Pre-print

Enhanced Presence Evaluation in Virtual Reality Feedback System with TOPSIS Model

Shuo Li[a], Hongtao Zheng[a], Ting han[a], Jun Hu[b], Chenqi Zhang[a] and Cong Yu[a]

[a]Department of Design, Shanghai Jiao Tong University, Shanghai, China;[b]Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands

ARTICLE HISTORY
Compiled July 18, 2023

ABSTRACT
Researcher has been trying to optimize the method for evaluating presence in virtual reality (VR) to address variability and uncertainty in quick evaluations using questionnaires. We recommend using the Technique for Order Performance by Similarity to Ideal Solution (TOPSIS) model to calculated the iGroup Presence Questionnaire to measure users' overall presence in VR prototyping, instead of the weighted sum approach. The effects of two presence factors (scene density and motion trajectory technology) on a self-developed VR swimming virtual system were tested using the TOPSIS model with 20 participants each 12 tasks in a user experiment. The results were compared using two different weighting methods, fuzzy hierarchical analysis and uniform weighting methods. TOPSIS had a narrower range of data within the 95% confidence interval and a significantly lower coefficient of variation (CV). This indicates enhanced precision in evaluating presence and can be used to compare different technique setting of virtual systems.

KEYWORDS
Virtual reality; presence evaluation; fuzzy analytic hierarchy process; human computer interaction; technique order performance by similarity to ideal solution

1. Introduction

Virtual reality (VR) technologies provide users with immersive and realistic experiences by enabling them to actively interact with the virtual environment(VE). Presence, defined as the level of similarity that a person experiences between the real and VEs, is vital for a VR system. In prior VR evaluations, several studies have evaluated the effect of presence to measure the quality of the VE and discussion the relationship between the devices (such as Head-mounted displays, HMD) of VR and users. In human–computer interaction applications such as driving simulators (Weidner et al., 2017), autonomous driving agents (Grasso et al., 2020), and trainer flight simulators(Matthews et al., 2020), the user must maintain a high sense of presence obtain more proximate to the actual situation in the simulation. Identically, a high level of presence can be established to improve the task performance in VEs (Ariza N. et al., 2017) for purposes such as physical training (Rose et al., 2000) and rehabil-

CONTACT Ting Han (hanting@sjtu.edu.cn). Design Department, Shanghai Jiao Tong University, 800 Dongchuan RD, Minhang District, Shanghai, China; Jun Hu (j.hu@tue.nl). Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands.

itation/clinical exercises (Farrow et al., 2019; Mazzone et al., 2013). Additionally, presence is significantly influenced by the emotional variations that can be elicited by alterations in the VE (Diemer et al., 2015; Shiban et al., 2016).

Although studies have established that VEs are positively correlated with changes in mood, such as scene density (Felton and Jackson, 2022) and haptic immersion techniques (Bailenson and Yee, 2008; Sevinc and Berkman, 2020), large-sample user testing is a challenge for stable presence evaluation due to the dynamic and non-linear changes in environmental characteristics (Felton and Jackson, 2022). In particular, small-sample test data can lead to instability and uncertainty in presence evaluation during the iterative development of VR systems. Therefore, in order to effectively guide the design of VEs, an objective and quantitative method for presence evaluation is required.

This study is based on the IPQ developed by Schubert and Friedmann (2001).The technique order performance by similarity to ideal solution (TOPSIS) method was utilized to improve the measurement precision of presence, and compared the strengths and weaknesses of the presence evaluation of various VR design solutions. Based on the initial exploration conducted using TOPSIS, we assumed that the selection of the weighting method would influence the evaluation results. Thus, we introduced fuzzy Analytic Hierarchy Process (AHP) to compare the measurement precision. The TOPSIS method was applied based on two distinct weighting techniques, and an IPQ score was derived to investigate its effects on precision improvement. Through this study, a valid assessment criterion was established for VR designers to optimize their solutions.

This study is not to develop a new evaluation model based on the IPQ evaluation method and to deduce a mathematical tool for measuring presence, but rather, to use existing optimization models to improve the quality of presence evaluation. The improvement would enable VR designers and experimenters to more precisely compare the variations among different settings and optimize their applications. In particular, the weighting factors and decision-making methods were employed to evaluate the IPQ scores and improve the interpretability in the comparison.

For conducting the experimental case study, we designed an immersive water-free swimming simulation system, as explained in Section 4. The validity of the presence evaluation and results of various weighting sum methods were analyzed, and the TOPSIS calculation results for two weighting methods are discussed in Section 5. The coefficient of variation of the IPQ data and the factors influencing presence are discussed and elaborated in Section 6. The concluding remarks on the advantages of the IPQ-based new framework are summarized in Section 7.

2. Background

2.1. Dimensions of presence

Presence refers to the feeling of "being there" in a mediated environment, which is a subjective experience and is limited to a specific medium, such as VR technology. Lombard *et al.* (1997) defined it as "the perceptual illusion of non-mediation," where individuals perceive or recognize their existence in a technologically mediated experience but are unsure how to respond. However, in VEs, there is an increasing tendency to use technological means to create more realistic environments, which Lombard *et al*. (2015) defined as spatial presence. Slater (2016; 2009; 2018) conceptualized it as a "sense of being in a real place" or a "sensation of being in a real location". Spatial presence is a decisive characteristic of VR experiences, and inducing a complete sense of spatial presence to make the VE indistinguishable from reality is a significant goal of VR technology (Mazuryk and Gervautz, 1996; Nash et al., 2000; Felton and Jackson, 2022).

In VEs, the factors that determine presence are diverse, just as the real world provides us with complex sensory stimuli. Felton (2022) proposed an inclusive classification of the origins of virtual presence, which extensively lists the determinants of presence. These determinants include two external factors (sensory determinants and content determinants) and three internal factors (psychological determinants, demographic determinants, and cultural determinants). External factors primarily encompass sensory variables related to multi-sensory interaction, such as display field-of-view (visual immersion), display resolution, perception of depth, haptic feedback, auditory cues, olfactory cues, and head-tracking (Slater et al., 2003; Kaul et al., 2017). Additionally, the realism of the VE and narrative elements are considered as internal variables. The other three internal factors are inter-subjective variable that vary due to individual differences among participants.

2.2. Presence measures and evaluation

Measuring virtual presence is complex, to quantify presence, researchers have attempted to identify the measurable subjective and objective criteria (Feng et al., 2020). Riva (2003) have reviewed subjective measures, behavioral metrics, and physiological identification of measuring presence in VEs. There is consensus in existing studies that objective stimuli in a VR environment can generate physiological and behavioral responses that are strongly correlated with presence (Hale and Stanney, 2014). Insko (2001) tracked the reflex movements to assess participants' reactions toward unexpected events, because the reaction-evoking behaviors are associated with stressors indicating the level of presence. Biological measures such as skin conductance, heart rate, eye tracking, and electromyography (EMG), were investigated as presence indicators. Although the correlation between the objective variables and sense of presence has been established, the application of objective measures is highly affected by the signal noise induced by the movements of participants (Clemente et al., 2014). In addition, the objective measurement method poses several limitations related to precision, sensor performance (Qiu et al., 2022; Zheng and Li, 2022), and the possibility of errors caused by the interference with the signal.

Currently, subjective measures still primarily rely on questionnaires and interviews, such as the Slater–Usoh–Steed (SUS) (Slater et al., 1994), presence questionnaire (PQ) (Witmer and Kline, 1998) and other researchers summarized some relevant standardized questionnaires (Qiu et al., 2020, 2023) . In addition, Schloerb and Stanney employed matching comparisons (Schloerb, 1995) and cross-modal matching to assess the users' subjective perceptions of presence (Stanney et al., 1998). Among the well-known questionnaires, the SUS contains six questions (7-point Likert scale) regarding the participants' perceptions of presence and limitations in a VE. The SUS measures the extent to which participants conceptualize a given VE (Slater and Wilbur, 1997). Moreover, PQ is an assessment tool that validates four impact factors on presence using a structural equation approach, including control, sensory, distraction, and realism factors. The tool evaluates the sum of presence perceptions using a 7-point Likert scale (Witmer and Kline, 1998). Schubert and Friedmann developed the iGroup presence questionnaire (IPQ) based on the embodied cognition framework (Lako, 1987), and prior research has been conducted on the PQ and SUS. The presence evaluation scale contains three separate multidimensional components of the 14 items (Schubert et al., 2001). The three independent variables were considered in the IPQ: subjective experience in a virtual space, subjective experience of technology and interaction possibilities, and subjective experience of authenticity.

Based on the results reported by Souza (2021), it is evident that the IPQ method was employed in 31 presence measurement tests of VEs to obtain suitable test results. In contrast to other assessment methods such as PQ and SUS, the IPQ applies four-dimensional visualization technology, as illustrated by the Presence Profile. Recent studies have also pointed out that IPQ is an effective tool for presence evaluation in VEs (Strojny et al., 2022). In particular, they used items including Involvement (INV), Experimental Realism (REAL), and Spatial Presence (SP) to form 3D coordinates, and subsequently, combined them with the range of variations in the Sense of Being There (G1) to elicit the level of the perceived presence of the VEs. For comparative analysis, other researchers have utilized a unified weight calculation method based on an expert system perspective and aimed to compare the VE solutions, analyze the defects in VE design (Zhang et al., 2019; Akdere et al., 2021), and improve the performance of VR systems (Chang et al., 2022). Although IPQ can be conveniently implemented, which yields inaccurate results owing to variations in the subjective perceptions across individuals because of psychological, social, and environmental differences (Casner and Gore, 2010). Therefore, the method is a needed to ensure relatively small sample sizes with satisfactory precision of presence measurements (Hennink and Kaiser, 2022).

2.3. Fuzzy analytic hierarchy process and TOPSIS

The method we develop is different from the unified weight calculation method of expert systems. Based on Hart weighting method (Hart, 2006), we try to compare the different factors of IPQ in paired comparison. This method determines which factors in the VE are more important to the sense of presence, so as to adapt to different VEs development. However, in practice, if participants subjectively judge which factors in IPQ are more important than another, this method introduces uncertainty and can become inaccurate. Instead we introduce the semantic scale (Yu, 2002a) and the fuzzy method of pairwise comparison (Deng, 1999), and use the fuzzy analytic hierarchy process proposed by Chang to determine the weight coefficients of different factors in IPQ. Here we briefly describe the basic concept behind the fuzzy analytic hierarchy process. More theoretical details can be found in the literature (Chang, 1996).

For instance, complex issues like sense of presence (goal layer) can be classified into several levels of criteria, and the indicator on level one can be decomposed into multiple levels with different indicators. This can be further decomposed into Presence in VR (Fig 1). Although the number of criteria for assessment is continuously expanding, the refinement of these criteria exponentially reduces their total number for comparison owing to the hierarchical structure(Harris et al., 2020).

To improve the precision and quality of presence measurements, related studies suggested the Technique for Order Performance by Similarity to Ideal Solution (TOPSIS) as a practical approach for considering the uncertainty inherent in evaluation (Zhou and Chan, 2017; Yoon and Hwang, 1995). Notably, the TOPSIS model assists decision

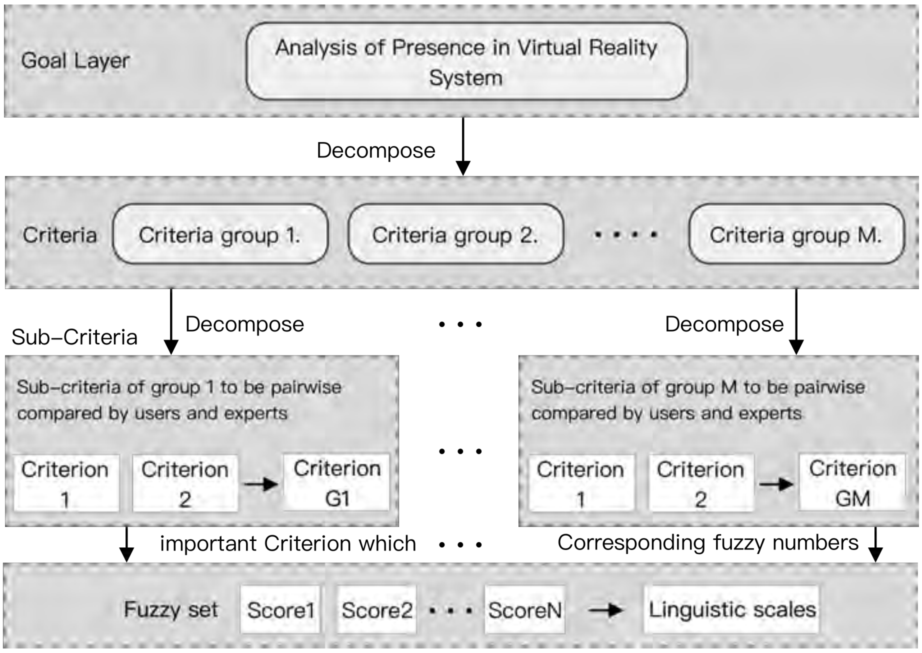

Figure 1. General hierarchical structure of Presence in VR; this structure can include related criteria from multiple groups or sources.

makers in analyzing and comparing the relative performance of multiple scenes, and thus, it has been widely applied in several complex domains, including project risk assessment in finance, patient safety assessment in health management, manufacturing system design, business management, and energy (Behzadian et al., 2012).

In summary, the aim of our study is to test the effect of different VEs on user immersion in the context of a small sample size, based on the definition of spatial presence, considering visual immersion variables and haptic feedback variables among external determinants, using the TOPSIS method as a new way to calculating overall presence in VR applications, introducing FAHP and weighted sum methods, and comparing the precision of presence measures.

3. Methods for computing the presence score

In this study, we will focus on discussing our method based on the IPQ table. Firstly, based on the existing IPQ table, a analytic hierarchy process (AHP) process was employed to improve the measurement quality of the VR presence to analyze and compare the degree of effect of the extra weight factors. Subsequently, the fuzzy AHP employed additional factor weights for each criterion to distinguish the significance level of various environmental variables, evaluate the overall presence, and modify the deficiencies in subjective evaluation methods. In Fig.2, we illustrate an overview of the presence evaluation in TOPSIS based on varying weighting methods.The framework can be divided into three steps. In the first step, we involved experts and evaluators in VR tasks to obtain IPQ data. Subsequently, we analyzed the data to obtain a Presence Profile. Thereafter, it applied uniform weighting coefficients to calculate the overall IPQ values. In step two, the weighting coefficients were determined through pairwise comparisons in accordance with fuzzy analytic hierarchy process recommendations. In the third step, the TOPSIS model computed the relative closeness coefficients based on the weights obtained in the first and second steps.

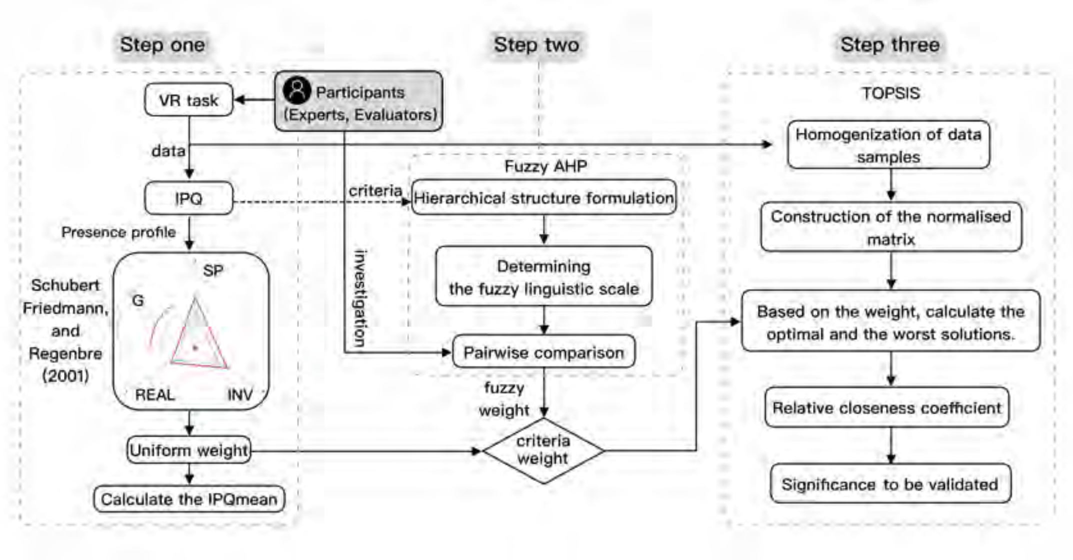

Figure 2. IPQ measurement method and schematic process the TOPSIS method use different weighting approaches.

3.1. Weighted sum method in IPQ

The calculation of the overall presence of IPQ has two steps. First, the participants of VE must perform a pairwise comparison of the criteria provided in the IPQ according to the conducted and experienced task. For instance, if the user perceives that Involvement is more important than Experienced realism, then the score (score on their relative importance) of Involvement is incremented by one, whereas that of Experienced realism remained unchanged. Second, after six comparisons, the weighting coefficients for the four criteria were obtained by normalization. These comparisons were applied to determine the weighting coefficients of each criterion. Thereafter, the overall presence score was computed as the sum of the scores for each criterion weighted by their respective weighting coefficients.

$$IPQ_{wsm} = W_{G1} * G1 + W_{SP} * SP + W_{INV} * INV + W_{REAL} * REAL \qquad (1)$$

where, W_{G1}, W_{SP}, W_{INV}, and W_{REAL} are denote corresponding weight coefficients for each criterion. In the overall evaluation, IPQ_{mean} can be regarded as a special case, indicating that W_{G1}, W_{SP} , W_{INV}, and W_{REAL} are equal to 25%

3.2. Analytic hierarchy process and fuzzy set

Herein, the AHP method was applied to extract various criteria at multiple levels in terms of IPQ. In practice, the process of comparison and decision making is associated with the strong vagueness of human thinking. Generally, evaluators partially or completely associate pairwise comparison values with an uncertainty degree instead of precise ratings, and such an uncertainty degree is represented by an appropriate semantic scale Yu (2002b). To this end, triangular fuzzy numbers are utilized for a pairwise comparison scale of the AHP.

3.2.1. Fuzzy numbers and a fuzzy synthetic extent analysis

Triangular fuzzy numbers (TFNs) were applied to transform the pairwise comparison table data into TFNs data. Thus, the membership function is defined by the following equation:

$$\mu_M(x) = \begin{cases} \frac{x-l}{m-l} & x \in [l, m] \\ \frac{x-u}{m-u} & x \in [m, u] \\ 0 & otherwise \end{cases} \tag{2}$$

where μ_M(x): $\mathbb{R} \rightarrow$[0,1] and l, m, and u denote the lower, modal, and upper values of the TFN, respectively.

According to the triangular fuzzy comparison matrix $\widetilde{A}$=(a$_{ij}$)$_{n\times n}$, the extent analysis sums up each row of this matrix, normalizes the sums with respect to the ith, and ultimately, calculates the amount of overlap.

$$S_i = \sum_{j=1}^{n} a_{ij} \otimes \left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\right]^{-1} \tag{3}$$

where a_{ij}=(l_{ij},m_{ij},u_{ij}) denote the triangular fuzzy number. According to the rules of operation, S_i defines a triangular fuzzy number. There are two triangular fuzzy numbers: S_1=(l_1,m_1,u_1) and S_2=(l_2,m_2,u_2). Upon comparing the degree of possibility, both S_2 and S_1 can be defined as follows:

$$V(S_2 \geq S_1) = \begin{cases} 1 & m_2 \geq m_1 \\ 0 & l_1 \geq u_2 \\ \frac{l_1-u_2}{(m_2-u_2)-(m_1-l_1)} & otherwise \end{cases} \tag{4}$$

Furthermore, to comparatively evaluate whether a convex fuzzy number S is more significant than k convex fuzzy numbers, S_i, *i=1,2,...k* can be defined as follows *minV* ($S \geq S_i$).

3.2.2. Procedures to fuzzy AHP method

Step 1:Presence analysis and hierarchical structure formulation. As outlined in Section 2.2, the Analytic Hierarchy Process (AHP) methodology developed by Saaty provides a standard for conducting hierarchical structure formulation. We can consider the study of presence in VR as a specific case of AHP, and structure it accordingly. Specifically, the goal layer is the problem of presence itself, while the indicator layer consists of factors that influence perceived presence, which can be further decomposed into more specific sub-indicators. In this study, we initially explored presence by identifying four influencing factors as indicators at the criteria layer. Through pairwise comparisons by experts and evaluators of different criterias in the criteria layer, we determined which criterias were more important. The corresponding hierarchical structure of IPQ is illustrated in Fig 3. However, due to the uncertainty of importance, we used a fuzzy linguistic scale to adjust the weightings.

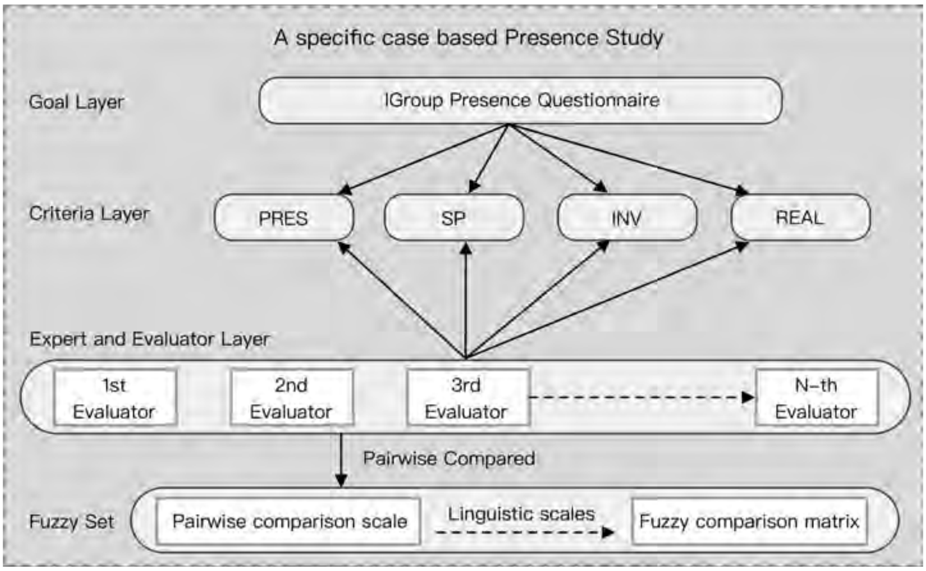

Figure 3. Specific case of hierarchical structure to measure presence with IPQ criteria.

Table 1. Linguistic rating scales and corresponding fuzzy numbers.

Linguistic scales	Score	TFNs
Both equally important	1	(1,1,1)
Weakly more important	2	(1,1,3/2)
Somewhat important	3	(3/2,2,5/2)
Remarkably more important	4	(5/2,3,7/2)
Very remarkably more important	5	(7/2,4,9/2)
Extremely more important	6	(9/2,5,11/2)
Absolute important	7	(11/2,6,13/2)

Step 2: Fuzzy comparison using fuzzy linguistic scale. After completing the VR task and the IPQ table, the experts and evaluators will also be invited to participate in pairwise comparisons table. In a previous study, the fuzzy linguistic scale (Saaty, 1988) was applied to transform uncertain and fuzzy information into a quantifiable form for subjective evaluation. As shown in Table 1, the linguistic scale of importance corresponds to fuzzy numbers. To understand the relative importance of criterion K_i *(i=1,2,...,m)* compared to K_j *(j=1,2,...,m, i≠j)* for presence measurement in a VE, a fuzzy comparison matrix was constructed based on the comparison table (Table 2), denoted as:

$$\widetilde{K} = \begin{pmatrix} 1 & k_{12} & k_{13} & k_{14} \\ k_{21} & 1 & k_{23} & k_{24} \\ k_{31} & k_{32} & 1 & k_{34} \\ k_{41} & k_{42} & k_{43} & 1 \end{pmatrix} \tag{5}$$

Table 2. Fuzzy comparison table to be filled with the linguistic expressions

	G1[a]	SP[b]	INV[c]	REAL[d]
G1	BI[e]			
SP		BI		
INV			BI	
REAL				BI

[a]Sense of being there
[b]Spatial presence
[c]Involvement
[d]Experienced realism
[e]Both equally important

where, $k_{ij} = 1/k_{ji}$, and

$$\begin{cases} k_{ij} = (l_{ij}, m_{ij}, u_{ij}) \\ \frac{1}{k_{ji}} = \left(\frac{1}{u_{ij}}, \frac{1}{m_{ij}}, \frac{1}{l_{ij}}\right) \end{cases} \tag{6}$$

The P evaluators yielded distinct fuzzy comparison matrices to form the final evaluation matrix $\widetilde{K_P}$, and all evaluations were aggregated. The average of all the fuzzy comparison matrices can be expressed as,

$$k_{ij} = \left(\frac{1}{P}\sum_{p=1}^{P} l\left(k_{ij}^{p}\right), \frac{1}{P}\sum_{p=1}^{P} m\left(k_{ij}^{p}\right), \frac{1}{P}\sum_{p=1}^{P} u\left(k_{ij}^{p}\right)\right) \tag{7}$$

where l, m and u represent the functions used for evaluating the lower, modal, and upper values of the TFN.

Step 3: Weighting vector determination. Using the extent analysis method formulated in Eq. 3 and 4, we can describe each criterion using a TFN. In principle, the comparison of fuzzy numbers must be used to determine the weighting vector of the criteria, where $w'(K_i) = \min V(S_i \geq S_j)$ for $j = 1, 2, ...M, i \neq j$,

$$w'_r = \left[w'(K_1), w'(K_2), \ldots, w'(K_M)\right]^T \tag{8}$$

where K_i , $i = 1, 2, ..., M$ reflect the criteria, and the final normalized weighting vector can be expressed as,

$$\begin{aligned} W &= \left[\frac{w'(K_1)}{\sum_{i=1}^{M} w'(K_i)}, \frac{w'(K_2)}{\sum_{i=1}^{M} w'(K_i)}, \ldots \frac{w'(K_M)}{\sum_{i=1}^{M} w'(K_i)}\right]^T \\ &= \left[w(K_1), w(K_2), \ldots, w(K_M)\right]^T \end{aligned} \tag{9}$$

3.3. TOPSIS method

The TOPSIS method utilizes a distance scale to measure differences among samples. To utilize this scale, it is necessary to normalize the index attributes in the same manner. Typically, this step is required in almost all evaluation methods to homogenize the raw data. For the IPQ evaluation of each VR system (object), the original data matrix was homogenized using the cosine distance measure, followed by the computation of the relative closeness coefficient.

3.3.1. Normalized matrix

Let n denote the number of objects to be evaluated; each object contains m attributes. Therefore, the original data matrix can be derived as,

$$O = \begin{pmatrix} o_{1,1} & o_{1,2} & \cdots & o_{1,m} \\ o_{2,1} & o_{2,1} & \cdots & o_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ o_{n,1} & o_{n,2} & \cdots & o_{n,m} \end{pmatrix} \tag{10}$$

To perform dimensionless calculations, we should construct a weighted canonical matrix in which the attributes are normalized vectors; i.e., each column element is segmented by the norm of the current column vector (according to the cosine distance measure):

$$h_{ij} = \frac{o_{ij}}{\sqrt{\sum_{i=1}^{n} o_{ij}^2}} \tag{11}$$

Accordingly, the normalized nondimensional matrix yields into,

$$H = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,m} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n,1} & h_{n,2} & \cdots & h_{n,m} \end{pmatrix} \tag{12}$$

3.3.2. Identification of optimal and worst solutions

There are two idealized goals, namely, the positive ideal goal or optimal goal, and the negative ideal solution or the worst goal. The positive optimal solution H^+ contains the maximum value of each column element in H:

$$H^+ = \begin{pmatrix} \max\left(h_{1,1}, h_{2,1}, \cdots, h_{n,1}\right) \\ \max\left(h_{1,2}, h_{2,2}, \cdots, h_{n,2}\right) \\ \vdots \\ \max\left(h_{1,m}, h_{2,m}, \cdots, h_{n,m}\right) \end{pmatrix} \tag{13}$$

The worst solution can be expressed based on Eq. 14:

$$H^{-} = \begin{pmatrix} \min\left(h_{1,1}, h_{2,1}, \cdots, h_{n,1}\right) \\ \min\left(h_{1,2}, h_{2,2}, \cdots, h_{n,2}\right) \\ \vdots \\ \min\left(h_{1,m}, h_{2,m}, \cdots, h_{n,m}\right) \end{pmatrix} \tag{14}$$

3.3.3. Calculation of separation distance

In principle, the separation distance is utilized to measure the distance between the current sample and the optimal/worst solution, which can be defined as follows:

$$d_i^{+} = \sqrt{\sum_{j=1}^{m} w_j \left(H_j^{+} - h_{i,j}\right)^2} \tag{15}$$

$$d_i^{-} = \sqrt{\sum_{j=1}^{m} w_j \left(H_j^{-} - h_{i,j}\right)^2} \tag{16}$$

where w_j denotes the weight of the jth attribute (importance), and it can be derived from the fuzzy AHP method, as expressed in Eq.9.

As the separation distance is a factor of two independent values (d_i^{+} and d_i^{-}), a unified measurement can be obtained in a single dimension through the relative closeness coefficient (X_i) of each entry, defined as,

$$X_i = \frac{d_i^{-}}{d_i^{+} + d_i^{-}}, \quad X_i \in [0, 1] \tag{17}$$

If $X_i \to 1$, d_i^{+} is small and d_i^{-} is large, implying that the measured object approaches the most optimal performance. The performance of each object can be defined by the values of X, ranking from large X_i to small X_i or in reverse order.

Applying this principle to the specific case based presence study, a larger X value suggests higher presence. In addition to the observation of the performance of each individual object, X can be grouped according to the factors tested in the experiment. Moreover, the variations in X can be compared among groups (e.g., the sense of presence created by varying physical feedback designs).

4. Case study

Herein, a case study was conducted to evaluate the effectiveness of the proposed method in improving the precision of the presence evaluation. A VR-based swimming simulator was developed to provide a vivid water-surface environment. In particular, two factors were predefined for designing the VR simulation: (a) immersive physical feedback (Jahn et al., 2020; Sevinc and Berkman, 2020); (b) scene density (Parsons et al., 2009). Thereafter, we conducted experiments based on the proposed modified-IPQ method to examine the influence of these two factors on the assessment of presence.

4.1. Experimental setup of the swimming simulator

The developed virtual swimming simulator (Li et al., 2022) features two motion track technologies (MTT) to compare the impact of immersive feedback technology (Bailenson and Yee, 2008) on user presence. We have designed a swimming simulation VR application with different scene densities for comparative analysis. Users can receive real-time feedback on their swimming posture through the HTC VIVE headset (display resolution: 2160 × 1200 px; refresh rate: 90hz). Additionally, the Head-mounted displays includes a breathing simulation device to adjust breathing speed. Furthermore, a metal protective frame with elastic force hardness was constructed to support the body in swimming postures and fluctuate the body by simulating floating and waves.

4.1.1. Tracker and Kinect motion

We explored the impact of immersive feedback technology and made different attempts in the virtual swimming simulator. Two motion tracking technologies were used: HTC Vive's Tracker and Kinect camera. In *MTT Scheme* A (MTT A), the Tracker (FOV is 240°) was worn on the participants' wrists and ankles to enable real-time tracking of their body movements. In contrast, *MTT Scheme B* (MTT B) used Kinect (depth resolution of 512 × 424 and 30 fps) to capture the swimming movements of the participants. More importantly, a posture recognition and matching module was developed, using a 10-point positioning method to facilitate real-time synchronization of human movements and virtual character movements in the HMD, including the simulation rendering of standard actions.

4.1.2. Respiration simulator

The HMD was equipped with a respiration simulator to regulate the respiratory rate. Additionally, a card slot structure was designed to attach the respiratory rate bags to the HTC VIVE VR headset. The bags was connected to the air pump through a Y-type tee-joint, and communication control was carried out through Arduino and WiFi. The breathing rate was synchronized with the task movements rate (Fig.4).

4.1.3. Elastic force harness and support device

To simulate the resistance generated by the arm action during swimming, we designed a wearable elastic force hardness device. The user was equipped with the harness that was attached to *Point A* in the support device to effectively simulate the resistance during the stroke. According to a swimming resistance study by Bixler and Riewald(Bixler and Riewald, 2002), the resistance faced by a standard-weight individual during swimming in water was approximately 24–40 N. Specifically, each harness was designed to surround the body in a direction without impacting mobility (Fig.5).

4.1.4. User interface for HMD

We designed a user interface for our virtual swimming simulation program, which includes a standard swimming pose example window, a real-time action feedback win-

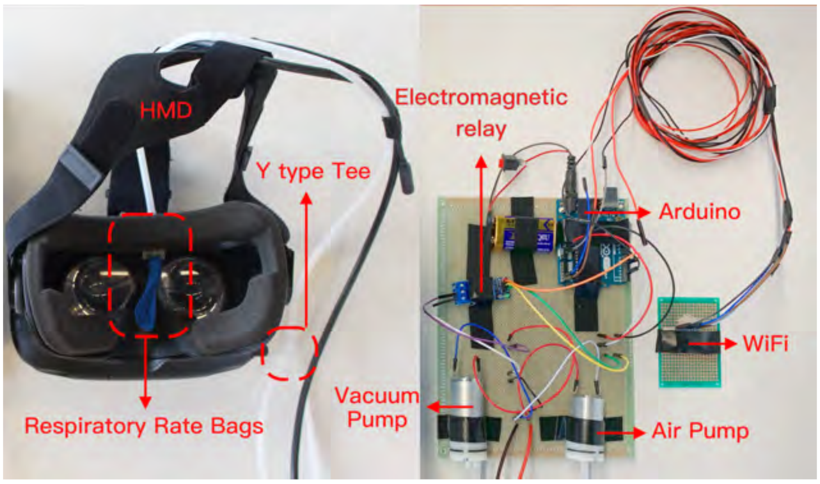

Figure 4. Respiration Simulator Details

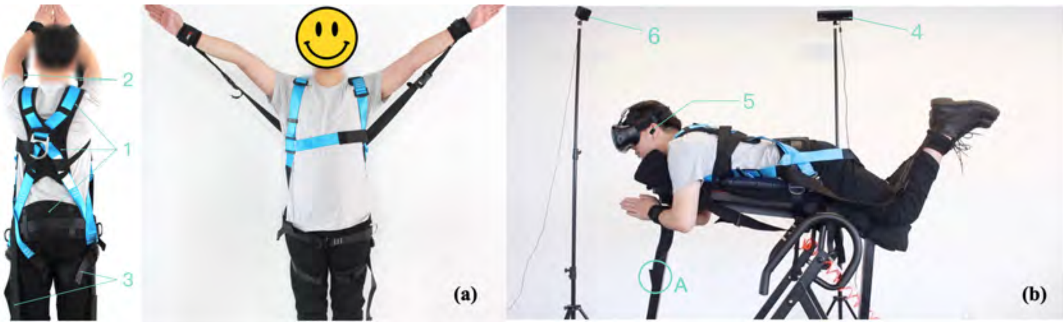

Figure 5. (a) Elastic force harness for limbs adheres to skin and includes a shoulder guard (1) and a soft pad for waist (1), whereas wrist guard (2) and ankle of limbs (3) are connected to back pad (1) via corresponding elastic force harness. (b) Support Device Diagram. Point A is the fixed point of the elastic force harness and bracket to ensure that the left and right hands are subjected to the same tension at the limit position. In addition, the device also includes Kinect (4), active noise-cancelling headphones (5) and locator for VR glasses(6).

dow, and a first-person perspective window (Fig.6). In the left window, users can learn the standard swimming poses and movements. In the right window, users can compare their real-time actions with the standard ones. Whenever the user's action exceeds the standard range, the program will display the incorrect action in red, reminding the user to correct it (Fig.6b). Additionally, we generated three different scenes with varying levels of detail by changing the textures and details of the water and environment in the VE, namely low (few objects), middle, and high (several objects). According to the definition provided in Section 2, we used this setup to differentiate users' sense of presence at different levels(Fig.7). For hearing, we used active noise-cancelling headphones to presentation audio signal. The sound density can be classified into three levels: low (minimal details), middle, and high (several details), and the density could be distinguished by the filtration of the sound samples (Cohen-Hadria et al., 2019)(Fig.8c).

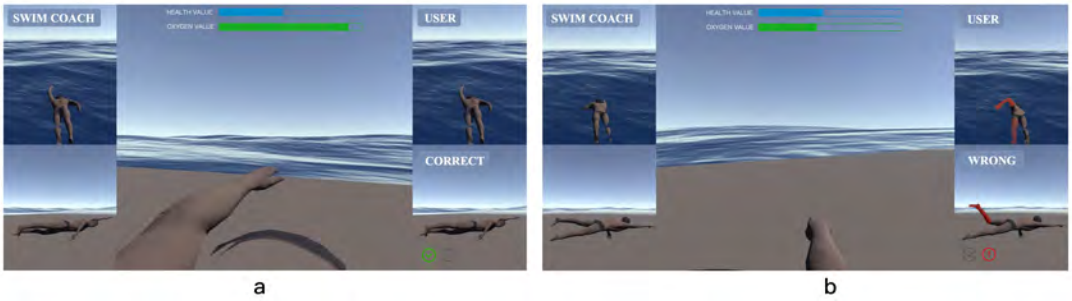

Figure 6. For the user interface with middle density details, we have the following components: (a) correct swimming style. The left column represents an example window of the standard swimming pose. The right column represents an example window of the real-time action feedback window, and Third-person perspective window. (b) Wrong swimming style.

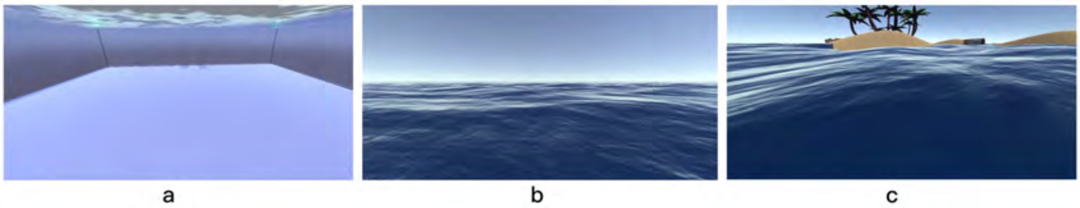

Figure 7. Scene with low(a), middle(b), and high(c) densities respectively.

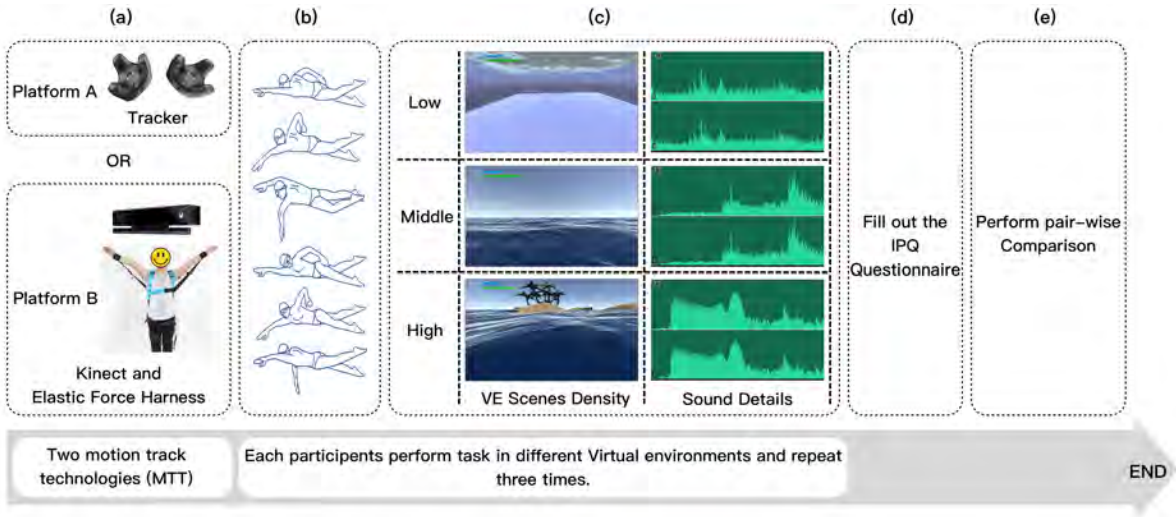

Figure 8. Experiment procedure: (a) each participant is provided an instruction manual of using the Swimming Simulator in various MTTs. (b) six components of freestyle stroke practice. (c) Visual scene and sound samples classified into three levels low (low details), middle, and high (high details). (d) participants were invited to fill out the IPQ questionnaire. (e) participants were invited to complete an individual pairwise comparison matrix.

Table 3. Experimental conditions

Experiment munber	1	2	3	4	5	6
Immersion feedback type	Tracker	Tracker	Tracker	Kinect and force feedback	Kinect and force feedback	Kinect and force feedback
Scene	Low	Middle	High	Low	Middle	High
HMD	HTC VIVE					

4.2. Participants

Twenty healthy participants were recruited to voluntarily participate in the experiment, including three professional swimming athletes and 17 non-athletes. The male-to-female ratio was 1:1 with average age of 25 years (SD = 1.82). Participants were requested to respond to a pre-exposure questionnaire for determining their familiarity with VR technology and acquiring the necessary health and background information. These participants were chosen because of their experience in the use of VR equipment. Moreover, the ethical considerations were upheld with written consent from the participants.

4.3. Experiment procedure

The experimental procedure is shown in Fig.8 and consists of five steps. Firstly, the participants responded to the pre-exposure questionnaire and were asked about their experience with VR. Subsequently, with the assistance of the experimenter, participants put on the equipment and were immersed in the VE. Secondly, participants practiced the stroke movements according to the predetermined stroke practice procedures. Specifically, users had to practice six components of freestyle and breath movements based on the standard movements in the left window (Fig. 8b). In the right window, if the movement was evaluated as incorrect or significantly deviated, the step would be restarted until the standard was achieved. Finally, when participants accurately completed all movements, they were asked to exit the virtual swimming simulator and fill out an IPQ questionnaire based on their experience and impressions (Fig. 8d). In addition, participants were invited to use the language expressions listed in Table 1 to complete a separate paired comparison matrix (Fig. 8e). To balance the experimental conditions, reduce random errors, and minimize mixing effects, the experimental conditions were set as shown in Table 3. Specifically, the researchers randomly tested six different conditions, each of which was repeated twice by a user on different days (for a total of six days). Each user performed a total of 12 tests (6×2). The procedure of this was that 1) the same swim test movement flow (same task) would effectively balance the cybersickness effect between participants (Fig. 7b)(Sevinc and Berkman, 2020). 2) The test was spread evenly over the 6 days in order to avoid multiple measurements that would not allow subjects to effectively differentiate between scene densities.

4.4. Weighting coefficients and statistical approach

In the final step of the experimental procedure, participants and two additional professors were invited to form a personal paired comparison matrix to derive the fuzzy comparison matrix. These evaluators were selected due to their experience in VR and were able to judge which factors were more important through firsthand experience. The detailed rating process was reported in Gumus (Gumus, 2009). We calculated the fuzzy weights according to the normalised weightings of the aggregated responses were derived from the fuzzy synthetic extent analysis, as described in Section 2.2. Subsequently, using these available weighting coefficients, the relative closeness coefficient X of the IPQ (IPQ_x) was extracted using the two sets of weights following the TOPSIS model through Eq.17. Similarly, according to various weighting coefficients (IPQ_{wsm} see Section 2.1), the IPQ score was evaluated using the weighted sum method. Furthermore, statistical analysis was performed to compare the differences methods in overall presence. However, as supplementary support, we considered the coefficient of variation (CV) in addition to the mean (M) and standard deviation (SD). CV is defined as the ratio of SD to M and represents a statistical measure of the dispersion of the data(Kesteven, 1946; Lovie, 2005). A higher CV corresponds to a higher dispersion and less reliable measurements. A flowchart of the validation process is illustrated in Fig.9.

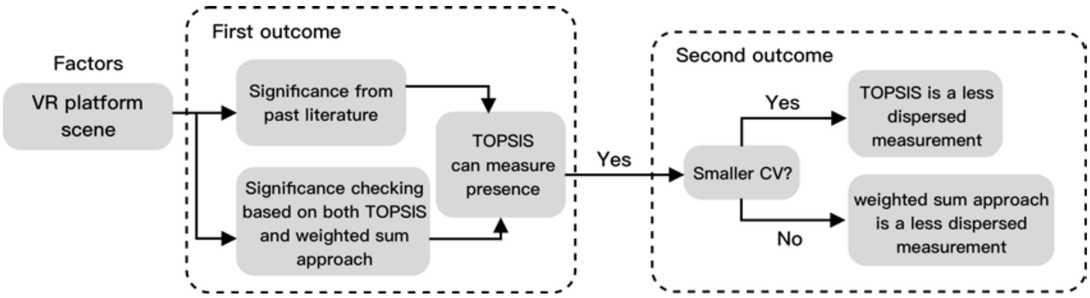

Figure 9. Flowchart of validation process for two factors

5. Results

First, the fuzzy weights were computed according to pair-wise comparisons for input in the IPQ_{wsm} and IPQ_x models. Second, the normalized and dimensionless evaluation data were input into the IPQ_{wsm} and IPQ_x models to evaluate the final score data. The two-way analysis of variance (ANOVA) of the IPQ score data was performed. Third, the coefficient of variation (CV) was computed based on the SD and M of the score data. Ultimately, a mixed-effect model analysis of the CV was performed.

5.1. Weighting coefficients

According to Eqs. 6, 8, 9, and 10, the pairwise comparison data were calculated as fuzzy weights. The SP and REAL factors were regarded as the fundamental factors influencing the sense of presence, with weight coefficients of 36.4 and 33.5%, respectively, which were much higher than those of the other two factors. Potentially, this occurred because the participants experienced scenes of varying densities during the experiment, which improved their understanding of SP and REAL, and subsequently, this resulted in the improvement of fuzzy weights for both terms. In comparison, the weight coefficient of G1 was less than the uniform weight (17.6%), whereas the weight of INV was relatively less (12.5%). In our opinion, the low INV resulted because the participants were not exposed to additional modes of interaction during the experiment. In particular, the participants were able to only assess the precision of the

Table 4. Results of data normality test

	Data	p-value[a]
IPQ_{wsm}	Fuzzing weighting	0.062
	Uniform weighting	0.054
IPQ_x	Fuzzing weighting	0.078
	Uniform weighting	0.063

[a]Kolmogorov–Smirnov test

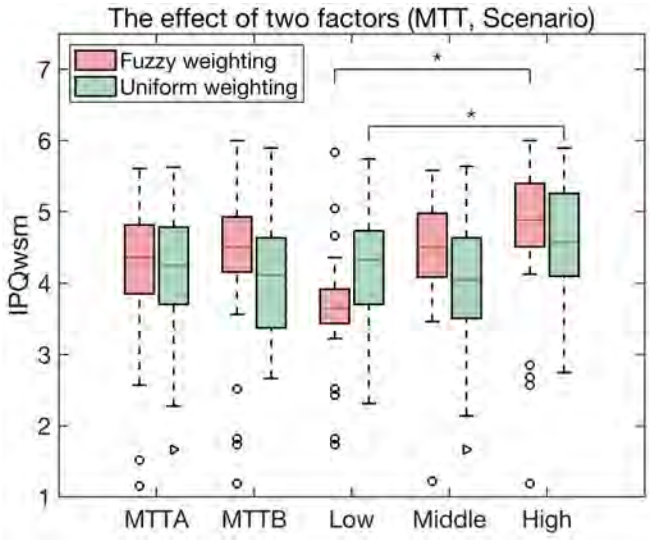

Figure 10. Effects of presence factors on IPQ_{wsm} evaluation

actions based on the task window. Accordingly, the pair-wise comparison table did not indicate that the participants considered the significance of INV factors as more important, which is consistent with the findings reported by Schubert(Schubert et al., 2001).

5.2. Results of weighting sum method and relative closeness coefficient

Additionally, the normality of the data was verified (Table 4), and the two-way repeated measures ANOVA was performed to determine the influence of two independent variables (MTT and scene) on the evaluation of presence, considering both the IPQ_{wsm} and IPQ_x methods, as depicted in (Fig.10 and 11) and Table 5. As illustrated in the figures, we initially considered the mean (M) and standard deviation (SD) obtained for each MTT and the scene type with two distinct weighting approaches and two evaluation methods. We determined that the fuzzing weighting generally yields a smaller standard deviation than uniform weighting, regardless of the IPQ_{wsm} or IPQ_x data.

Based on the IPQ_{wsm} evaluation (Fig.10), the level of scene density significantly influenced only the presence in case of measurement using the fuzzy AHP ($F_{2,38} = 4.25$, $p = 0.03$, $\eta_p^2 = 0.3$) and uniform weight approaches ($F_{2,38} = 2.58$, $p = 0.04$, $\eta_p^2 = 0.2$). In contrast, the VR MTT did not pose any significant effect, regardless of the weighting approach (fuzzy weighting: $F_{2,19} = 3.28$, $p = 0.6$, $\eta_p^2 = 0.1$; uniform weighting: $F_{2,19} = 2.10$, $p = 0.18$, $\eta_p^2 = 0.1$). However, the outcomes revealed an

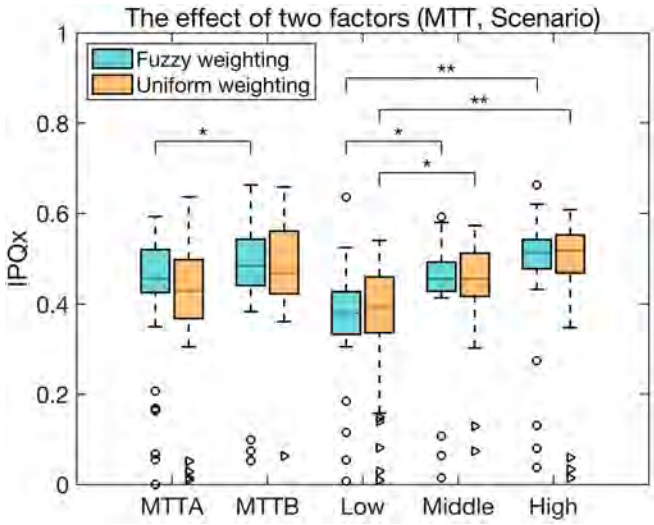

Figure 11. Effects of presence factors on IPQ_x evaluation

Table 5. Effect of factors on presence from various determination methods

	Fuzzy weight			Uniform weight		
IPQ_{wsm}						
Factors	MTT	scene	MTT: scene	MTT	scene	MTT: scene
Sum sq	4.11	15.72	0.83	3.57	12.85	1.54
Df	1	2	2	1	2	2
MS	4.11	7.86	0.42	3.57	6.43	0.77
F	3.28	4.25	1.02	2.10	2.58	2.37
P-value	0.60	0.03*	0.05*	0.18	0.04*	0.08
η_p^2	0.1	0.3	< 0.1	0.1	0.2	< 0.1
IPQ_x						
Factors	MTT	scene	MTT: scene	MTT	scene	MTT: scene
Sum sq	0.0004	0.08	0.0032	0.0008	0.05	0.0054
Df	1	2	2	1	2	2
MS	0.0004	0.04	0.0016	0.0004	0.03	0.0027
F	0.22	3.65	1.41	0.52	4.82	2.01
P-value	0.04*	0.009**	0.02*	0.82	0.03*	1.08
η_p^2	< 0.01	0.04	< 0.1	0.01	0.03	< 0.01

Note: Significance level: .05 (*), .01 (**), Sum sq: Sum square, Df: degree of freedom, MS: mean squared.

interaction effect between the MTT and scene density ($F_{2,38} = 1.02$, $p = 0.05$, $\eta_p^2 < 0.1$) in fuzzy AHP. More importantly, uniform weighting did not exhibit a predominant effect with no observation of additional interactions. In the IPQ_x evaluation (Fig.11), the level of scene density revealed a significant effect for both weighting approaches (fuzzy weighting: $F_{2,38} = 3.65$, $p = 0.009$, $\eta_p^2 = 0.4$; uniform weighting: $F_{2,38} = 4.82$, $p = 0.03$, $\eta_p^2 = 0.3$). In particular, the outcomes revealed a significant variation in the level of VR MTT ($F_{1,19} = 0.22$, $p = 0.04$, $\eta_p^2 < 0.1$) and an interaction effect between the MTT and scene density ($F_{2,38} = 1.41$, $p = 0.02$, $\eta_p^2 < 0.1$). Furthermore, the between-factor MTT and interaction ($F_{1,19} = 0.52$, $p = 0.82$, $\eta_p^2 = 0.1$) did not reveal any observable effect.

Thereafter, post-hoc analyses were performed to understand the variations between multiple levels of scene density. As the data were subjected to multiple pairwise comparisons, we used Tukey's test, and the results are listed in the table 6. Regarding significance (p-value), we considered the mean differences with 95%-confidence intervals. With the IPQ_{wsm} evaluation including two weighting approaches, the statistical significance was observed only between the low and high densities (fuzzy weighting: $p = 0.03$, $95\%CI$: -1.44 to -0.12, $\eta_p^2 = 0.03$; uniform weighting: $p = 0.04$, $95\%CI$: -1.89 to -0.05, $\eta_p^2 = 0.03$). In contrast, the IPQ_x evaluation reported further significant effects. The statistical significance was detected between the low and middle densities (fuzzy weighting: $p = 0.04$, $95\%CI$: -0.06 to -0.009, $\eta_p^2 = 0.01$; uniform weighting: $p = 0.05$, $95\%CI$:-0.05 to -0.004, $\eta_p^2 = 0.02$), as well as between the low and high densities with both the weighting methods (fuzzy weighting: $p = 0.005$, $95\%CI$: -0.07 to -0.03, $\eta_p^2 = 0.04$; uniform weighting: $p = 0.009$, $95\%CI$: -0.06 to -0.01, $\eta_p^2 = 0.04$). Note that the confidence intervals from the IPQ_x are much narrower than those from the IPQ_{wsm}.

5.3. Analyzing coefficient of variation

As listed in Table 6, the range of the confidence intervals was wider than that with the TOPSIS method if the presence was measured using the weighted sum method, implying that the TOPSIS provided more precise estimates than the weighted sum method. As supplementary support, we considered the coefficient of variation in addition to the mean (M) and standard deviation (SD) obtained for each MTT and scene type with the two distinct weighting approaches and two evaluation methods (Table 7). An additional mixed-effects model was employed to analyze the coefficient of variation, where the random effects included the MTT and scene density, and the fixed effects included the evaluation methods (IPQ_{wsm} and IPQ_x) and weighting approaches (fuzzy weighting and uniform weighting). The results of the statistical analyses are presented in Table 8.

Furthermore, we observed a significant deviation between the evaluation methods in terms of the CV, and the weighting approach significantly influenced CV ($F_{1,30} = 0.93$, $p < 0.01$, $95\%CI$: -0.08 to -0.04, $\eta_p^2 = 0.62$). The range of the confidence intervals revealed that IPQ_x produced significantly smaller CVs compared to IPQ_{wsm}, with 95%-confidence intervals ranging from –0.06 to –0.04. Comprehensively, as an exploratory study, we did not detect a difference at the given significance level between fuzzy weighting and uniform weighting, $F_{1,30} = 0.35$, $p > 0.05$, $\eta_p^2 = 0.99$. Therefore, the evaluation from TOPSIS illustrated a significantly lower dispersion of the data compared to the weighted sum method, which confirmed our expectation that the TOPSIS enhanced the precision of the subjective evaluation.

Table 6. Post-hoc analysis for scene type considering various determination methods, "Lower" and "Upper" represent the boundaries of 95%-confidence interval (CI).

	Fuzzy weight			Uniform weight		
IPQ_{wsm}						
Scene type	low	low	middle	low	low	middle
Scene type	middle	high	high	middle	high	high
Mean difference	-0.47	-0.85	-0.31	-0.53	-0.81	-0.28
Lower	-1.25	-1.44	-0.97	-1.47	-1.89	-0.83
Upper	0.23	-0.12	0.36	0.02	-0.05	0.05
P-value	0.11	0.03*	0.34	0.07	0.04*	0.30
η_p^2	0.02	0.03	< 0.01	0.01	0.03	< 0.01
IPQ_x						
Scene type	low	low	middle	low	low	middle
Scene type	middle	high	high	middle	high	high
Mean difference	-0.01	-0.04	-0.01	-0.03	-0.04	-0.02
Lower	-0.06	-0.07	-0.04	-0.05	-0.06	-0.05
Upper	-0.009	-0.03	0.02	-0.004	-0.01	0.01
P-value	0.04*	0.005**	0.39	0.05*	0.009**	0.31
η_p^2	0.01	0.04	< 0.01	0.02	0.04	< 0.01

Note: Significance level: .05 (*), .01 (**).

Table 7. Descriptive statistics the CV with the different approaches

			Fuzzy weighting	Uniform weighting
			CV	CV
IPQ_{wsm}	MTT	Tracker	0.45	0.41
		Kinect and force feedback	0.34	0.39
	Scene type	low	0.45	0.42
		middle	0.45	0.39
		high	0.34	0.40
IPQ_x	MTT	Tracker	0.31	0.30
		Kinect and force feedback	0.29	0.31
	Scene type	low	0.34	0.38
		middle	0.35	0.35
		high	0.32	0.36

Table 8. CV for the different evaluation methods and weighting approaches

Mixed-effects model		
Factors	IPQ_{wsm} v.s. IPQ_x	Fuzzy weighting v.s. Uniform weighting
Sum Sq	0.0018	0.0008
Df	1	1
MS	0.0018	0.0008
F	0.93	0.35
P-value	0.00**	0.06
η_p^2	0.62	0.99
95% CI		
Mean Difference	-0.08	-0.01
Lower	-0.06	-0.04
Upper	-0.04	0.012

Note: Significance level: .05 (*), .01 (**), Sum sq: Sum square, Df: degree of freedom, MS: mean squared.

6. Discussion

Based on the stated results, IPQ_x can be deemed as a more precise presence quantification method than the weighted-sum approach in our case study. We exploratively discuss the effectiveness of IPQ_x on regional presence strength under multiple virtual feedback types and scene densities. We studied the IPQ evaluation performance based on the TOPSIS method considering two perspectives, namely, and verify the precision of the two weight evaluation methods by comparing the CVs.

Overall, we detected no evidence to substantiate that different VR MTTs could produce significantly multiple levels of presence, and neither IPQ_x nor IPQ_{wsm} exhibited such an effect. Thus, our findings are consistent with those reported in previous studies(Dey et al., 2020; Freitas, 2018), confirming that the IPQ_x is useful for evaluating the level of presence. As listed in the table 5, we can finding that the mechanism of the IPQ_x and IPQ_{wsm} methods were consistent.

Regardless of the evaluation method, no significant difference was observed between the middle and higher density of the scenes. Previous studies revealed that the participants' presence in a VE varies according to the attraction in the visual stimulus task(Slater and Wilbur, 1997; Voinescu et al., 2020). In our experiments, we created the intensity of visual stimuli by setting varying scene densities to ensure distinct perception among the users. From the IPQ_x results, when tasks were performed in lower density scenes, the visual stimuli decreased with the degree of task adaptation owing to the relatively low density of visual scenes, thereby resulting in a significantly lower sense of presence than middle-saturation scenes. Weech *et al*. (2019) argued that the two factors, presence and cybersickness, were negatively related, and that the main reason for this result was that participants typically rated the two feelings subjectively in terms of sensory integration when self-reporting them. This is consistent with our findings. We suggest that the participants' task flow was identical and that the low-density scenes likely induced cybersickness in the subjects, laterally reducing the presence of the low-density scenes among the participants. The visual stimulus richness of the scenes satisfied their exploratory and cognitive demands for the current swimming task, explaining the absence of any significant difference between moderate and high-saturation scenes, regardless of the evaluation method used. However, this does not indicate that a higher density of the scene reveals a stronger presence. This conclusion is consistent with that of a study by Lee (2004), which argues that it is reasonable to assume that realistic resolution is ineffective for presence in VEs because the low acuity of our peripheral vision makes us tolerant of low-resolution visual scenes.

Regarding the scene type, the experimental results obtained the weighted sum method revealed that the statistical significance could be detected only between the low and high densities, implying inadequate precision of this method for discriminating presence at smaller scales. In contrast, more significant effects were detected using the TOPSIS method, because differences were detected between the low and middle densities. Therefore, the TOPSIS approach can more precisely detect the marginal variations in presence, because of the CV of the IPQ_x was significantly smaller and indicated a lower dispersion of the data. Thus, an improved quality of discrimination among data could be achieved using the TOPSIS approach.

Regarding the weighting approaches, we did not detect a significant deviation in the measurement between the fuzzy hierarchy method and the unified weight method, because the application of both IPQ_{wsm} and IPQ_x yielded similar significant results. However, the CV of the fuzzy hierarchy method differed significantly from that of the uniform weight method. Thus, the measurement results of the fuzzy AHP were effective with reference to the unified weighting method. In particular, the fuzzy weighting method can be used as an alternative for the unified weighting method to determine the weighting coefficients of each criterion in the total presence evaluation. This further substantiates that the importance (weight) of each criterion is a necessary step in measuring and calculating the overall presence rating. Moreover, the smaller CV signified a higher precision of the presence measured using the TOPSIS approach, which is particularly important for those wishing to compare and manage the presence difference at smaller scales.Nevertheless, a considerable limitation pertains to the consideration of only two factors for determining the effectiveness of the proposed approach. Thus, future research will consider additional factors with more use cases to test the generality of this approach.

7. Conclusion

Given the rapid advancement of VR technology, the construction of presence is of paramount importance. In this study, we have employed the TOPSIS method to optimize the measurement of presence in the VR domain environment. Building upon the IPQ measurement method, we have tested a new presence calculation method that aims to improve the quantitative metrics for presence and reduce the coefficient of data variation. The model has been examined using two weighting methods: fuzzy weighting and uniform weighting. By incorporating the fuzzy analytic hierarchy process (FAHP) with a universal hierarchical analysis structure into the new calculation framework, it becomes possible to enable small-scale, rapid measurement of presence, even with limited sample sizes. We applied this method to test a VR swimming simulator that is currently under development. We investigated two factors in simulation tasks based on VR: the VR MTT type and scene density. The purpose was to validate the performance of the method. It is noteworthy that when using the TOPSIS evaluation to quantify the level of presence constructed by different VR immersion technologies, represented as IPQ_x, there was a decrease in the coefficient of variation (CV) of the evaluation scores compared to the weighted method. This suggests that the new calculation framework can effectively reduce the subjective uncertainty in measuring the presence of VR and enhance measurement accuracy. Moreover, based on the detailed experimental procedures described in this paper, this method can further benefit the measurement of presence in other domains.

Although the current results indicate the effectiveness of the new calculation framework, we acknowledge that our evaluation data were obtained from low-fidelity experimental scenarios and only involved 20 participants in the analysis of presence sample variation. Additionally, we did not extensively discuss the differences in presence evaluation between males and females. In future research, a more comprehensive validation of the model's performance will be conducted, considering larger sample sizes. Furthermore, it is important to note that this method only considered two influencing factors in VR design, which represents a significant limitation. In future studies, we will consider more factors and use cases to validate the generality of this method.

About the Authors

Shuo Li is a Ph.D. candidate at the School of Design, Shanghai Jiao Tong University (SJTU). She received her master's degree in industrial design engineering from Zhejiang University (ZJU). Her research interests include human-computer interaction, VR and olfactory interface.

Hongtao Zheng is a Ph.D. candidate at the School of Design, Shanghai Jiao Tong University (SJTU). He received his master's degree in design science from Harbin Institute of Technology (HIT). His research interests include persuasive technique and human–autonomy interaction.

Ting Han received the Ph.D. degree from the Department of Industrial Design, Zhejiang University (ZJU). He is a Professor of the School of Design, Shanghai Jiao Tong University (SJTU). He is also a Distinguished Professor with the Institute of Medical Robotics, SJTU. His research interests include human-computer interaction and healthcare.

Jun Hu received the Ph.D. degree from the Department of Industrial Design, Eindhoven University of Technology (TU/e). He is the Head of designed intelligence and an Associate Professor in design research on social computing with the Department of Industrial Design, TU/e. He has over 240 publications in conferences and journals.

Chenqi Zhang is a Ph.D. candidate at the School of Design, Shanghai Jiao Tong University (SJTU). She received her master's degree in design from SJTU. Her research interests include human-computer interaction and health promotion.

Cong Yu received her Bachelor's degree from the School of Electronic Information and Electrical Engineering at Shanghai Jiao Tong University (SJTU). She is currently pursuing a Master's degree from the School of Design, SJTU. Her expertise lies in Python and Unity, and she has developed a strong interest in human-computer interaction.

Funding

This work was supported by the National Key Research and Development Program under Grant [2022YFB3303303].

References

Mesut Akdere, Yeling Jiang, and Flavio Destri Lobo. Evaluation and assessment of virtual reality-based simulated training: exploring the human–technology frontier. *European Journal of Training and Development*, 46(5/6):434–449, 2021.

Oscar J. Ariza N., Markus Lange, Frank Steinicke, and Gerd Bruder. Vibrotactile assistance for user guidance towards selection targets in vr and the cognitive resources involved. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 95–98, 2017. .

Jeremy N Bailenson and Nick Yee. Virtual interpersonal touch: Haptic interaction and copresence in collaborative virtual environments. *Multimedia Tools and Applications*, 37:5–14, 2008.

Majid Behzadian, S Khanmohammadi Otaghsara, Morteza Yazdani, and Joshua Ignatius. A state-of the-art survey of topsis applications. *Expert Systems with applications*, 39(17): 13051–13069, 2012.

Barry Bixler and Scott Riewald. Analysis of a swimmer's hand and arm in steady flow conditions using computational fluid dynamics. *Journal of Biomechanics*, 35(5):713–717, 2002. .

Stephen Casner and Brian Gore. Measuring and evaluating workload: A primer, 2010.

Ching-Yi Chang, Han-Yu Sung, Jong-Long Guo, Bieng-Yi Chang, and Fan-Ray Kuo. Effects of spherical video-based virtual reality on nursing students' learning performance in childbirth education training. *Interactive Learning Environments*, 30(3):400–416, 2022.

Da-Yong Chang. Applications of the extent analysis method on fuzzy ahp. *European Journal of Operational Research*, 95(3):649–655, 1996. .

Miriam Clemente, Alejandro Rodríguez, Beatriz Rey, and Mariano Alcañiz. Assessment of the influence of navigation control and screen size on the sense of presence in virtual reality using eeg. *Expert Systems with Applications*, 41(4, Part 2):1584–1592, 2014. .

Alice Cohen-Hadria, Mark Cartwright, Brian McFee, and Juan Pablo Bello. Voice anonymization in urban sound recordings. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2019.

Hepu Deng. Multicriteria analysis with fuzzy pairwise comparison. *International journal of approximate reasoning*, 21(3):215–231, 1999.

Arindam Dey, Jane Phoon, Shuvodeep Saha, Chelsea Dobbins, and Mark Billinghurst. A neurophysiological approach for measuring presence in immersive virtual environments. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 474–485, 2020. .

Julia Diemer, Georg W. Alpers, Henrik M. Peperkorn, Youssef Shiban, and Andreas Mühlberger. The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, 6, 2015.

Matthew Farrow, Christof Lutteroth, Peter C. Rouse, and James L. J. Bilzon. Virtual-reality exergaming improves performance during high-intensity interval training. *European Journal of Sport Science*, 19(6):719–727, 2019. .

William M Felton and Russell E Jackson. Presence: A review. *International Journal of Human–Computer Interaction*, 38(1):1–18, 2022.

Yuan Feng, Emilia I Barakova, Suihuai Yu, Jun Hu, and GW Matthias Rauterberg. Effects of the level of interactivity of a social robot and the response of the augmented reality display in contextual interactions of people with dementia. *Sensors*, 20(13):3771, 2020.

José Diogo Camacho Freitas. Kave - kinect cave: Design, tools and comparative analysis with other vr technologies. 2018.

Giorgio M. Grasso, Giovanni d'Italia, and Sebastiano Battiato. A flexible virtual environment for autonomous driving agent-human interaction testing. In *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6, 2020. .

Alev Taskin Gumus. Evaluation of hazardous waste transportation firms by using a two step fuzzy-ahp and topsis methodology. *Expert Systems with Applications*, 36(2, Part 2):4067–4074, 2009. .

Kelly S Hale and Kay M Stanney. *Handbook of Virtual Environments: Design, Implementation, and Applications*. CRC Press, 2014.

David Harris, Mark Wilson, and Samuel Vine. Development and validation of a simulation workload measure: The simulation task load index (sim-tlx). *Virtual Reality*, 24(4):557–566, 2020. .

Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

Monique Hennink and Bonnie N Kaiser. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 292:114523, 2022.

Brent Edward Insko. *Passive Haptics Significantly Enhances Virtual Environments*. The University of North Carolina at Chapel Hill, 2001.

Katharina Jahn, Bastian Kordyaka, Caroline Ressing, Kristina Roeding, and Bjoern Niehaves. Designing self-presence in immersive virtual reality to improve cognitive performance—a research proposal. In Fred D. Davis, René Riedl, Jan vom Brocke, Pierre-Majorique Léger, Adriane Randolph, and Thomas Fischer, editors, *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation, pages 83–91, Cham, 2020. Springer International Publishing. ISBN 978-3-030-28144-1. .

Oliver Beren Kaul, Kevin Meier, and Michael Rohs. Increasing presence in virtual reality with a vibrotactile grid around the head. In *Human-Computer Interaction–INTERACT 2017: 16th IFIP TC 13 International Conference, Mumbai, India, September 25-29, 2017, Proceedings, Part IV 16*, pages 289–298. Springer, 2017.

GL Kesteven. The coefficient of variation. *Nature*, 158(4015):520–521, 1946.

George Lako. Women, fire, and dangerous things. *U-niversity of Chicago Press, Chicago*, 1987.

Kwan Min Lee. Why presence occurs: Evolutionary psychology, media equation, and presence. *Presence: Teleoperators & Virtual Environments*, 13(4):494–505, 2004.

Shuo Li, Hongtao Zheng, Wenyu Yuan, and Ting Han. Virtual reality for delivering swimming practice through water-free immersive training system. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 86212, page V002T02A086. American Society of Mechanical Engineers, 2022.

Matthew Lombard and Theresa Ditton. At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, 3(2):JCMC321, 1997.

Matthew Lombard and Matthew T Jones. Defining presence. *Immersed in media: Telepresence theory, measurement & technology*, pages 13–34, 2015.

Pat Lovie. Coefficient of variation. In *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd, 2005. ISBN 978-0-470-01319-9. .

Shawn Matthews, Alvaro Uribe-Quevedo, and Alexander Theodorou. Rendering optimizations for virtual reality using eye-tracking. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, pages 398–405, 2020. .

Tomasz Mazuryk and Michael Gervautz. History, applications, technology and future. *Virtual Reality*, 72(4):486–497, 1996.

B Mazzone, L Lighthall Haubert, S Mulroy, P Requejo, M Gotsis, V Lympouridis, B Lange, R Profitt, and C Winstein. Intensity of shoulder muscle activation during resistive exercises performed with and without virtual reality games. In *2013 International Conference on Virtual Rehabilitation (ICVR)*, pages 127–133, 2013. .

Eric B Nash, Gregory W Edwards, Jennifer A Thompson, and Woodrow Barfield. A review of presence and performance in virtual environments. *International Journal of human-computer Interaction*, 12(1):1–41, 2000.

Thomas D. Parsons, Louise Cosand, Christopher Courtney, Arvind Iyer, and Albert A. Rizzo. Neurocognitive workload assessment using the virtual reality cognitive performance assessment test. In Don Harris, editor, *Engineering Psychology and Cognitive Ergonomics*, Lecture Notes in Computer Science, pages 243–252, Berlin, Heidelberg, 2009. Springer. ISBN 978-3-642-02728-4. .

Sen Qiu, Hongkai Zhao, Nan Jiang, Zhelong Wang, Long Liu, Yi An, Hongyu Zhao, Xin Miao, Ruichen Liu, and Giancarlo Fortino. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, 80:241–265, 2022.

Shi Qiu, Jun Hu, Ting Han, Hirotaka Osawa, and Matthias Rauterberg. Social glasses: simulating interactive gaze for visually impaired people in face-to-face communication. *International Journal of Human–Computer Interaction*, 36(9):839–855, 2020.

Shi Qiu, Emiran Kaisar, Renbo Ding, Ting Han, Jun Hu, and Matthias Rauterberg. Social balance ball: Designing and evaluating an exergame that promotes social interaction between older and younger players. *International Journal of Human–Computer Interaction*, pages 1–24, 2023.

G Riva, F Davide, and WA IJsselsteijn. Measuring presence: Subjective, behavioral and physiological methods. *Being there: Concepts, effects and measurement of user presence in synthetic environments*, 5:110–118, 2003.

F David Rose, Elizabeth A Attree, Bennett M Brooks, David M Parslow, and Paul R Penn.

Training in virtual environments: Transfer to real world tasks and equivalence to real task training. *Ergonomics*, 43(4):494–511, 2000.

Thomas L Saaty. What is the analytic hierarchy process? In *Mathematical Models for Decision Support*, pages 109–121. Springer, 1988.

David W Schloerb. A quantitative measure of telepresence. *Presence: Teleoperators & Virtual Environments*, 4(1):64–80, 1995.

Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001.

Volkan Sevinc and Mehmet Ilker Berkman. Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments. *Applied ergonomics*, 82:102958, 2020.

Youssef Shiban, Julia Diemer, Simone Brandl, Rebecca Zack, Andreas Mühlberger, and Stefan Wüst. Trier social stress test in vivo and in virtual reality: Dissociation of response domains. *International Journal of Psychophysiology*, 110:47–55, 2016. .

Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364 (1535):3549–3557, 2009.

Mel Slater. Immersion and the illusion of presence in virtual reality. *British journal of psychology*, 109(3):431–433, 2018.

Mel Slater and Maria V Sanchez-Vives. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3:74, 2016.

Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.

Mel Slater, Martin Usoh, and Anthony Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.

Mel Slater, Andrea Brogni, and Anthony Steed. Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th annual international workshop on presence*, volume 157. Citeseer, 2003.

Vinicius Souza, Anderson Maciel, Luciana Nedel, and Regis Kopper. Measuring presence in virtual environments: A survey. *ACM Computing Surveys*, 54(8):163:1–163:37, 2021. .

Kay M. Stanney, Ronald R. Mourant, and Robert S. Kennedy. Human factors issues in virtual environments: A review of the literature. *Presence*, 7(4):327–351, 1998. .

Paweł Strojny, Natalia Lipp, and Agnieszka Strojny. Polish adaptation of the igroup presence questionnaire. 2022.

Alexandra Voinescu, Liviu Andrei Fodor, Danaë Stanton Fraser, and Daniel David. Exploring attention in vr: Effects of visual and auditory modalities. In Tareq Ahram and Christianne Falcão, editors, *Advances in Usability, User Experience, Wearable and Assistive Technology*, Advances in Intelligent Systems and Computing, pages 677–683, Cham, 2020. Springer International Publishing. ISBN 978-3-030-51828-8. .

Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology*, 10:158, 2019.

Florian Weidner, Anne Hoesch, Sandra Poeschl, and Wolfgang Broll. Comparing vr and non-vr driving simulations: An experimental user study. In *2017 IEEE Virtual Reality (VR)*, pages 281–282, 2017. .

Bob G. Witmer and Paul B. Kline. Judging perceived and traversed distance in virtual environments. *Presence*, 7(2):144–167, 1998. .

K Paul Yoon and Ching-Lai Hwang. *Multiple Attribute Decision Making: An Introduction.* Sage publications, 1995.

Chian-Son Yu. A gp-ahp method for solving group decision-making fuzzy ahp problems. *Computers & Operations Research*, 29(14):1969–2001, 2002a.

Chian-Son Yu. A gp-ahp method for solving group decision-making fuzzy ahp problems. *Computers & Operations Research*, 29(14):1969–2001, 2002b. .

Hui Zhang, Xueqiu He, and Hani Mitri. Fuzzy comprehensive evaluation of virtual reality mine safety training system. *Safety Science*, 120:341–351, 2019.

Hongtao Zheng and Shuo Li. Analysis of persuasive design mechanism based on unconscious calculation. *Scientific Programming*, 2022, 2022.

Ronggang Zhou and Alan H. S. Chan. Using a fuzzy comprehensive evaluation method to determine product usability: A proposed theoretical framework. *Work*, 56(1):9–19, 2017. .

8. Appendices

Appendix A. iGroup Presence Questionnaire

The current version of the IPQ includes three subscales and one additional general item, which emerged from principal component analyses and can be regarded as fairly independent factors. The three subscales are:

• Spatial Presence - the degree to which one feels physically present in the VE.

• Involvement - measuring the attention devoted to the VE and the level of involvement experienced.

• Experienced Realism - measuring the subjective experience of realism in the VE
Additionally, there is one general item that does not belong to any subscale.

See details in : http://www.igroup.org/pq/ipq/index.php

Appendix B. Fuzzy comparison questionnaire

B.1. Questionnaire design

Questionnaire (see Fig. 7 step: Perform pair-wise comparison):

• Name:

• The iGroup Presence Questionnaire (IPQ) is a widely used, subjective, multidimensional assessment tool that rates perceived effectiveness or other aspects of performance for tasks, systems, or teams. More specifically, it measures Spatial presence (SP), Experienced realism (REAL), Involvement (INV), and Sense of being there (G1).

• To determine the importance of different influencing factors, please fill in the following table based on what you believe was more important during the task you just performed?

• Language expressions: Both equally important (BI), Weakly more important (WI), Somewhat important (SI), Remarkably more important (RI), Very remarkably more important (VI), Extremely more important (EI), and Absolute important (AI).

B.2. Example of a filled questionnaire

See Table B1.

Table B1. One filled questionnaire from the experiment (G1: Sense of being there; SP: Spatial presence; INV: Involvement; REAL: Experienced realism); the cells with "-" are automatically filled during the data analysis as they have a reciprocal relationship with the item from the other side of the diagonal.

	G1	SP	INV	REAL
G1	BI	-	-	AI
SP	SI	BI	SI	-
INV	RI	EI	BI	-
REAL	-	RI	RI	BI