average scores. Also, the importance of each Presence construct can vary depending on the application field. Also, the age distribution of the sample does not allow us to proceed with analysis by age group/range. While literature is scarce and, yet, not consensual regarding the impact of age on the reported sense of presence (e.g. [69], [70], [71]), this is an important point to address in future studies. Also, a factor that was not possible to accomplish was to output a qualitative scale for each type of VR setup (non-immersive, semi-immersive and fully immersive) due to the dataset distribution. Further work shall investigate the variance between different VR setups to create an analogous qualitative scale. Also, some of the data provided by the authors that contributed to the database consisted of the individual scores for each subscale. The individual scoring for each item was unavailable, preventing us from performing a reliability analysis. Nevertheless, this is mitigated by the fact that all the considered studies were conducted using a validated version of the IPQp, which maintains the integrity of the data.

Another limitation of this work is that our qualitative scale was based on the scores from IPQ, so it cannot be applied to other questionnaires. Therefore, future work should also address other presence questionnaires to create their corresponding qualitative scales. In addition, the unification of Presence scores from the more adopted Presence questionnaire and the creation of the corresponding qualitative scales should be done to allow direct comparison between studies with different questionnaires.

V. CONCLUSION

This work intends to provide VR researchers and developers with a tool to evaluate their products by giving meaning to a score obtained from the administration of the well-known Presence questionnaire IPQ. For this purpose, a comprehensive database of responses to the questionnaires from published studies was built. Based on the analysis of the resulting database and by adopting the analogous university grading system, it is possible to classify VR applications in terms of acceptability and to qualify the applications using an adjective scale. Furthermore, the fact that the subscales were also included in this analysis allows us to extract richer information about the evaluation process and identify possible room for improvements regarding specific features of the VR application regarding Spatial Presence, Involvement, and Experienced realism. Also, the possibility of assessing the application without having to compare it against other applications enables more insightful iterative development cycles. At each iteration, it is possible to analyse the different factors given by the subscales and focus on the priorities accordingly.

REFERENCES

[1] K. Kim, "Is virtual reality (VR) becoming an effective application for the market opportunity in health care, manufacturing, and entertainment industry?" *Eur. Sci. J., ESJ*, vol. 12, no. 9, p. 14, Mar. 2016. [Online]. Available: https://eujournal.org/index.php/esj/article/view/7179

[2] D. Narciso, M. Melo, S. Rodrigues, J. P. Cunha, J. Vasconcelos-Raposo, and M. Bessa, "A systematic review on the use of immersive virtual reality to train professionals," *Multimedia Tools Appl.*, vol. 80, no. 9, pp. 13195–13214, Apr. 2021. [Online]. Available: https://link.springer.com/10.1007/s11042-020-10454-y

[3] B. Peixoto, R. Pinto, M. Melo, L. Cabral, and M. Bessa, "Immersive virtual reality for foreign language education: A PRISMA systematic review," *IEEE Access*, vol. 9, pp. 48952–48962, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9386131/

[4] MASSIVE. (2021). *Resources—Igroup Presence Questionnaire—Translated and Validated to Portuguese (IPQP)*. [Online]. Available: https://massive.inesctec.pt/publications/resources/

[5] G. Gonçalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa, "Systematic review of comparative studies of the impact of realism in immersive virtual experiences," *ACM Comput. Surveys*, vol. 55, no. 6, pp. 1–36, Dec. 2022, doi: 10.1145/3533377.

[6] R. Skarbez, F. P. Brooks Jr., and M. C. Whitton, "A survey of presence and related concepts," *ACM Comput. Surv. (CSUR)*, vol. 50, no. 6, p. 96, 2018.

[7] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton, "A psychophysical experiment regarding components of the plausibility illusion," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 4, pp. 1369–1378, Apr. 2017.

[8] M. Minsky, "Telepresence," Omni Magazine, Jun. 1980, pp. 45–52. [Online]. Available: https://isfdb.org/cgi-bin/pl.cgi?59862

[9] J. Steuer, "Defining virtual reality: Dimensions determining telepresence," *J. Commun.*, vol. 42, no. 4, pp. 73–93, 1992.

[10] W. Wirth, S. Wolf, U. Mogerle, and S. Bocking, "Measuring the subjective experience of presence with think-aloud method: Theory, instruments, implications," in *Proc. 7th Annu. Int. Workshop Presence*, 2004, pp. 351–358.

[11] A. Spagnoli and L. Gamberini, "The sense of being 'there': A model for the space of presence," in *Proc. 7th Annu. Int. Workshop Presence*, M. A. Raya and B. R. Solaz, Eds. Valencia, Spain: Universidad Politecnica de Valencia, Oct. 2004, pp. 48–53.

[12] M. Slater, "A note on presence terminology," *Presence Connect*, vol. 3, no. 3, pp. 1–5, 2003.

[13] 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.

[14] F. Biocca, "Inserting the presence of mind into a philosophy of presence: A response to Sheridan and Mantovani and Riva," *Presence, Teleoperators Virtual Environ.*, vol. 10, no. 5, pp. 546–556, Oct. 2001.

[15] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *J. Comput.-Mediated Commun.*, vol. 3, no. 2, Jun. 2006, Art. no. JCMC321.

[16] M. Slater and M. Usoh, "Representations systems, perceptual position, and presence in immersive virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 2, no. 3, pp. 221–233, Jan. 1993.

[17] M. Parola, S. Johnson, and R. West, "Turning presence inside-out: Meta-Narratives," *Electron. Imag.*, vol. 28, no. 4, pp. 1–9, Feb. 2016.

[18] E. L. Waterworth and J. A. Waterworth, "Focus, locus, and sensus: The three dimensions of virtual experience," *CyberPsychol. Behav.*, vol. 4, no. 2, pp. 203–213, Apr. 2001.

[19] C. Coelho, J. Tichon, T. J. Hine, G. Wallis, and G. Riva, "Media presence and inner presence: The sense of presence in virtual reality technologies," *From Commun. Presence, Cognition, Emotions Culture Towards Ultimate Communicative Exper.*, vol. 11, pp. 25–45, Jan. 2006.

[20] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," *Nature Rev. Neurosci.*, vol. 6, pp. 332–339, Apr. 2005.

[21] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philos. Trans. Roy. Soc. B, Biol. Sci.*, vol. 364, no. 1535, pp. 3549–3557, 2009.

[22] R. Schroeder, "Copresence and interaction in virtual environments: An overview of the range of issues," in *Proc. 5th Int. Workshop*, 2002, pp. 274–295.

[23] J. Short, E. Williams, and B. Christie, *The Social Psychology of Telecommunications*. London, U.K.: Wiley, 1976.

[24] G. Riva, B. K. Wiederhold, and F. Mantovani, "Neuroscience of virtual reality: From virtual exposure to embodied medicine," *Cyberpsychology, Behav., Social Netw.*, vol. 22, no. 1, pp. 82–96, Jan. 2019.

[25] F. Pianzola, G. Riva, K. Kukkonen, and F. Mantovani, "Presence, flow, and narrative absorption: An interdisciplinary theoretical exploration with a new spatiotemporal integrated model based on predictive processing," *Open Res. Eur.*, vol. 1, p. 28, Mar. 2021.

[26] B. Witmer and M. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence, Teleoperators Virtual Environ.*, vol. 7, no. 3, pp. 225–240, Jun. 1998, doi: 10.1162/105474698565686.

[27] S. Bocking, A. Gysbers, W. Wirth, C. Klimmt, T. Hartmann, H. Schramm, J. Laarni, A. Sacau, and P. Vorderer, "Theoretical and empirical support for distinctions between components and conditions of spatial presence," in *Proc. 7th Int. Workshop Presence-Presence*, 2004, pp. 224–231.

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

[29] R. Skarbez, F. P. Brooks, and M. C. Whitton, "Immersion and coherence: Research agenda and early results," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 10, pp. 3839–3850, Oct. 2021.

[30] C. Dillon, E. Keogh, and J. Freeman, "It's been emotional': Affect, physiology, and presence," in *Proc. 5th Annu. Int. Workshop Presence, Porto, Portugal*, 2002, pp. 1–10.

[31] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, "Physiological measures of presence in stressful virtual environments," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 645–652, Jul. 2002.

[32] M. P. Huang and N. E. Alessi, "Presence as an emotional experience," in *Medicine Meets Virtual Reality*. Amsterdam, The Netherlands: IOS Press, 1999, pp. 148–153.

[33] J. Freeman, S. Avons, R. Meddis, D. Pearson, and W. I. Jsselsteijn, "Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli," *Presence*, vol. 9, no. 2, pp. 149–164, Apr. 2000.

[34] S. Nichols, C. Haldane, and J. R. Wilson, "Measurement of presence and its consequences in virtual environments," *Int. J. Hum.-Comput. Stud.*, vol. 52, no. 3, pp. 471–491, Mar. 2000.

[35] J. Van Baren and W. I. Jsselsteijn, "Measuring presence: A guide to current measurement approaches," Inf. Soc. Technol., Eindhoven Univ. Technol., The Netherlands, Tech. Rep., IST-2001-39237, 2004. [Online]. Available: https://cordis.europa.eu/project/id/IST-2001-39237/

[36] G. Goncalves, P. Monteiro, H. Coelho, M. Melo, and M. Bessa, "Systematic review on realism research methodologies on immersive virtual, augmented and mixed realities," *IEEE Access*, vol. 9, pp. 89150–89161, 2021.

[37] M. Slater, M. Usoh, and A. Steed, "Depth of presence in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 3, no. 2, pp. 130–144, 1994.

[38] 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, 2001.

[39] J. Vasconcelos-Raposo, M. Bessa, M. Melo, L. Barbosa, R. Rodrigues, C. M. Teixeira, L. Cabral, and A. A. Sousa, "Adaptation and validation of the igroup presence questionnaire (IPQ) in a Portuguese sample," *Presence, Teleoperators Virtual Environ.*, vol. 25, no. 3, pp. 191–203, Dec. 2016, doi: 10.1162/PRES_a_00261.

[40] (2016). *Igroup Presence Questionnaire (IPQ) Overview*. [Online]. Available: http://www.igroup.org/pq/ipq/index.php

[41] J. Dawes, "Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales," *Int. J. Market Res.*, vol. 50, no. 1, pp. 61–104, 2008.

[42] L. Barbosa, P. Monteiro, M. Pinto, H. Coelho, M. Melo, and M. Bessa, "Multisensory virtual environment for firefighter training simulation: Study of the impact of haptic feedback on task execution," in *Proc. Encontro Português de Computaçáo Gráfica e Interação (EPCGI)*, 2017, pp. 1–7.

[43] M. Melo, J. Vasconcelos-Raposo, and M. Bessa, "Presence and cybersickness in immersive content: Effects of content type, exposure time and gender," *Comput. Graph.*, vol. 71, pp. 159–165, Apr. 2018.

[44] H. Coelho, M. Melo, L. Barbosa, J. Martins, M. S. Teixeira, and M. Bessa, "Authoring tools for creating 360 multisensory videos—Evaluation of different interfaces," *Exp. Syst.*, vol. 38, no. 5, Aug. 2021, Art. no. e12418.

[45] 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.

[46] 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. Hum.-Mach. Syst.*, vol. 51, no. 2, pp. 65–74, Apr. 2021.

[47] P. Monteiro, D. Carvalho, M. Melo, F. Branco, and M. Bessa, "Application of the steering law to virtual reality walking navigation interfaces," *Comput. Graph.*, vol. 77, pp. 80–87, Dec. 2018.

[48] G. Gonçalves, M. Melo, L. Barbosa, J. Vasconcelos-Raposo, and M. Bessa, "Evaluation of the impact of different levels of self-representation and body tracking on the sense of presence and embodiment in immersive VR," *Virtual Reality*, vol. 26, no. 1, pp. 1–14, Mar. 2022.

[49] M. Melo, T. Rocha, L. Barbosa, and M. Bessa, "Presence in virtual environments: Objective metrics vs. subjective metrics—A pilot study," in *Proc. 23rd Portuguese Meeting Comput. Graph. Interact. (EPCGI)*, 2016, pp. 1–6.

[50] D. Narciso, M. Melo, J. Vasconcelos-Raposo, and M. Bessa, "The impact of olfactory and wind stimuli on 360° videos using head-mounted displays," *ACM Trans. Appl. Perception*, vol. 17, no. 1, pp. 1–13, Feb. 2020, doi: 10.1145/3380903.

[51] G. Goncalves, P. Monteiro, M. Melo, J. Vasconcelos-Raposo, and M. Bessa, "A comparative study between wired and wireless virtual reality setups," *IEEE Access*, vol. 8, pp. 29249–29258, 2020.

[52] G. Goncalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa, "Correlation between game experience and presence in immersive virtual reality games," in *Proc. Int. Conf. Graph. Interact. (ICGI)*, Nov. 2019, pp. 107–114.

[53] G. Goncalves, 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.

[54] D. Narciso, M. Melo, J. V. Raposo, J. Cunha, and M. Bessa, "Virtual reality in training: An experimental study with firefighters," *Multimedia Tools Appl.*, vol. 79, no. 9, pp. 6227–6245, 2020.

[55] D. Narciso, M. Bessa, M. Melo, and J. Vasconcelos-Raposo, "Virtual reality for training—The impact of smell on presence, cybersickness, fatigue, stress and knowledge transfer," in *Proc. Int. Conf. Graph. Interact. (ICGI)*, Nov. 2019, pp. 115–121.

[56] M. Bessa, H. Coelho, M. Melo, and M. Pinto, "Impact of different display devices and types of virtual environments on emotions and feeling of presence," in *Proc. Encontro Portugues de Computação Gráfica e Interação (EPCGI)*, 2017, pp. 1–7.

[57] M. Pinto, M. Melo, and M. Bessa, "Use of the physiological response to improve the gaming experience," in *Proc. Int. Conf. Graph. Interact. (ICGI)*, 2018, pp. 1–7.

[58] M. Bessa, M. Melo, A. A. D. Sousa, and J. Vasconcelos-Raposo, "The effects of body position on reflexive motor acts and the sense of presence in virtual environments," *Comput. Graph.*, vol. 71, pp. 35–41, Apr. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0097849317301863

[59] M. Magalhaes, M. Melo, M. Bessa, and A. F. Coelho, "The relationship between cybersickness, sense of presence, and the users expectancy and perceived similarity between virtual and real places," *IEEE Access*, vol. 9, pp. 79685–79694, 2021.

[60] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *Int. J. Hum.-Comput. Interact.*, vol. 24, no. 6, pp. 574–594, 2008, doi: 10.1080/10447310802205776.

[61] J. Brooke, "SUS—A quick and dirty usability scale," in *Usability Evaluation in Industry*, vol. 189, no. 194, 1996, pp. 4–7.

[62] J. Sauro, *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Denver, CO, USA: Measuring Usability LLC, 2011.

[63] J. Clement. (2020). *Share of Gamers Worldwide as of November 2020, by Gender and Age*. [Online]. Available: https://www.statista.com/statistics/272327/mobile-social-and-traditional-gaming-users-by-gender/

[64] M. Koch, K. Von Luck, J. Schwarzer, and S. Draheim, "The novelty effect in large display deployments—Experiences and lessons-learned for evaluating prototypes," in *Proc. 16th Eur. Conf. Comput.-Supported Cooperat. Work-Explor. Papers*, 2018, pp. 1–19.

[65] H. Regenbrecht and T. Schubert, "Real and illusory interactions enhance presence in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 11, no. 4, pp. 425–434, Aug. 2002.

[66] A. A. Kinkhabwala, Y. Gu, D. Aronov, and D. W. Tank, "Visual cue-related activity of cells in the medial entorhinal cortex during navigation in virtual reality," *Elife*, vol. 9, Jan. 2020, Art. no. e43140.

[67] A. Dietrich, "Neurocognitive mechanisms underlying the experience of flow," *Consciousness Cognition*, vol. 13, no. 4, pp. 746–761, Dec. 2004.

[68] D. R. Gruber, *Brain Art and Neuroscience: Neurosensuality and Affective Realism*. London, U.K.: Routledge, 2020.
[69] S. Bangay and L. Preston, ''An investigation into factors influencing immersion in interactive virtual reality environments,'' in *Virtual Environments in Clinical Psychology and Neuroscience*. Amsterdam, The Netherlands: IOS Press, 1998, pp. 43–51.
[70] M. Schuemie, B. Abel, C. Van Der Mast, M. Krijn, and P. Emmelkamp, ''The effect of locomotion technique on presence, fear and usability in a virtual environment,'' in *Euromedia*. Ghent, Belgium: Eurosis-Eti, 2005.
[71] T. L. Mitzner, S. A. McGlynn, and W. A. Rogers, ''Understanding spatial presence formation and maintenance in virtual reality for younger and older adults,'' *Gerontechnology*, vol. 20, no. 2, pp. 1–14, Jan. 2021.

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

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

JOSÉ VASCONCELOS-RAPOSO is a Full Professor with the Education and Psychology Department, University of Trás-os-Montes e Alto Douro, specializing in performance psychology. He is also a Senior Researcher with INESC TEC. His research interest includes the use of technology to improve human performance applied in sports and health care.

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