Research Article

Evaluating the Authenticity of Virtual Environments: Comparison of Three Devices

Aila Kronqvist,[1] Jussi Jokinen,[2] and Rebekah Rousi[1]

[1]*University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland*
[2]*Aalto University, 02150 Espoo, Finland*

Correspondence should be addressed to Aila Kronqvist; aila.j.kronqvist@student.jyu.fi

Immersive virtual environments (VEs) have the potential to provide novel cost effective ways for evaluating not only new environments and usability scenarios, but also potential user experiences. To achieve this, VEs must be adequately realistic. The level of perceived authenticity can be ascertained by measuring the levels of immersion people experience in their VE interactions. In this paper the degree of authenticity is measured via an *authenticity index* in relation to three different immersive virtual environment devices. These devices include (1) a headband, (2) 3D glasses, and (3) a head-mounted display (HMD). A quick scale for measuring immersion, feeling of control, and simulator sickness was developed and tested. The HMD proved to be the most immersive device, although the headband was demonstrated as being a more stable environment causing the least simulator sickness. The results have design implication as they provide insight into specific factors which make experience in a VE seem more authentic to users. The paper emphasizes that, in addition to the quality of the VE, focus needs to be placed on ergonomic factors such as the weight of the devices, as these may compromise the quality of results obtained when examining studying human-technology interaction in a VE.

1. Introduction

Advancements in virtual environment (VE) technology have enabled new ways to design, prototype, and evaluate other technologies [1–4]. The latest VE laboratories seem promising in their ability to create environments rich in detail and fidelity, while allowing the researcher to retain experimental control [5, 6]. This is encouraging for topics in human-technology interaction (HTI) research, such as user experience. Furthermore, the particular benefit of this in HTI research can be seen in cases which more or less demand the observation of interactions in realistic use contexts [7]. Timeliness and situatedness (context and environment) are two key facets of user experience [7]. For this reason, VEs offer researchers and designers the opportunity to create (simulate) the conditions of life-like human-technology interactions, for the purposes of observing and scientifically analyzing the affects and emotions involved in these interactions, via a variety of user experience methods applicable to their own research and design goals and intentions. However, in order to study life-like experiences in VEs in an

ecologically valid way [8, 9], the simulated environments and interactions need to be experienced as authentic [1, 6, 7, 10].

While in experience research, studies using VEs have acknowledged the need to consider validity [1, 6, 11, 12], common frameworks for assessing the validity of VEs in user experience studies are still lacking. This paper presents the development and piloting of a "quick and dirty" methodological assessment framework for measuring the authenticity of experience during a VE experiment. The framework requires common metrics for measuring the fidelity and naturalness of the VE experience. The framework introduced here, with its standardized measures for assessing the authenticity of the VE experiment, can be used as a benchmark to assess the strength of the experimental results and inferences made from the VE in relation to real-life human-technology interactions.

When considering the utilization of VEs for the purposes of studying usability, or user experience, the experience of presence is paramount. Through achieving a high sense of presence, there are possibilities to simulate interactions in

a number of environments and contexts, without having to physically leave the laboratory. The process of studying usability, for instance, requires that a user, or participant, is able to naturally interact and utilize a device or system in order to ascertain whether or not the design in question is learnable, efficient, memorable, error-free (or that the user has the ability to recover from errors quickly), and satisfying Nielsen [13]. These interactions necessitate that both actors—device/system and user—are present in a natural way. Moreover, research delving into more detail regarding user experience as a broader paradigm in HTI requires that interactions are examined in as natural settings as possible [14]. This research includes studies which focus on going beyond the instrumental and physical, social, and psychological analysis and previous as well as anticipated user experiences [7]. Thus, users, designers, and even researchers need to experience the sense of *being there*.

2. Presence, Affordance, and Control in Virtual Environments

A VE is a technologically generated environment, which allows the user to experience presence in a place other than where she or he is physically located [15]. In order to have an authentic experience in a VE, the user needs to be able to *feel presence*, that is, experience "being there," or being immersed in the environment [5, 16]. The feeling of presence is fundamental to our sense of what is real. Thus, presence is closely related to how people tangibly interact with objects in an environment [15]. Differences can subsequently be drawn between reality and images in contrast to reality (or virtual reality) in which the potential to interact with the environment and its contents is paramount; images only reflect and represent a version of reality. Consequently, another main difference between the two is the perceiver's inability to locate actual objects in space in a way which allows for this tangible or concrete interaction [17]. One cannot grab and smoke a pipe in a painting, and one cannot climb the mountains in a landscape photograph, even if techniques such as induced perspective and an implied third dimension allude to the character of *affordance*, an understanding of the ways in which objects (even pictorial) afford use and thus benefit the user through action and function [18].

Presence has been referred to as a "normal awareness phenomenon" [19] that demands focused attention and comprises sensory stimulation; factors in the environment which allude to and entice involvement to facilitate immersion; and internal tendencies within the person to become involved [19]. In his chapter "Immersion in Virtual Worlds" Calleja [20] describes the tendency and problematics of employing the terms "presence" and "immersion" to characterise the involved nature of people within virtual environments. Calleja [20] argues that the concept of "immersion" has been too widely used in relation to vastly different domains from artistic experience, to literature, cinema, and gaming, and due to this diversity in application has caused a lack of consensus regarding its specific definition. "Presence," originating in Minsky's [21] and "telepresence" utilized in

reference to telerobotics were already applied to the character of *authenticity* in VEs. Previous application of the term occurred in reference to Sutherland's [22] prototype of "The Ultimate Display" to describe the experience of *being there in the environment*.

Similarly to immersion, discussions on presence are highly interdisciplinary, which in turn causes debate and discrepancy in definition [23]. The presence debate has attracted scholars from the fields of psychology, philosophy, design, and communication to name some, who have adopted the concept in various forms to suit alternative theories such as simulation theory [24]; involvement [25]; the theory of play in psychology [26]; and affective disposition theory [27]. While debate continues, an overarching understanding of presence has been characterised not only as the sensation of *being there*, but also as that of *interacting there*, in the spaces of both the virtual and the physical worlds [28]. Thus, the feeling of presence is action-related and relates to a person's propensity to act, their ability to situate themselves in a social or physical space [29]. Presence is the experience of being able to transfer one's own knowledge about the world into interaction with the world [30]. Riva et al. [31] have gone even further to describe presence as "a selective and adaptive mechanism" allowing one to establish action-related boundaries through distinguishing between the "internal" and "external" facets of sensory flow [32].

Presence in the realm of human-computer interaction can be categorized into several different types: media presence, "the perceptual illusion of nonmediated presence" [33]; social presence, the feeling of existing and interacting with other intelligent beings [34]; copresence, the sensation of being in a mediated room with a mediated person [35]; physical, spatial presence and telepresence, the feeling of being there in the distal environment, and spatial awareness through the interaction of spatial representations [36]. Nonetheless, in cognitive science presence is seen as an embodied intuitive metacognitive process which bestows agency, enabling people to control their own actions through the process of comparison between perception and intention [32]. Presence consists of dynamic and continuous interaction between human sensory perception and cognitive and affective processes [33, 34]. The key to producing a sense of presence in technologically mediated environments is to develop the design to a standard to which a person loses the ability to detect the technological medium through which the environment (virtual) is communicated or represented [33].

In the case of the present study, technology on the one hand aids interactions and facilitates the sensation of presence within the virtual environments, as *invisible technology* (this is one perspective towards viewing what Streitz and Nixon [37] term as the "disappearing computer") or the "illusion of nonmediation" [33], while on the other hand, here the VEs themselves provide environments in which people (study participants) can come into contact with other technologies, the artefacts, and services under scrutiny—the *visible technology*—in order to experience these in simulated use situations. Thus, from the perspective of *affordance*, technology helps with this interaction: it is in our way of considering technological artefacts as a means to an end that makes

these artefacts appear to us as technological in character. Therefore, technological artefacts in VEs need to be able to satisfy the expected affordance, which refers to the actions for which the object allows [11, 18, 38, 39]. Coined by Gibson [18] the term affordance refers to the way in which people do not see things (or in this case environments) in and of themselves. Rather, people see and understand objects and systems in terms of what they afford them to do, and how this will potentially result in reward or detriment. In terms of design, experienced affordance may include easy to life (how well the design suits a person's life situation), reachability and tactile qualities, status symbol, feeding excitement, and reliability.

In other words, affordance refers to how people understand environments and associated phenomena as not objects, or isolated pieces of information, but rather, in terms of what they do, and their expected repercussions. Thus, perceived phenomena are understood in terms of how they will influence and support people's actions. In fact, regarding Gibson's "ecological approach to visual perception," it may be observed that people possess two ways of *seeing* or perceiving VEs [18]. On the one hand, people perceive the affordances of what is inside the VE (or picture), and on the other hand they can perceive the VE itself in terms of what *it* affords as a design. The idea behind increasing presence is to reduce the gap between these two modes of seeing, by heightening a person's sense of affordance in virtual properties. Thus, VEs which seek to induce *authentic* experiences of presence need to suggest affordances to technologies represented in the environment. For example, in the physical world the design of a chair alludes to the affordance of supporting someone who sits on it. Likewise, in the VE, the observer should also experience the chair as having the same affordance as in the physical world, even though in the VE the chair does not necessarily possess this functional property.

For this reason, control or the sense of control is important when experiencing presence if a person feels as though they can control and influence their environment or utilize/manipulate what is afforded to them through the representations of, for example, technological devices and interfaces ready for testing; they are more likely to experience what is inside the VE as *real* [30]. Some of the mechanisms for inducing a sense of control and affordance include feedback, space, distance, form, size, and movement [5, 40]. Thus, the definition of presence as action-related immersion has two clear implications for the design of an authentically experienced VE. The user of a VE should, in order to feel presence, be able to locate objects of the VE relative to her or himself and to experience a degree of control over the objects. Both implications involve technical requirements, which the VE needs to pass in order to be useful in studying how people experience technologies. This entails sufficient resolution and depth of the simulation [41, 42], the ability of the user to move in the environment and study it from multiple perspectives [11], and the means of letting the actor interact fluently with the environment [6].

Achieving sufficient simulation resolution has been shown to be one of the main requirements of an authentically experienced VE. Yet only recently the level of technology has been adequate enough for the reliable projection of

high resolution VEs [42–44]. High resolution in itself is not enough to create the experience of a rich and realistic environment. In order to better facilitate the user's ability to pinpoint the locations of the objects in the VE, the depth of the simulation can be increased by creating three-dimensional (3D) stereoscopic projections [41, 45]. Increased depth can be added by utilizing different modalities [43], such as surround sound, which has been shown to increase the level of immersion in the environment [46, 47].

Sufficient resolution and depth of the VE are required for authentic environments, but they do not necessarily provide sufficient action-orientation, which is essential for the feeling of presence. In order to fully experience presence in an environment, the user needs to be able to move in it. Otherwise, there would be no difference between looking at a picture and truly being in an environment [17, 48]. To facilitate naturalistic movement in a VE, the simulation needs to react to the movements of the user, and the visual presentation needs to be shifted accordingly in order to simulate movement in the VE [11]. This coupling of the user and the simulation increases the feeling of presence by connecting the act of moving to the visual presentation of the environment. Further, the actor should be able to interact with the simulated environment in order to truly experience it [6, 11]. One way to do this is to have the user wear special gloves or hold a pointer, which can be tracked and visually represented to the user in the simulation [49, 50]. By visualizing these handheld pointers as objects of the VE, the user is given the means to interact with the simulated environment and gain visual feedback from this interaction [6]. Other mechanisms such as haptic feedback can boost the experience of interacting with the VE; however, this experiment focused on the effects of visual constructions of VE authenticity. This meant that the gloves and pointers were apt for this type of investigation as they enabled interaction, without influencing the evaluation of the primarily visual experience.

The ecological validity of HTI studies conducted in VEs depends, therefore, largely on the amount of the feeling of presence, and lack of simulator sickness, a side effect of the visual perception of movement, of which the vestibular understanding of the body is being stationary. Although the need to assess the validity of the experience in a VE has been noted before [11, 49], a common framework for evaluating the authenticity of VE experience is still lacking. This framework needs to incorporate the factors described above in clearly operationalized metrics that are quick and easy to use, producing data, which are useful for comparing different VE implementations. In the current experiment, subjective measurement scales for the authenticity of the experience are developed and tested in comparison of different VE devices. This is an emerging field, and although theories such as presence [19] and immersion [20] have proven influential, there is still a lack of efficient and practical tools for measuring experienced authenticity in VEs, while reporting the level of simulator sickness induced.

2.1. Simulator Sickness. VEs have been promising in their ability to simulate situations and environments in which people experience the sensations of *being there* and *interacting*

there. One major side effect however has been the onset of simulator sickness, or cybersickness. Awareness of simulator sickness has existed since the mid-1950s from the time of the first helicopter flight simulators [51]. A widely used explanation for simulator sickness is Sensory Conflict Theory [52–54]. Sensory Conflict Theory posits that simulator, or cybersickness, results from conflict in the two primary sensory systems involved in perceiving the VE: the vestibular system, providing information regarding movement as well as the head's orientation in space [55], and the perceived visual sense of self-motion, vection which is the deceptive feeling of self-motion although a person is stationary (e.g., such as when one is sitting on a train and the adjacent train begins to move) [56]. Thus, simulator sickness is seen to be caused by discrepancy in what is perceived and consciously experienced (i.e., self-movement), and the actual physical state of the body (stationary position). Thus, the visual system informs the individual about specific details and movement in the visual environment, but the vestibular system, used to regulate the head's orientation, does not correlate with the visual information, as the person is *not actually moving* [53].

The physical symptoms resulting from this conflict of sensory information include headache, vertigo, eye strain, cold sweats, fullness of stomach, nausea and vomiting, disorientation, dryness of mouth and ataxia, numbness, tingling in the hands and feet, and difficulty in coordinating muscle movements [57]. Correlations have been observed between an individual's age and the likelihood; they are to experience simulator sickness [58]. Thus, the older an individual is, the more likely they are to experience simulator sickness [51]. The effects of simulator sickness rarely last longer than 12 hours [51]. Various techniques have been proposed and tested to alleviate the likelihood of simulator sickness, such as the introduction of independent visual backgrounds to decrease balance disturbance [59]; ensuring that VE interaction sessions do not exceed two hours [51]; and adaptation [58] or habituation [60], a brief or trial acclimation session in the VE several days before actual testing or experimentation. Thus, the ideal result of employing VEs in the systematic measurement of user experience and other human-technology interaction observations is the optimal sensation of presence, combined with the least possible (or *no*) reporting of simulator sickness.

3. Authenticity Index and Framework

Through their operationalization of the *presence question-naire* (PQ) and *immersive tendencies questionnaire* (ITQ), Witmer and Singer's study on presence [19], shows that there is a slight correlation between task performance and presence in VEs. This aspect can be seen as one of the explanatory factors for research concentrating on VE authenticity in a variety of contexts ranging from learning [30, 61–63]; work and collaboration tools [64]; tools for experimental psychology and neuroscience [65, 66]; the arts [67]; and gaming and entertainment [68]. Thus, a unifying factor is the act of interacting through doing and achieving action goals in these environments. While PQ [19] is widely used in the study of presence in immersive technologies, it has the

drawback of being extremely long and arduous to complete. This not only induces fatigue in participants responding to the questionnaire but also takes away from the immediacy of the experience being reported in the evaluation form. Thus, the aim of developing the *authenticity index* was to create a tool for measuring the degree to which people (participants) experience reality in VEs, which was both succinct and accurate. The authenticity index consists of a questionnaire designed to measure immersion, control, and the side effect of simulator sickness. The questionnaire needed to be short enough for subjects to answer easily and efficiently, but detailed enough to provide rich data to measure authenticity. Lightness, in terms of easy comprehension and efficiency to answer, is desirable since the feeling of presence needs to be assessed relatively quickly. The required components contributing to the characteristics of immersion and control can be interpreted from themes presented by Witmer and Singer [19], which are seen in Table 1.

The authenticity index was generated from two different categories: (1) describing how well the subject was immersed in the environment and (2) the level of control she or he experienced when working in the environment. The factors measured in order to create the authenticity index are seen in Table 2. The number of experienced technical problems was added to indicate the degree to which these technical disturbances decreased the feeling of control.

4. Method and Materials

4.1. The Virtual Environment. The experiment was conducted in a VE laboratory, which was a large room with one wall-sized video screen (size 358 × 221 cm, resolution 1280 × 720) and eight cameras. The VE used in the experiment was modelled using the Unity game engine (Unity3D, version 3). Three different devices were used to project the simulated environment to the participants (Figure 1). The first device, a headband, had markers attached to it to track the movements of the user, allowing the coupling of the user and the visual projection. OptiTrack VR trackers with a MiddleVR software platform were used to accomplish the coupling. The VE was projected onto the video screen as a normal two-dimensional projection. For a 3D effect, stereoscopic *3D glasses* (XPAND 3D; size: large) were used as the second device. The glasses included markers to track the participant. It was hypothesized that the glasses would receive a higher score for the feeling of presence in comparison to the headband due to the added third-dimensional depth. However, it was also noted that stereoscopy might induce additional simulator sickness [69, 70]. Hence, it was hypothesized that the score for simulator sickness would be higher for the 3D glasses as compared to the headband.

The third device used in the VE projection was a *head-mounted display* (HMD). The participants wore a Sony HMD (Sony HMZ-T2; HD oled panel, FOV 45 degree), completely covering their vision. The HMD also used stereoscopy and thus allowed for a three-dimensional representation of the environment [71, 72]. The wall screen was not seen by the participants while wearing the HMD but was left on for the experimenters' observations. Tracking and coupling of

TABLE 1: Factors hypothesized to contribute a sense of presence [19].

Control factors	Sensory factors	Distraction factors	Realism factors
Degree of control	Sensory modality	Isolation	Scene realism
Immediacy of control	Environmental richness	Selective attention	Information consistent with objective world
Anticipation of events	Multimodal presentation	Interface awareness	Meaningfulness of experience
Mode of control	Consistency of multimodal information		Separation anxiety/disorientation
Physical environment modifiability	Degree of movement perception Active search		

TABLE 2: Factors used to create authenticity index.

Category	Factor
Immersion	Feeling of presence
	Anticipated affordances compared to fulfilled affordances
Control	Feeling of control
	Discovery ratio
	Amount of technical problems experienced

the user movements were implemented as with the other devices. Of the three devices, the HMD was hypothesized to have the best score for the feeling of presence, because it completely occluded the participant's vision and obscured any visibility of the laboratory itself [71]. However, because of the stereoscopy and immersiveness, the HMD was also hypothesized to result in the highest level of reported simulator sickness [71].

In all of the task conditions, the participants used a handheld controller (FragFX Shark PS3 Classic with self-added trackers) to interact with the VE [49, 50]. The controller had markers attached, which enabled a projection of the controller in the simulation fashioned as a hand-sized cylinder with a white line protruding from it. The line helped the participant to point at objects, and its color changed to green when the object which was being pointed at possessed interactive features (e.g., an interactive button, or the ability to pick up an object). The physical controller had a button for triggering these actions.

The VE environment used in the experiment consisted of a garage with a car inside. The physical skeletal structure of the interior of the car was constructed in the middle of the room and consisted of a chair, a steering wheel, and a gearstick (see Figure 1). Figure 2 shows the projection of the car dashboard onto the laboratory wall. Most of the tasks required the participant to sit in the car. However, there were additional tasks in which the participant was asked to walk around the vehicle's exterior.

4.2. Participants and Tasks. $N = 15$ participants (8 women and 7 men) were recruited for the experiment. Their mean age was 31.9 years, SD was 12.5, and age range was 20–63. This is a highly varied age sample and as mentioned above in relation to simulator sickness, increased age can have an impact on the degree of symptoms experienced. However, during this pilot study, the objective was to test the effectiveness of the *authenticity index* as a reliable instrument for measuring the level of authenticity experienced in the simulated environments, in accordance with the tested devices. Thus, in a study focusing specifically on the VEs and devices, a higher degree of control would be implemented in relation to the selection criteria of participants (such as age). All of the participants had a driver's license and most drove daily or weekly (two reported that they drove less than monthly). The experimental design was a counter-balanced within-subjects design. Each participant conducted tasks in three blocks, using all three devices in a counter-balanced order. In each task block, the participants conducted nine tasks, which were similar but not identical (except the first task, see below) between the blocks.

The experiment started with the participant sitting in the chair (inside the virtual car model), and the experimenter asking the participant to name all visible instruments on the virtual dashboard of the car. This task, called the *discovery task*, was used to test the ability of participants to recognize objects in the simulated environment, as well as give them the chance to familiarize themselves with the VE. The discovery task was conducted at the beginning of all three task blocks, with all three devices. For the purposes of the analysis, only the first task block of each participant will be referred to as in the succeeding blocks the dashboard was already familiar.

After the participant had completed the discovery task, the experimenter verbally presented the rest of the tasks one by one. The tasks were as follows: checking the odometer reading; checking the fuel tank reading; reporting the gear-stick configuration; putting the gear into reverse; testing the loudspeakers of the car by using the radio; adjusting the lights, checking, or changing the current radio frequency; naming the buttons on the steering wheel; selecting the highest gear; locating the seat warmer controls; and opening the window. In addition to the tasks which the participants conducted while sitting inside the car, there were tasks requiring the participants to step outside the car model, and into the garage. These tasks were checking for dents or rust in the exterior of the car; visually inspecting the tire pressures or wheel rims; and visually inspecting the windows.

From the above listed tasks, in addition to the discovery task, this paper reports the analysis and results of the *loudspeaker task* and the task of inspecting the exterior of the car (*inspection task*). The loudspeaker task was chosen because it highlights interaction with the environment.

Figure 1: The devices used in the experiment: upper left: headband; lower left: 3D glasses; upper centre: head-mounted display (HMD); lower centre: the controller; right: car skeleton.

Figure 2: The VRE wall screen behind a participant.

The inspection task was chosen as it required participants to step out of the car, which hence entailed movement inside the VE. As with the discovery task, only the results of the first task block of each participant are discussed here.

4.3. Data and Analysis. Two main sources of data were used in this study. Firstly, participants were asked to complete questionnaires, which allowed for the statistical examination of the hypotheses. Secondly, the participants were asked to verbally think aloud during the tasks. Additionally, they were interviewed at the end of the experiment. This resulted in textual data about the participants' experiences, complementing the numerical, more standardized data. This paper focuses on reporting the results of the questionnaires, utilizing some of the findings resulting from the thinking aloud data and interviews to explain the results. However, due to the length and detailed nature of the results, the thinking aloud data scrutinized via protocol analysis are not included in this current paper.

4.4. Questionnaires. Presence questionnaire (PQ) [19] is a widely used method for measuring the feeling of presence in VEs. For the purposes of the experiment reported here, PQ was considered too long and detailed to be truly effective in capturing the dynamic and ephemeral impressions of authenticity in VEs. The measures presented in this paper are intended to be used to give validity to user experience studies conducted in VEs. It is probable that in these kinds of studies other questionnaires, closer to the focus of the study, such as specifically targeted user experience measurements, are also utilized. This makes a "quick and dirty" alternative to the tradition PQ even more appealing, as it frees space (and time) for researchers and designers to focus on the issues specifically at hand. Hence, the feeling of presence needs to be assessed relatively quickly, which was the goal of the scales presented here. PQ was used as a basis for the creation of the faster, easier to fill questionnaire.

Based on the discussion in the introduction, the feeling of presence was operationalized in two subjective scales: feeling of control and immersion. To measure the participant's sense of being in the VE, and ability to locate objects, a scale of *immersion* was created. This contained the following four items, adapted from the PQ [19]:

 (i) I was immersed in the environment.

 (ii) The visual elements of the environment felt natural.

(iii) The experience in the virtual environment was congruent with a real world experience.

(iv) I could inspect the objects of the environment.

To measure the participant's ability to interact with the simulated environment, a scale of *control* was created. This contained the five following items, adapted from the PQ:

(i) I could control what happened in the environment.

(ii) I felt the environment reacts to my actions.

(iii) I felt my actions in the environment were natural.

(iv) I could anticipate the results of my actions.

(v) My actions to control the environment were natural.

The questionnaire was completed by the participants after each task block in order to enable within-subjects comparison of the devices. The scale of the items for immersion and control was from one ("very much disagree") to five ("very much agree"). The scale was presented to the participants as numbers (1–5), as well as text. Cronbach's alphas, indicating the reliability of the scales, were calculated separately for immersion and control between the three devices. The immersion alphas were for the headband $\alpha = 0.77$, the 3D glasses $\alpha = 0.78$, and the HMD $\alpha = 0.90$. The control alphas were for the headband $\alpha = 0.81$, the 3D glasses $\alpha = 0.70$, and the HMD $\alpha = 0.82$. The reliability of the scales was considered sufficient ($\alpha > 0.70$), and the items were calculated into summated scales by averaging the sum of the items. This procedure retained the original scale of the variables (from one to five), which made comparing the three conditions easier.

In order to test the hypothesis that the different devices had different scores for the feeling of presence, nonparametric repeated measures Friedman tests were conducted for the two summated scales. Dunn-Bonferroni tests were conducted for pairwise comparisons between the conditions. Nonparametric testing instead of the analysis of variance was conducted, because the number of the participants was small.

For measuring simulator sickness, a modified simulator sickness questionnaire was constructed by choosing eleven items from the standardized simulator sickness questionnaire [70]. The items were general discomfort, fatigue, headache, eyestrain, difficulty focusing, increased salivation, nausea, difficulty concentrating, blurred vision, dizziness (eyes open), and dizziness (eyes closed). The scale of the items was from one ("not at all") to five ("very much"). The simulator sickness questionnaire was presented to the participants after each task block, and at the very beginning of the experiment for a baseline measure.

In order to analyze the experiences of simulator sickness, the nine questionnaire items were summed together and averaged to create a summated scale. The sum variable, simulator sickness, retained the original scale of the modified questionnaire (1 = "no sickness at all," 5 = "extreme sickness"). The comparison of simulator sickness between the conditions was analyzed using the Friedman test. Further, exploratory item-level analysis was used to reveal the most prominent items of simulator sickness by individually comparing the means of the conditioned responses to the baseline responses. In order to compare the conditions using one measurement, a principal components analysis (PCA) including the three sum variables (immersion, control, and simulator Sickness) was conducted. The component scores of the three sum variables were used to calculate a standardized index value for each of the three devices. The constructed authenticity index was then compared between the conditions using Friedman test. Based on the hypotheses above we predicted that the HMD would receive the best index score.

4.5. Think Aloud Protocols and Interviews. In addition to the questionnaires, the think aloud method was used to collect data about the behaviour of the participants during the tasks. The participants were asked to verbalize their thinking at all times while conducting the tasks. At the beginning of the experiment, the participants practiced thinking aloud by performing simple calculations in their head while verbalizing the contents of their thought. The participants were also constantly reminded to think aloud by asking them to verbally repeat each task instruction.

The verbal reports of the participants were analyzed using protocol analysis [73]. The methodological assumption is that the participants verbalized the contents of their working memory, and these data are useful in understanding the participants' mental representations of the environment in which they acted [74]. The sequences of actions were gathered from verbalizations in order to see how the tasks were solved and if any problems were encountered during the tasks. The sequences were used to identify problematic tasks which were then analyzed in more detailed level (the radio task and the inspection task). At the end of the experiment, the participants were also shortly interviewed about their experience and asked to give feedback from all three VE techniques. The interview answers were compared to the results of the statistical analyses of the questionnaire answers.

4.5.1. The Discovery Task. The first task of the participants was to freely name all visible controls in the dashboard of the car. Out of the 28 possible objects, the lowest number named by a participant was five and the highest 22. On average, the participants named 11.1 instruments (mdn 10.5). The average number of instruments detected with the headband, when it was the first condition, was 7.8 (mdn 7.5), the 3D glasses 13.0 (mdn 13.0), and the HMD 13.0 (mdn 15). A post hoc Wilcoxon test revealed the difference between the detected instruments in the headband condition and the 3D glasses condition to be statistically significant, $p = 0.011$. The difference between the headband and the HMD was not statistically significant, $p = 0.056$. These results partly support the hypothesis that adding the third dimension increases the feeling of presence of the participants, as they are able to locate more objects in the environment.

As an interesting notion concerning the action-relatedness of the sense of presence, the verbalizations revealed that, in addition to the visual search, the participants also tried to interact with the discovered instruments (e.g., activating the windscreen wipers after discovering the wiper lever). The participants were eager to try the instruments while discovering them, but unfortunately most of the instruments had no functionality in the VE model. Although the discovery task was only to name the visible instruments of the dashboard, the lack of interaction with the instruments was judged as a hindrance for completing the task. This lack of interaction most likely caused decreased experienced authenticity in all conditions.
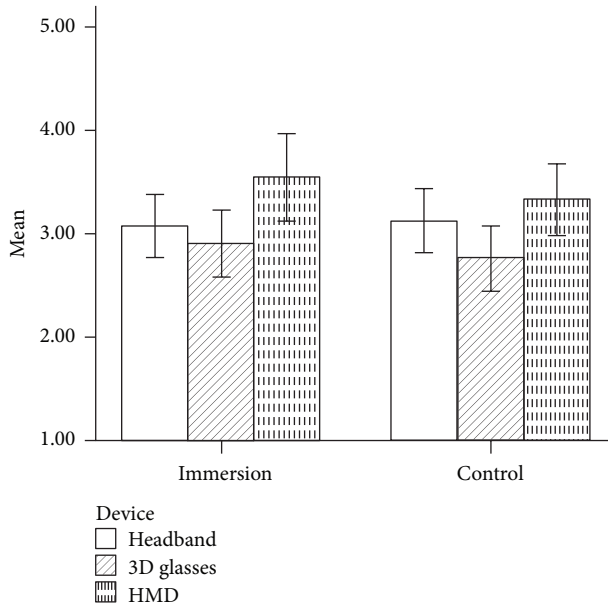
FIGURE 3: The means of immersion and control summated scales between the three devices. The lines represent 95% confidence interval.

FIGURE 4: The mean of the standardized VE authenticity index between the devices. The lines represent 95% confidence interval.

5. Results

5.1. Immersion, Control, and Simulator Sickness. The means of the summated scales between the three devices are shown in Figure 3. The distributions of immersion between the devices were different, $\chi^2(2) = 8.8$, $p = 0.012$. Pairwise comparisons revealed that the differences between the HMD and the 3D glasses and the HMD and the headband were statistically significant, $p = 0.008$ and $p = 0.022$. The difference between the 3D glasses and the headband was not statistically nonsignificant, $p = 0.715$.

The difference in the distribution of control between the conditions was also statistically significant, $\chi^2(2) = 12.3$, $p = 0.002$. Pairwise comparisons revealed that the difference between the HMD and the 3D glasses was statistically significant, $p = 0.001$. The other pairwise comparisons were statistically nonsignificant (3D glasses and headband $p = 0.121$, and HDM and headband, $p = 0.068$). The results partly support the hypothesis that the HMD has the best score for feeling of presence. However, contrary to the hypotheses, the 3D glasses did not perform well in immersion and control scales, although the differences between the conditions were not very large.

The simulator sickness scale scores were low for all devices. The mean of the baseline simulator sickness was 1.14, for the headband 1.12, the 3D glasses 1.25, and the HMD 1.14. The distribution was nevertheless different between the devices, $\chi^2(2) = 12.0$, $p = 0.002$. In pairwise comparisons, the difference between the headband and the 3D glasses was statistically significant, $p = 0.003$, but the others were not. The result suggests that none of the devices caused notable amounts of simulator sickness, but the 3D glasses seemed to introduce slight sickness. At the item-level analysis,
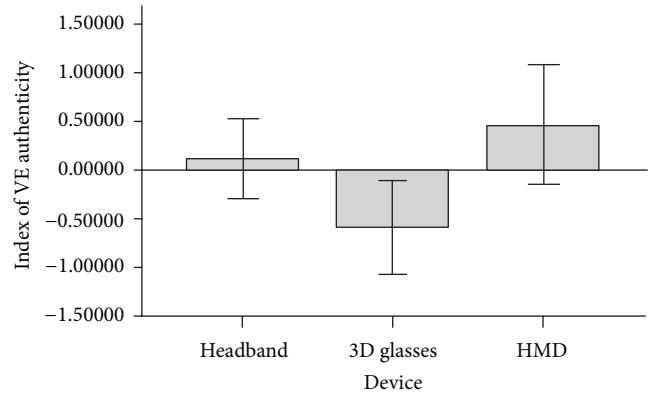
the main items contributing to the mildly higher simulator sickness score of the 3D glasses were eyestrain, concentration difficulties, general discomfort, and blurred vision.

The three sum variables were then combined into an index of authenticity by using single-component PCA. The component explained 61.9% of the total variance. The component loadings of the sum variables were control 0.90, immersion 0.86, and simulator sickness −0.55, indicating strong positive correlation between the two scales of feeling of presence, and a negative correlation between feeling of presence and simulator sickness. Due to standardization, the mean of authenticity was 1.00, and hence less than average values were negative. This is visible in Figure 4, which shows the comparison between the devices. The distribution of authenticity which was different between the devices was statistically significant, $\chi^2(2) = 9.7$, $p = 0.008$. Pairwise comparisons revealed that only the difference between the 3D glasses and the HMD was statistically significant, $p = 0.002$. While effect-sizes for nonparametric tests are not commonly reported, it is possible to visually inspect Cohen's d from Figure 4. Because the VE authenticity index is scaled, a difference of 1.0 in the means of two conditions equals the effect of $d = 1.0$. Thus, the difference between 3D glasses and HMD is large ($d > 0.8$) and clear to detect.

5.2. Think Aloud Protocols and Interviews

5.2.1. The Discovery Task. The first task of the participants was to freely name all visible controls in the dashboard of the car. Out of the 28 possible objects, the lowest number named by a participant was five and the highest 22. On average, the participants named 11.1 instruments (mdn 10.5). The average number of instruments detected with headband, when it was the first condition, was 7.8 (mdn 7.5), the 3D glasses 13.0 (mdn 13.0), and the HMD 13.0 (mdn 15). A post hoc Wilcoxon test revealed the difference between the detected instruments in the headband condition and the 3D glasses condition to be statistically significant, $p = 0.011$. The difference between the headband and the HMD was not statistically significant, $p = 0.056$. These results partly support the hypothesis that adding the third dimension increases feeling of presence of

FIGURE 5: The radio used in the radio task. Red arrow points to the correct button for turning the radio on. Blue arrow shows where the participants tried first to start the radio.

the participants, as they are able to locate more objects in the environment.

As an interesting notion concerning the action-relatedness of sense of presence, the verbalizations revealed that, in addition to the visual search, the participants also tried to interact with the discovered instruments (e.g., activating the windscreen wipers after discovering the wiper lever). The participants were eager to try the instruments while discovering them, but unfortunately most of the instruments had no functionality in the VE model. Although the discovery task was only to name the visible instruments of the dashboard, the lack of interaction with the instruments was judged as a hindrance for completing the task. This lack of interaction most likely caused decreased authenticity in all conditions.

5.2.2. The Radio Task. The radio task served as a demonstration for interaction with the VE. As such, it revealed problems associated with the feeling of presence in simulated environments. Resolution details, problems with tracking, and the lack of feedback from physical interaction were all observed to hinder participants' immersion in the environment. In addition, the verbalizations revealed that the controller was not perceived as a natural replacement for hands, but as an artificial and relatively unstable interface between the subject and the VE. This caused utterances of distrust towards the controller.

The first subtask of the participants in the radio task was to start the radio. Finding the correct button to turn on the radio was not easy: the participants first tried to press the large round button in the middle, highlighted with a blue arrow in Figure 5. However, the correct button was the smaller button to the left, highlighted with a red arrow. The confusion was partly caused by the insufficient resolution of the simulation: the correct button was labelled as "Radio," but the participants were not able to read the label clearly. Therefore, they had to hunch and move their head closer to the radio in order to inspect it and read the small text. Moving, especially with the 3D glasses, was considered unnatural, and such an easy task as leaning forward to inspect a radio proved to be difficult.

The second problem in the radio task was in the lack of feedback from incorrect button pressing. Bringing the controller close to the correct button changed the coloring of the controller pointer from white to green, but with

noncorrect buttons, the color stayed white. This leads to confusion, as the participants were uncertain if the button they were trying to press was incorrect, or if they just failed to position the controller properly. The participants became frustrated, as they felt that the controller did not adequately let them to interact with the radio.

5.2.3. The Inspection Task. At the beginning of the inspection task, participants were told that they now needed to see if the car had dents, if it was rusty, or if the tires or windows were broken. Many of the participants did not realize that they could step outside of the car model. Some participants tried to put their head through the door of the car while seated. After trying to solve the problem from inside the car model, some asked if they could step outside the car, but others had to be prompted to do so. The novelty of the idea of stepping outside the model was also evident in the immediate verbalizations of the participants when they stepped out of the car.

Technical problems concerning tearing and freezing of the simulation were observed when the participants walked next to the car. Because the tracking area was limited, any movement outside the specified area, or crouching to see under the car model, was prone to cause tracking problems and subsequent simulation errors. After the initial astonishment at being able to roam the VE, the technological constraints of the laboratory served as a disillusioning and disappointing element.

6. Discussion

The authenticity of a virtual environment (VE) was evaluated between three devices by asking participants (N = 15) to conduct tasks in a virtual car model. The devices were a headband, 3D glasses, and a head-mounted display (HMD). The participants were asked to detect objects in the environment, interact with the environment, and move around in the environment. Their actions during the tasks were investigated, and responses to posttask questionnaires were analyzed. Comparisons between the devices resulted in a proposal for a VE authenticity index.

VE authenticity was measured by subjective questionnaire scales for the feeling of presence (operationalized into the feelings of immersion and control). The first hypothesis was that introducing 3D-stereoscopy by having the participants wear 3D glasses or an HMD, instead of a tracker headband, would increase the feeling of presence [73]. Using 3D glasses or the HMD resulted in more detected instruments in the detection task when compared to using only the headband, which partly supported the hypothesis. One possible explanation for this is that the participants experienced the depth of the 3D-environment as more interesting and thus focused more on its exploration [41]. However, although the 3D glasses resulted in more detailed detection, the device received the lowest score for the immersion and control scales. The HMD on the other hand received the highest score on both scales, although when compared to the headband, the difference was not statistically significant.

Although lengthy, widely used questionnaires for assessing the feeling of presence exist [19], our goal was to construct a short but still reliable scale for constant validation of VE experiments. The scales for measuring the feeling of presence worked well with only a few items (four or five), as supported by proven reliability. The scales also highlighted understandable differences between the devices, giving support for the construct validity of the scales. The benefit of a small "authentication questionnaire" is that it leaves room for longer questionnaires concerning the experiences of the system being evaluated in the VE, while still providing input for the evaluation of the authenticity of the measured experience.

The metrics for VE authenticity were combined in an authenticity index, which allowed an easy comparison between the devices. Principal component analysis provided support for the claim that the feeling of presence is dependent on the ability to both locate objects (immersion) and interact with them (control). Comparison of the authenticity index between the devices suggests that experiences with the HMD were most authentic. However, the authenticity score for the HMD was not as stable across participants as with the headband, as was evident in the larger confidence intervals for the HMD. This encourages future studies exploring the individual differences in how authentic a VE is experienced.

In postexperiment interviews, participants reported that while the HMD provided a novel, immersive experience, the headband was less cumbersome and still capable of providing sufficient immersive experience. 3D glasses were reported to have problems with resolution and clarity. The HMD and the headband were both favoured: the HMD gave the best support for the feeling of presence, but longer tasks were more pleasant to conduct while wearing the less invasive headband. We suggest that even at the risk of reduced immersion, a noninvasive combination of headband and wall screen may work better for VE studies in HTI, at least until HMD technology has better matured in terms of both virtual fidelity and physical ergonomics.

The authenticity index variable proved to be capable of revealing differences between the devices. As such, it provides a promising start to creating a valid and reliable measure for VE authenticity. The next steps are to combine it with the observations made from the talk aloud verbalizations. For example, regarding the handheld controller as a tool rather than an actual bodily extension can be used as a questionnaire item. It should also be possible to combine the result of the discovery task with the subjective scales. From the definition of presence it should follow that the subjective feeling of presence is positively correlated to the number of detected items. In the experiment reported here, the number of discovered items was not included in the authenticity index, as the number of participants was relatively small: for each device, there were only five participants using the device in the first task block; thus the discovery task resulted in the clearest differences between the device. In the subsequent task blocks, the car interior was familiar, and the discovery task was relatively trivial. However, below it is suggested that a discovery task should be operationalized as a standard measure for the VE authenticity index.

The results from the discovery task serve as a reminder of the action-relatedness of the feeling of presence [17, 48]. Participants were eager to interact with the items they discovered. Finding out that a windscreen wiper lever, for example, does not actually activate the wipers is detrimental to the feeling of presence in the VE. Thus, while locating the lever in space increases immersion, not being able to interact with it decreases the feeling of control and nullifies perceived affordance. In other words, participants found that the expected affordances in the environment were not in fact present. The problems with the lack of interaction were observed also with the radio task. Participants were not able to interactively explore the radio interface and started to doubt if the handheld controller was still functional, or if they were using it properly. The notion of expectations regarding the affordances of the VE should be operationalized in a measure, such as questionnaires before and after use. This measure could then be incorporated in the VE authenticity index.

Lack of physical feedback from interaction may, however, also be beneficial for HTI research in VEs. One important finding arose in the qualitative data which is not much reported in this paper. This referred to the relationship between motor functions and conscious experience and how this in turn supports the weighting of affordance and control on experience of presence. Motor functions are often implicit and an interruption in the implicitly expected tactile feedback causes explicit awareness of these motor functions [75]. This was visible in the verbalizations of the participants during the radio task. While the lack of tactile feedback decreases immersion, it also makes the interaction more explicit to the user. The increased awareness may cause the participant to notice problems, which she/he would have passed over in a real-life experiment. This notion may be useful for VE usability studies, focusing on certain interactions, since high fidelity prototypes are not necessarily a requirement for identifying usability problems [10].

In addition to the radio task, the inspection task, requiring the participants to step outside the car model, indicated problems with the validity of the VE experience. Navigating the space was only possible, if the participants' mental maps of the environment afforded this action-relation [76]. The participants had trouble understanding that they should have exited the car, which indicates that they were not completely —on the level of mental representation—immersed in the environment. Their representations of the VE did not afford the possibility of stepping outside the vehicle, or the impossibility of moving one's head through a solid door.

Problems associated with the technical limitations of the VE used in the study serve as a reminder for designers of VE experiments. Devising tasks, which involve the risk of technical problems, should be discouraged. Not only do technical problems during the interaction confound the results relating to the problematic task itself, but also the resulting distrust towards the environment may make all subsequent tasks confounded. Not only is it therefore the function of the environment in itself which allows for its authentic experience, but also authenticity is dependent on how well the tasks have been designed to conform to the limitations of the environment.

Table 3: Authenticity index.

Category	Factor
Immersion	Feeling of presence
	Anticipated affordances compared to fulfilled affordances
Control	Feeling of control
	Discovery ratio
	Amount of technical problems experienced

7. Conclusion

The goal of the study was to propose and develop a framework for evaluating the authenticity of a VE laboratory experiment, especially in the context of HTI. Analysis of questionnaire responses for the feeling of presence and feeling of control revealed several underlying authenticity subfactors. The suggested subfactors for the VE authenticity index are listed in Table 3.

In future studies, these subfactors should be included in the authenticity index as means of more accurately ascertaining the experience of presence in VEs. Furthermore, due to the indications of high reliability and construct validity indicated in the measures, the operationalization of the *feeling of presence* and *feeling of control* in this study can be viewed as sufficient. The factors should together capture the subjective experience of actually "being in" the VE and being able to interact with it naturally. *Discovery ratio* should be operationalized as the number of functions or features discovered either at the start of the experiment or throughout the experiment, divided by the number of discoverable functions or features. This ratio would hence indicate how well the participants were able to find the intended functions or features of the environment.

Anticipated affordances compared to fulfilled affordances should be operationalized as either a questionnaire, presented to the participants before and after the use, or a number of expected but not fulfilled affordances during the experiment. While a questionnaire scale would be easier to be reliably operationalized, the actual number of expected but not fulfilled affordances would be a more objective measure of how well the VE realized the interactive expectations of the participants. Of course, the latter method requires thinking aloud, and this might not always be a viable choice in an experiment. *Technical problems* should also be operationalized either as subjective assessment after use or in terms of the number of encountered problems, with the same caveats as in the affordances measure.

Although the above subfactors of authentic VE experience are considered as existing on an equal level, when computing the final factor using either principal component analysis or factor analysis, it should be noted that it is possible to construct more internally complex representations of authentic VE experience. For example, one could hypothesize that the discovery rate is the causal antecedent of the feeling of presence, while fulfilled affordances are the causal antecedent of the feeling of control. Using structural equation modelling

for such hypotheses would be, while an interesting study, very demanding on resources, as such a study would necessitate a large number of participants in order to be valid. However, understanding the causalities of VE experience is necessary for designing authentic environments, and hence these kinds of studies should be pursued.

The reason for using a VE laboratory in HTI research is that it simultaneously provides the flexibility of a simulated environment and the control of a laboratory experiment. A VE can be used to evaluate design prototypes in early stages of development. Changing the context of the evaluation, such as varying the scenarios, is relatively easy. Thus, VEs enable the study of user experience in diverse use contexts and situations already in the early stages of development. This is highly valuable from both the industrial and scientific perspectives as (a) substantial savings can be made from user experience findings and subsequent glitches before too much investment has been made in development and production and (b) contextual factors can be isolated and experimented in terms of gaining more precise data relating to influential elements in user experience and other interaction dynamics.

One aspect to remember however is that when using a VE laboratory to evaluate prototypes and interaction scenarios, it is important to indicate the extent to which the participant experiences reality in the simulated environment. While technological improvements and clever task design, which takes the technological limitations into consideration, potentially improve the authenticity of a VE study, the validity of the experiment should always be evaluated with a standardized framework. There are design implications for the authenticity index in relation to its capacity to provide data regarding the level of authenticity experienced in the VE contingent to specific purposes. Different validity measures for VE authenticity have been proposed, such as participant performance [49] and long questionnaires [19], but what is lacking is a common, easy, and fast framework for assessing the authenticity of the VE experience. The experiment reported here serves as a pilot study in constructing a metric for this purpose.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
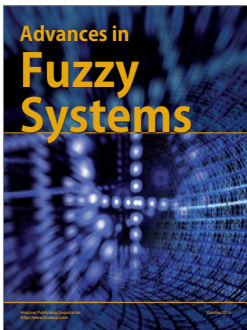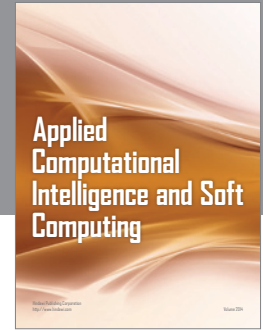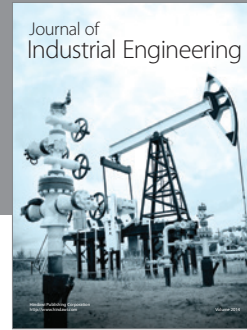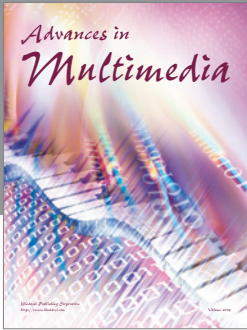
Acknowledgments

References

[1] I. D. Bishop and C. Stock, "Using collaborative virtual environments to plan wind energy installations," *Renewable Energy*, vol. 35, no. 10, pp. 2348–2355, 2010.

[2] A. Covaci, C. C. Postelnicu, A. N. Panfir, and D. Talaba, "A virtual reality simulator for basketball free-throw skills

development," in *Technological Innovation for Value Creation*, pp. 105–112, Springer, Berlin, Germany, 2012.

[3] M. J. Tarr and W. H. Warren, "Virtual reality in behavioral neuroscience and beyond," *Nature Neuroscience*, vol. 5, pp. 1089–1092, 2002.

[4] H. Wu, Z. He, and J. Gong, "A virtual globe-based 3D visualization and interactive framework for public participation in urban planning processes," *Computers, Environment and Urban Systems*, vol. 34, no. 4, pp. 291–298, 2010.

[5] J. M. Loomis, J. J. Blascovich, and A. C. Beall, "Immersive virtual environment technology as a basic research tool in psychology," *Behavior Research Methods, Instruments, & Computers*, vol. 31, no. 4, pp. 557–564, 1999.

[6] K. M. Stanney, M. Mollaghasemi, L. Reeves, R. Breaux, and D. A. Graeber, "Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design," *International Journal of Human Computer Studies*, vol. 58, no. 4, pp. 447–481, 2003.

[7] M. Hassenzahl and N. Tractinsky, "User experience—a research agenda," *Behaviour and Information Technology*, vol. 25, no. 2, pp. 91–97, 2006.

[8] L. Berkowitz and E. Donnerstein, "External validity is more than skin deep," *American Psychologist*, vol. 37, no. 3, pp. 245–257, 1982.

[9] U. Neisser, *Cognition and Reality*, W. H. Freeman and Company, San Francisco, Calif, USA, 1976.

[10] J. Sauer, K. Seibel, and B. Rüttinger, "The influence of user expertise and prototype fidelity in usability tests," *Applied Ergonomics*, vol. 41, no. 1, pp. 130–140, 2010.

[11] B. Dalgarno and M. J. W. Lee, "What are the learning affordances of 3-D virtual environments?" *British Journal of Educational Technology*, vol. 41, no. 1, pp. 10–32, 2010.

[12] F. Rebelo, P. Noriega, E. Duarte, and M. Soares, "Using virtual reality to assess user experience," *Human Factors*, vol. 54, no. 6, pp. 964–982, 2012.

[13] J. Nielsen, *Usability Engineering*, Elsevier, Amsterdam, The Netherlands, 1994.

[14] V. Roto, E. Law, A. P. O. S. Vermeeren, and J. Hoonhout, *User Experience White Paper. Bringing Clarity to the Concept of User Experience*, 2011, http://www.allaboutux.org/files/UX-White-Paper.pdf.

[15] R. Schroeder, "Defining virtual worlds and virtual environments," *Journal of Virtual Worlds Research*, vol. 1, no. 1, pp. 1–3, 2008.

[16] K. M. Stanney, G. Salvendy, J. Deisinger et al., "Aftereffects and sense of presence in virtual environments: formulation of a research and development agenda," *International Journal of Human-Computer Interaction*, vol. 10, no. 2, pp. 135–187, 1998.

[17] M. Matthen, "Two visual systems and the feeling of presence," in *Perception, Action, and Consciousness*, N. Gangopadhyay, M. Madary, and F. Spicer, Eds., pp. 107–124, Oxford University Press, Oxford, UK, 2010.

[18] J. J. Gibson, *An Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, Mass, USA, 1978.

[19] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: a presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[20] G. Calleja, "Immersion in virtual worlds," in *The Oxford Handbook of Virtuality*, pp. 222–236, Oxford University Press, Oxford, UK, 2014.

[21] M. Minsky, "Telepresence," OMNI Magazine, 1980, http://web.media.mit.edu/~minsky/papers/Telepresence.html.

[22] I. Sutherland, "The ultimate display. Information processing techniques," in *Proceedings of the IFIP Congress*, pp. 506–508, New York, NY, USA, 1965, http://worrydream.com/refs/Sutherland%20-%20The%20Ultimate%20Display.pdf.

[23] C. Klimmt and P. Vorderer, "Media Psychology 'is not yet there': introducing theories on media entertainment to the presence debate," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 4, pp. 346–359, 2003.

[24] K. Oatley, "A taxonomy of the emotions of literary response and a theory of identification in fictional narrative," *Poetics*, vol. 23, no. 1-2, pp. 53–74, 1995.

[25] P. Vorderer and U. Ritterfeld, "Children's future programming and media," in *The Faces of Televisual Media: Teaching, Violence, Selling to Children*, E. L. Palmer and B. Young, Eds., pp. 241–262, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2003.

[26] R. Oerter, *The Psychology of Play: An Activity Oriented Approach*, Quintessenz, Munich, Germany, 1993.

[27] D. Zillmann, "The psychology of suspense in dramatic exposition," in *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*, P. Vorderer, H. J. Wulff, and M. Friedrichsen, Eds., pp. 199–231, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.

[28] G. Riva, "The psychology of ambient intelligence: activity, situation and presence," in *Ambient Intelligence*, G. Riva, F. Vatalaro, F. Davide, and M. Alcañiz, Eds., chapter 2, pp. 17–33, IOS Press, Amsterdam, Netherlands, 2005.

[29] A. Spagnolli and L. Gamberini, "A place for presence. Understanding the human involvement in mediated interactive environments," *PsychNology Journal*, vol. 3, no. 1, pp. 6–15, 2005.

[30] T. Reiners, H. Teras, V. Chang et al., "Authentic, immersive, and emotional experience in virtual learning environments: the fear of dying as an important learning experience in a simulation," in *Proceedings of the Teaching and Learning Forum*, Perth, Australia, January 2014, https://ctl.curtin.edu.au/events/conferences/tlf/tlf2014/refereed/reiners.html.

[31] G. Riva, J. A. Waterworth, E. L. Waterworth, and F. Mantovani, "From intention to action: the role of presence," *New Ideas in Psychology*, vol. 29, no. 1, pp. 24–37, 2011.

[32] G. Riva and F. Mantovani, "Extending the self through the tools and the others: a general framework for presence and social presence in mediated interactions," in *Interacting with Presence: HCI and the Sense of Presence in Computer-Mediated Environments*, G. Riva, J. Waterworth, and D. Murray, Eds., pp. 9–31, DeGruyter, Berlin, Germany, 2014.

[33] M. Lombard and T. Ditton, "At the heart of it all: the concept of presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, 1997.

[34] K. Lee and C. Nass, "Designing social presence of social actors in human computer interaction," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*, pp. 289–296, Fort Lauderdale, Fla, USA, April 2003.

[35] W. A. Ijsselsteijn, J. Freeman, and H. de Ridder, "Presence: where are we?" *Cyberpsychology and Behavior*, vol. 4, no. 2, pp. 179–182, 2001.

[36] M. Matelli and G. Luppino, "Parietofrontal circuits for action and space perception in the macaque monkey," *NeuroImage*, vol. 14, no. 1, pp. S27–S32, 2001.

[37] N. Streitz and P. Nixon, "The dissappearing computer," *Communications of the ACM*, vol. 48, no. 3, pp. 32–35, 2005.

[38] D. A. Norman, *The Psychology of Everyday Things*, Basic Books, New York, NY, USA, 1988.

[39] D. A. Norman, "Affordance, conventions, and design," *Interactions*, vol. 6, no. 3, pp. 38–43, 1999.

[40] W. Barfield, D. Zeltzer, T. Sheridan, and M. Slater, "Presence and performance within virtual environments," in *Virtual Environments and Advanced Interface Design*, W. Barfield and T. Furness, Eds., pp. 473–513, Oxford University Press, New York, NY, USA, 1995.

[41] J. Häkkinen, T. Kawai, J. Takatalo, R. Mitsuya, and G. Nyman, "What do people look at when they watch stereoscopic movies?" in *Stereoscopic Displays and Applications XXI*, vol. 7524 of *Proceedings of SPIE*, San Jose, Calif, USA, February 2010.

[42] T. Ni, D. A. Bowman, and J. Chen, "Increased display size and resolution improve task performance in information-rich virtual environments," in *Proceedings of the Graphics Interface (GI '06)*, pp. 139–146, Canadian Information Processing Society, Quebec, Canada, June 2006.

[43] H. Q. Dinh, N. Walker, L. F. Hodges, C. Song, and A. Kobayashi, "Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments," in *Proceedings of the IEEE Virtual Reality (VR '99)*, pp. 222–228, March 1999.

[44] C. H. Lewis and M. J. Griffin, "Human factors consideration in clinical applications of virtual reality," in *Virtual Reality in Neuro-Psycho-Physiology*, G. Riva, Ed., pp. 35–56, IOS Press, Amsterdam, The Netherlands, 1997.

[45] W. Ijsselsteijn, H. de Ridder, J. Freeman, S. E. Avons, and D. Bouwhuis, "Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 298–311, 2001.

[46] D. Murphy and F. Neff, "Spatial sound for computer games and virtual reality," in *Game Sound Technology and Player Interaction: Concepts and Developments*, M. Grimshaw, Ed., pp. 287–312, Information Science Reference, Hershey, Pa, USA, 2011.

[47] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Proceedings of the IEEE Virtual Reality Conference (VR '10)*, pp. 147–153, IEEE, Waltham, Mass, USA, March 2010.

[48] G. Riva, "Is presence a technology issue? Some insights from cognitive sciences," *Virtual Reality*, vol. 13, no. 3, pp. 159–169, 2009.

[49] E. Dubois, L. P. Nedel, C. M. D. S. Freitas, and L. Jacon, "Beyond user experimentation: notational-based systematic evaluation of interaction techniques in virtual reality environments," *Virtual Reality*, vol. 8, no. 2, pp. 118–128, 2005.

[50] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst, "Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques," *Computer Graphics Forum*, vol. 17, no. 3, pp. 41–52, 1998.

[51] D. M. Johnson, *Introduction to and Review of Simulator Sickness Research*, Army Research Institute, Field Unit, Fort Rucker, Ala, USA, 2005.

[52] G. E. Riccio and T. A. Stoffregen, "An ecological theory of motion sickness and postural instability," *Ecological Psychology*, vol. 3, no. 3, pp. 195–240, 1991.

[53] J. LaViola Jr., "A discussion of cybersickness in virtual environments," *ACM SIGCHI Bulletin*, vol. 32, no. 1, pp. 47–56, 2000.

[54] J. Reason and J. Brand, *Motion Sickness*, Academic Press, London, UK, 1975.

[55] H. Gleitman, *Basic Psychology*, Norton and Company, New York, NY, USA, 1992.

[56] J. Dichgans and T. Brandt, "Visual-vestibular interaction: effects on self-motion perception and postural control," in *Perception*, R. Held, H. Leibowitz, and H. Teuber, Eds., vol. 8 of *Handbook of Sensory Physiology*, pp. 755–804, Springer, Berlin, Germany, 1978.

[57] J. Casali, "Vehicular stimulation-induced sickness, vol. 1: an overview," IEBOR Technical Report 8501, Naval Training Systems Center, Orlando, Fla, USA, 1986.

[58] J. E. Domeyer, N. D. Cassavaugh, and R. W. Backs, "The use of adaptation to reduce simulator sickness in driving assessment and research," *Accident Analysis & Prevention*, vol. 53, pp. 127–132, 2013.

[59] H. B.-L. Duh, D. E. Parker, and T. A. Furness, "An 'independent visual background' reduced balance disturbance envoked by visual scene motion: implication for alleviating simulator sickness," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, pp. 85–89, ACM, Seattle, Wash, USA, March-April 2001.

[60] P. A. Howarth and S. G. Hodder, "Characteristics of habituation to motion in a virtual environment," *Displays*, vol. 29, no. 2, pp. 117–123, 2008.

[61] C. Dede, "The evolution of constructivist learning environments: immersion in distributed, virtual worlds," *Educational Technology*, vol. 35, no. 5, pp. 46–52, 1995.

[62] C. Guetl, T. Reiners, S. Gregory et al., "Operationalising gamification in an educational authentic environment," in *Proceedings of the IADIS International Conference on Internet Technologies & Society (ITS '12)*, Perth, Australia, September 2012.

[63] M. Roussou, "Learning by doing and learning through play: an exploration of interactivity in virtual environments for children," *Computers in Entertainment*, vol. 2, no. 1, p. 10, 2004.

[64] S. Y. Hu and G. M. Liao, "Scalable peer-to-peer networked virtual environment," in *Proceedings of 3rd ACM SIGCOMM Workshop on Network and System Support for Games (NetGames '04)*, pp. 129–133, Portland, Ore, USA, August 2004.

[65] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology," *Psychological Inquiry*, vol. 13, no. 2, pp. 103–124, 2002.

[66] G. Riva, "Virtual reality in neuroscience: a survey," in *Virtual Environments in Clinical Psychology and Neuroscience*, vol. 58 of *Studies in Health Technology and Informatics*, pp. 191–199, IOS Press, Amsterdam, The Netherlands, 1998.

[67] M. A. Moser and D. MacLeod, *Immersed in Technology: Art and Virtual Environments*, MIT Press, Cambridge, Mass, USA, 1996.

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

[69] J. Häkkinen, M. Pölönen, J. Takatalo, and G. Nyman, "Simulator sickness in virtual display gaming: a comparison of stereoscopic and non-stereoscopic situations," in *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06)*, pp. 227–230, ACM, Espoo, Finland, September 2006.

[70] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[71] R. Patterson, M. D. Winterbottom, and B. J. Pierce, "Perceptual issues in the use of head-mounted visual displays," *Human Factors*, vol. 48, no. 3, pp. 555–573, 2006.

[72] T. Shibata, "Head mounted display," *Displays*, vol. 23, no. 1-2, pp. 57–64, 2002.

[73] K. A. Ericsson and H. A. Simon, *Protocol Analysis*, MIT Press, Cambridge, Mass, USA, 1993.

[74] D. Murphy and F. Neff, "Spatial sound for computer games and virtual reality," in *Game Sound Technology and Player Interaction: Concepts and Developments*, M. Grimshaw, Ed., pp. 287–312, Information Science Reference, Hershey, 2011.

[75] A. Baddeley, *Working Memory, Thought, and Action*, Oxford University Press, Oxford, UK, 2007.

[76] B. S. Santos, P. Dias, A. Pimentel et al., "Head-mounted display versus desktop for 3D navigation in virtual reality: a user study," *Multimedia Tools and Applications*, vol. 41, no. 1, pp. 161–181, 2009.