Emotional Response–Based Approach for Assessing the Sense of Presence of Subjects in Virtual Building Evacuation Studies

Hao Zou[1]; Nan Li, M.ASCE[2]; and Lijun Cao[3]

Abstract: Using the latest advances in virtual reality (VR) technologies, it is possible to create immersive virtual environments (IVEs) that can be used for behavioral experiments in controlled setups to examine individual evacuee behaviors. As a first step to achieving sufficient ecological validity, this study proposes an approach for assessing the sense of presence of the subjects performing virtual evacuation tasks, and this assessment can serve as an indirect indicator of the overall validity of the experiments. The proposed approach integrates a set of subjective and objective measures to assess the subjects' emotional responses, including emotional valence and emotional arousal, which are highly reflective of their sense of presence in the IVEs. Three IVEs are developed, presenting an apartment under normal or fire emergency situations with different levels of realism. The subjects were asked to perform an evacuation task in all IVEs, and their emotional responses were monitored and analyzed throughout the experiments. The results show that the proposed approach can not only detect changes in subjects' emotions triggered by the IVEs but can differentiate subjects' emotions when they were immersed in IVEs with different levels of realism. These findings suggest that it is feasible to use the proposed approach to assess the sense of presence experienced by subjects conducting virtual evacuation tasks and ultimately the level of ecological validity of IVE-based evacuation studies. **DOI: 10.1061/(ASCE)CP.1943-5487.0000679.**

Author keywords: Emergency evacuation; Virtual reality; Ecological validity; Sense of presence; Emotional assessment.

Introduction

The behaviors of evacuees have significant impact on the overall efficiency of evacuation tasks during building emergencies. A wide range of evacuation simulation models has been proposed in prior research (Zheng et al. 2009) and a number of commercial simulation tools have been developed which, by implementing these models, can simulate crowd behaviors of evacuees during building emergencies and assess the efficiency of evacuation processes (Ozel 1987; Feinberg and Johnson 1995; Thompson and Marchant 1995; Santos and Aguirre 2004). However, such simulation models and tools generally bear a major limitation: an oversimplification and lack of justification of various assumptions about behavioral patterns of individual particles and agents in the models, which represent individual evacuees in reality. This limitation prevents, to a large extent, fine-grained simulation and prediction of evacuation processes and evaluation of evacuation management measures. For example, the evacuees' personal histories, skills, interpersonal relationships, and other attributes may influence their responses to emergencies during the crisis, and these attributes and their impacts on evacuees' behaviors are difficult to fully take into account in current simulations. Consequently, many of the individual behavioral patterns that have been observed in past evacuation instances, such as herding, competition and collaboration, and leadership following, cannot be fully explained and properly modelled. Because human systems are emergent systems where the emergence of behavioral rules on a global scale is formed through individual interactions on a local scale (Charlotte 2005), the individual behaviors are fundamental to collective behaviors and require extensive further research. To examine the behaviors of individual evacuees, several approaches, including postemergency investigations (Zhao et al. 2009; Urbina and Wolshon 2003), evacuation drills (Cheng et al. 2009; Kobes et al. 2010), and animal subject–based experiments (Soria et al. 2012; Saloma et al. 2003), have been used in prior research. However, these approaches all suffer from certain limitations, such as scarcity and incompleteness of real behavioral data, high cost, difficulty in setting a controlled behavioral experiment environment, and debatable similarity between human and animal subjects, preventing comprehensive examination of individual evacuees' behaviors.

Recent development of virtual reality (VR) technologies has provided a promising alternative approach to conduct evacuation experiments for evacuee behavioral studies. Virtual reality is a "real or simulated environment in which the perceiver experiences telepresence" (Steuer 1992). It is advantageous over existing approaches because it allows for the creation of low-risk, cost-efficient, and highly controllable immersive virtual environments (IVEs) for conducting virtual evacuation experiments. The advantages of IVEs have been well demonstrated in various behavioral psychology studies (Yu et al. 2013; Ioannou et al. 2014), including several studies that focused on evacuation behaviors (Ronchi et al. 2015, 2016). One limitation of IVE-based behavioral experiments for which they are mostly criticized, however, is the uncertainty about their ecological validity. Ecological validity refers to the extent to which experiment subjects' perceptions and responses can be generalized to real-life settings (Brewer et al. 2000). Ecological

[1]Graduate Student, Dept. of Construction Management, Tsinghua Univ., Beijing 100084, China.

[2]Associate Professor, Dept. of Construction Management, Tsinghua Univ., Beijing 100084, China (corresponding author). E-mail: nanli@tsinghua.edu.cn

[3]Graduate Student, Dept. of Construction Management, Tsinghua Univ., Beijing 100084, China.

Note. This manuscript was submitted on September 6, 2016; approved on January 9, 2017; published online on March 25, 2017. Discussion period open until August 25, 2017; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

validity represents the authenticity, applicability, and generalizability of the research; experiments should have high ecological validity so that conclusions drawn based on the experiment results can be generalized to real-life scenarios instead of being limited to the experimental settings. Although ecological validity can be assessed in most cases by directly comparing experiment subjects' perceptions and responses in IVEs and in real-life scenarios against which the IVEs are created, it is almost impossible to do so for evacuation studies, because, as previously mentioned, creating such real-life scenarios and exposing the subjects to these scenarios is ethically and costly prohibitive. Therefore there is an urgent need for an approach to assess the ecological validity of IVE-based evacuation experiments in order to establish the applicability of IVEs in evacuee behavioral studies and justify the generalizability of their findings.

Prior research has pointed out that the sense of presence of human subjects is a fundamental factor that determines the usefulness of virtual environments and the ecological validity of IVE-based studies (Slater et al. 2009). Therefore the level of presence perceived by human subjects in a virtual environment can be an effective indicator for evaluation of the ecological validity of IVE-based studies (Schuemie et al. 2001). Accordingly, this study proposes a new emotional response–based approach for assessing the sense of presence of subjects in IVE-based evacuation behavioral studies. The approach integrates subjective and objective measures, including an emotion scale and two physiological indicators, to assess both emotional valence and emotional arousal of the subjects, which are reflective of their sense of presence when they are immersed in the virtual environments. The feasibility of using this approach for sense-of-presence assessment is demonstrated in this study with its application in a virtual building-evacuation experiment which involved three IVEs that presented a building under normal or fire emergency situations with different levels of realism.

Literature Review

Evacuee Behavioral Studies

A wide range of evacuation simulation models has been developed in prior research. Based on their rationale and level of detail in modeling evacuee behaviors, these models can be broadly categorized as cellular automata models, lattice gas models, social force models, fluid-dynamic models, agent-based models, and game theoretic models (Zheng et al. 2009). These models simulate the evacuation process by defining various parameters of the pedestrians and environments, such as the velocity of movement and the number of exits, or by defining the rules of interactions between environments and pedestrians and among pedestrians. A number of commercial simulation tools have also been developed to simulate crowd behaviors of evacuees. Typical tools include *BGRAF* (Ozel 1987), *FIRESCAP* (Feinberg and Johnson 1995), *SIMULEX* (Thompson and Marchant 1995), and *EXODUS* (Santos and Aguirre 2004). Widely adopted in fire-fighting training and evacuation drills and education, these evacuation models and simulation tools have been able to explain and predict evacuees' behaviors and the overall evacuation processes to some extent. However, due to an oversimplification regarding behavioral patterns of individual particles and agents, and lack of justification of various assumptions about evacuee behaviors, most of these models and tools are not able to provide fine-grained simulation and accurate prediction of evacuee behaviors and overall evacuation processes.

To examine the behaviors of individual evacuees, the most widely used approach is postemergency investigation. Such investigations usually involve interviews with survivors (Zhao et al. 2009) and analysis of incident data such as surveillance videos and official incident reports (Urbina and Wolshon 2003) to restore the emergency scenes and evacuation processes. This approach can yield reasonable behavioral analysis based on real data but is largely limited by the scarcity and incompleteness of such data due to the lack of existing records and poor availability (Tobin 1997). Alternatively, evacuation drills, including announced drills (Cheng et al. 2009) and unannounced drills (Kobes et al. 2010), have been used in evacuee behavioral studies. In these drill-based studies, experiment subjects are instructed to respond to imagined building emergencies. The emotional and behavioral responses of the subjects are analyzed based on videos and questionnaires collected during the drills. However, such drills, especially the announced drills that involve scripts provided to subjects in advance, are usually criticized for their significant difference compared with real fire scenarios and their inability to cause the same stress-coping behaviors of evacuees (Cheng et al. 2009). Moreover, it is costly to conduct drills and challenging to exactly maintain all control variables, making it difficult to reveal and justify causal relationships between dependent and independent variables, which are key to explaining evacuee behaviors. To address these issues and avoid possible ethical and legal concerns that conducting evacuation experiments in real manufactured emergencies may cause physiological and psychological damages to human subjects, recent research has also used animal subjects, such as ants (Soria et al. 2012) and mice (Saloma et al. 2003), in evacuation behavioral studies. By exposing the animal subjects to real hazards, such as fire and smoke, and observing their responses, researchers are able to examine their individual and collective evacuation behaviors. Then, based on the assumption that both human and other biological entities have similar emergency reactions (Shiwakoti et al. 2009), researchers postulate the behaviors of human subjects if they were exposed to the same hazardous environments. While these experiments are effective in overcoming the limitation of experiments that involve human subjects, the behavioral similarities and dissimilarities between humans and animals are still largely debatable (Shiwakoti and Sarvi 2013), calling for significant caution when generalizing conclusions of these experiments to human evacuees.

Application of VR Technologies in Emergency Evacuation Research

The use of VR technologies in evacuation studies can date back to the 1990s when the serious games, regarded by some researchers as the first generation of VR, became popular due to the development of human–machine interaction and image processing technologies. Serious games refers to games designed for primary purposes such as training, simulation, or education rather than pure entertainment (Susi et al. 2007). They have been widely used by various domains such as education, scientific exploration, and engineering. For emergency management in particular, serious games can provide a virtual environment in which people can conduct evacuation tasks similar to those in real-life settings. For instance, evacuation training and drill games were developed to help people become familiar with emergency environments and learn necessary evacuation skills (Chittaro 2012; Smith and Trenholme 2009). Major limitations of serious games, however, include their two-dimensional, nonimmersive presentation of computer-generated environments and their limited tools to enrich human–computer interactions that are essential to enhancing users' sense of presence. The more recent development of VR technologies has significantly enriched the tools that researchers can use to build realistic, immersive, and interactive IVEs. For instance, creating virtual environments with more-immersive senses of vision, audio, and touch, and more-realistic human–computer interaction has become possible by using various VR equipment such as head-mounted displays (HMD), cave automatic virtual environment (CAVE), and motion capture systems. The advancement of VR technologies has resulted in an increasing volume of interest in applying IVE to evacuation studies. A number of IVE-based evacuation systems have been developed in prior studies, such as *InterFIRE VR* (Putorti Jr and McElroy 2000), *SGEM* (Lo et al. 2004), and *Vegas* (Xi and Smith 2014). These applications were mostly developed for firefighting training (Lee et al. 2010) and evacuation training (Wiederhold and Wiederhold 2008), with a particular focus on environments with high potential for accidents, such as mines (Tichon and Burgess-Limerick 2011) and tunnels (Sharma et al. 2014).

More recently, with their increased sense of realism, IVEs have been applied to evacuee behavioral studies. For instance, evacuees' decision-making processes, such as route choice (Kinateder et al. 2014a), waiting time (Andrée et al. 2016), and helping behavior (Gamberini et al. 2015), were investigated with IVE-based experiments. The impact factors of evacuee behaviors, such as social impact factors (Kinateder et al. 2014b) and environmental impact factors (Duarte et al. 2014), were also studied in experiments conducted in IVEs. Techniques such as Bayesian networks were adopted to model the relationship between the impact factors and evacuees' behaviors (Musharraf et al. 2016). Moreover, by integrating VR with building information modeling (BIM) technology, which can provide detailed data about geometry, materials, and functions of building elements (Azhar 2011), the effect of building conditions on occupants' behavior during the evacuation process has also been studied (Rüppel and Schatz 2011). Examples include the VR platform developed by Wang et al. (2014) which can provide real-time fire-evacuation guidance based on the dynamically changing building information provided by BIM technology.

Despite the improved sense of realism of IVEs enabled by the latest technological advancement, the gap between IVEs and real evacuation scenarios remains a major challenge to the ecological validity of IVE-based evacuation studies. To address this challenge, rather than to eliminate the gap, it is more important and urgent to develop an approach to assess the level of ecological validity of such studies. The assessment can be used to establish benchmarks for different evacuation studies with varying requirements and evaluate the effectiveness of various VR technologies that can contribute to the level of realism of building evacuation IVEs.

Measurement of the Sense of Presence in IVEs

The ecological validity of IVE-based studies can be regarded satisfactory if experiment subjects' behavioral reactions in the IVEs are similar to those in the real world (Anderson and Bushman 1997). The level of similarity is usually measured by comparing subjects' behaviors in IVEs to their behaviors in corresponding real-world settings in studies about, e.g., building design optimization (Heydarian et al. 2015), driving-behavior analysis (Shechtman et al. 2009), and interpersonal-behavior analysis (Iachini et al. 2016). However, this direct measurement of ecological validity is much less applicable in evacuation behavioral studies because it is highly difficult, if possible at all, to create real building emergency scenarios, without causing significant damages and public panic, to match the scenarios in the IVEs for benchmarking purpose. Therefore alternative approaches are needed for ecological validity assessment of IVE-based evacuation behavioral studies.

As was previously mentioned, the level of presence perceived by human subjects is an effective indicator of the ecological validity of IVE-based studies (Schuemie et al. 2001). Several approaches have been proposed in the literature to measure the level of presence. The most widely used approach is using either questionnaires (Witmer and Singer 1998) or interviews (De Leo et al. 2014). Although this approach is simple and convenient to implement, it is highly subjective and lacks sufficient precision. Alternatively, because the sense of human beings is controlled by the brain, the reaction of the brain to the virtual environments is considered to be a more objective and sensitive reflection of humans' sense of presence. Based on this assumption, some researchers have used electroencephalogram (EEG) devices to measure brain activations as an indirect metric of the sense of presence (Baumgartner et al. 2006; Clemente et al. 2014). However, the EEG devices are usually costly, cumbersome, and difficult to operate (Rodríguez et al. 2015). A third approach is to measure the emotional response of the subjects interacting with virtual environments. There is increasing evidence in the literature that shows noticeable correlation between the sense of presence of the subjects and their emotions when immersed in virtual environments (Riva et al. 2007). Assessing the emotional response of the subjects in IVE-based experiments provides a promising way of assessing the subjects' sense of presence and ultimately the level of ecological validity of IVE-based evacuation experiments.

Integrated Approach for Assessing Emotional Responses

Emotional assessment in general includes the assessment of emotional valence, which is the hedonic value of a specific emotion, and the assessment of emotional arousal, which is the intensity of a specific emotion. There is no single measure that can provide fine-grained assessment of both the emotional valence and emotional arousal at the same time. An emotion scale is a typical subjective measure for emotional assessment in psychological studies (Watson et al. 1988). Several self-reporting emotion scales have been proposed and tested in prior research (Luigi et al. 2015). An emotion scale allows subjects to assess the type of their own emotions by completing a deliberately designed survey; therefore it can be used as an effective measure to assess the emotional valence of the subjects. Although they are convenient to use, emotion scales are sometimes criticized due to their subjectivity because subjects may differ in their comprehension of the evaluation standard (Mardaga et al. 2006). Emotion scales also lack sufficient sensitivity to subtle emotional fluctuations (Wilson and Sasse 2000), which prevents accurate assessment of emotional arousal.

Interpretation of certain types of physiological data from the peripheral nervous system (PNS) of the human body is another emerging method that has attracted noticeable attention in recent years due to its objectivity. The PNS can be divided into the somatic nervous system (SNS) and the autonomic nervous system (ANS). The SNS can be controlled by the human mind, and it regulates body movements consciously. The ANS acts largely unconsciously, which makes it a promising objective measure of emotional responses in psychological studies (Stephens et al. 2010), and it has been widely used in recent studies. There are several types of physiological indicators, such as cardiovascular, respiratory, and electrodermal measures, that can be used to assess ANS activities. Kreibig (2010) summarized the use of these physiological indicators and their varying patterns under different emotions, and concluded that distinguishing different types of positive and negative valence based on ANS activities remains a challenging task. However, the magnitude of emotional arousal has been proven to have clear relevance to certain types of ANS activity.

For instance, skin conductance (SC) (Kallinen 2004) was reported to have positive correlation with the magnitude of emotional arousal. The low-frequency (LF) and high-frequency (HF) power of heart-rate variability (HRV) were reported as able to respectively reflect the activities of two parts of ANS, the sympathetic and parasympathetic nervous systems (Appelhans and Luecken 2006).

Built on current literature in emotional assessment, this study proposes the integration of both subjective measures (emotion scale) and objective measures (physiological indicators) to assess the emotional valence and the emotional arousal, respectively, of subjects in IVE-based evacuation experiments. It is hypothesized that these two measures when used together can effectively distinguish specific types of positively and negatively valenced emotions and can accurately capture subtle changes in the magnitude of emotional responses. The assessment reported by the proposed approach can be used as an indicator of the subjects' sense of presence in virtual building-evacuation tasks, which is reflective of the overall ecological validity of the IVE-based evacuation studies. A preliminary version of this approach was reported in (Zou et al. 2016), after which the approach has been improved by a new set of experiments that have been conducted for the present study.

Specifically, the emotion scale used in the proposed integrated approach is the positive affect and negative affect scale (PANAS), which is a widely adopted emotion scale in the field of psychology. Designed by Watson et al. (1988), PANAS is composed of twenty adjectives, including ten adjectives describing positive emotions and ten adjectives describing negative emotions. A subject is asked to evaluate the extent of each of these emotions using a scale from 1 to 5, with 1 meaning the least and 5 the strongest, according to his/her present emotional state. The total scores of all positive and all negative emotions respectively represent the extent of the subject's overall positive and negative emotional state. The original PANAS survey, along with a Chinese translation that was proposed in (Zhang 2001) and well cited by following studies, was presented to experiment subjects, whose native language was Chinese, in this study to ensure precise comprehension of the adjectives by the subjects. The physiological indicators used in the proposed approach to measure the ANS activity included galvanic skin response (GSR) and HRV. Galvanic skin response refers to changes in the electrical properties of the skin, and HRV refers to variation of the time interval between heartbeats. Both indicators can be assessed with different measures. Because these two indicators are applied here to the emotional responses triggered by virtual evacuation tasks for the first time in the literature to the best of the authors' knowledge, the effectiveness of these measures was unknown, and their variation in response to different types of emotion could differ in patterns that are yet to be investigated (Kreibig 2010). Therefore different measures of the GSR and HRV were tested in this study, and the effective measures were selected based on the experiment results. Specifically, the GSR can be reflected by the level of SC, which is used to describe the conductivity of skin. The SC accounts for both skin conductance level (SCL), which is a slowly varying tonic activity and can be measured by skin conductance tonic (SCT) score, and skin conductance response (SCR), which is a fast-varying phasic activity and can be measured by skin conductance phasic (SCP) score. The SC score is equal to the SCT score plus the SCP score. These three GSR measures were captured using a multichannel physiological recorder in the experiments. The HRV can be analyzed from both the time domain and the frequency domain. Given that the typical duration of a virtual evacuation task in this study was around one minute, short-term measures of HRV in the time domain, including mean heart rate (HR), standard deviation of the normal-to-normal intervals (SDNN), and square root of the mean squared differences of successive normal-to-normal intervals (RMSSD), and short-term measures of the frequency domain, including LF power, HF power, and LF/HF, were tested in this study.

Virtual Evacuation Experiment

A virtual evacuation experiment was conducted in this study to validate the feasibility of using the proposed approach for sense-of-presence assessment. In order to determine the main factors that impact people's perception of building emergency scenes, semistructured interviews were first conducted with first responders from three fire stations in Beijing, staff from two fire museums in Beijing and Qingdao that had virtual fire scenario exhibits, and developers from a leading VR company in China. Based on these interviews, flame, smoke, and sound associated with, e.g., burning and fire alarms were identified as the main factors that impact how people perceive fire emergency scenes and that could be modeled with current VR technologies. In addition, most interviewees were positive about the feasibility of introducing IVEs in evacuee behavioral studies and were excited about the potential value of the likely outcomes of such studies in improving building evacuation preparedness and supporting onsite evacuation management operations. Taking the interview results into consideration, three IVEs were developed for the experiment based on a single-bedroom apartment scenario. The basic IVE (hereafter referred to as IVE-b) presented the apartment under a normal situation. The other two IVEs presented the apartment under fire emergency situations. Intended to create different levels of realism, these two IVEs had exactly the same visual representation of building elements, but different visual and audio representations of hazard elements. Specifically, these two IVEs differed in terms of the visual effects and audio effects associated with fire, smoke, and fire alarms, which were selected based on the results of the aforementioned interviews. In one IVE (hereafter referred to as IVE-l), the visualized flame and smoke effects had a limited diffusion range and relatively blurry rendering, and there was no background sound at all. In the other IVE (hereafter referred to as IVE-h), the visualized flame and smoke effects had an extended diffusion range blocking half of the passage, and the rendering was done in high resolution. In addition, fire alarms were added to the background sound in IVE-h. Screenshots of all three IVEs are shown in Fig. 1. Basic geometries of the apartment were modeled and rendered in *3ds Max*. The rendered model was imported into *Unity 3D*, in which the components of fire hazards, including flame, smoke, and glow effects, were created using the embedded particle system, and stereo combustion sound effects and fire alarms were added using audio sources and the embedded listener system. A first-person controller was created in *Unity 3D* to enable user-IVE interactions such as navigation and door opening. An Oculus Rift DK2 HMD (Oculus VR, Menlo Park, California) and a Bose QC20 noise cancelling headphone (Framingham, Massachusetts) were used to immerse subjects into the virtual environment, and a Microsoft Xbox joystick (Redmond, Washington) was used by the subjects to navigate in the virtual environment. The computer used for creating the IVEs and driving them during the experiment was a Dell Precision T7800 workstation with Microsoft Windows 7 operating system, Intel Xeon E5-2603 processors, NVIDIA Quadro K620 graphics card, and 16 GB memory.

The physiological data were collected with GSR sensors that were bound around subjects' index and middle fingers, and HRV sensors that were attached to their earlobes. The time window for reporting GSR and HRV data in each virtual evacuation task was one minute starting from the moment the subject started moving in

J. Comput. Civ. Eng., 2017, 31(5): 04017028

IVE-b IVE-l IVE-h

Fig. 1. Screenshots of all three IVEs

the IVE. One minute was the average time required for each subject to complete the task. During the entire experiment procedure the subjects' physiological conditions, interactions with IVEs, and first-person views from the HMD were monitored in real time and recorded using ERGOLAB platform (Kingfar International, Beijing) (Fig. 2).

Nine subjects were recruited for a pilot experiment in which the subjects provided feedback that helped improve the design of the IVEs and experiment procedure. Then a total of 40 subjects participated in the experiment. These subjects included 28 undergraduate students and 12 graduate students from Tsinghua University. They included 27 males and 13 females between 17 and 29 years old with majors including civil engineering, architecture, electrical engineering, and journalism. The experiments were conducted in a sound-attenuated room with temperature set to 25°C. If a subject felt any motion sickness in the IVE, he/she could ask to terminate the experiment at any time. Each subject went through the following experiment procedure: at the beginning of the experiment, the subject put on the GSR and HRV sensors and the HMD, and then trained in IVE-b, in order to become familiar with the view in the HMD and learn how to navigate. This training was limited to the bedroom space of IVE-b in order to minimize the practice effect (Heiman 1995) of the subject that might influence his/her behaviors in the next steps of the experiment. After the training was completed, the subject took a five-minute rest, after which his/her baseline physiological conditions were recorded. Next, the subject was instructed to repeat an evacuation task in IVE-b, IVE-l, and IVE-h. The pilot experiment only included IVE- l and IVE-h, the comparison between which allowed for the assessment of the proposed approach in differentiating the emotional responses under IVEs with different levels of realism. However, the pilot experiment results showed that the emotions of the subjects could be impacted by their sense of curiosity as they navigated through a virtual environment for the first time. Therefore, the evacuation task in IVE-b was added to the formal experiment in order to eliminate the impact of curiosity when the subjects repeated the task in IVE-l and IVE-h. In addition, to avoid the impact of the order of tasks on subjects' emotions, each subject was randomly assigned to either Order 1 (IVE-b, IVE-l, IVE-h) or Order 2 (IVE-b, IVE-h, IVE-l). The evacuation task was purposely designed to be simple so that the subject would find himself/herself always under comparable situations when repeating the task in different IVEs: the subject would be positioned in a bedroom at the beginning of each task, with a health warning message about the use of a HMD displayed in front of him/her.

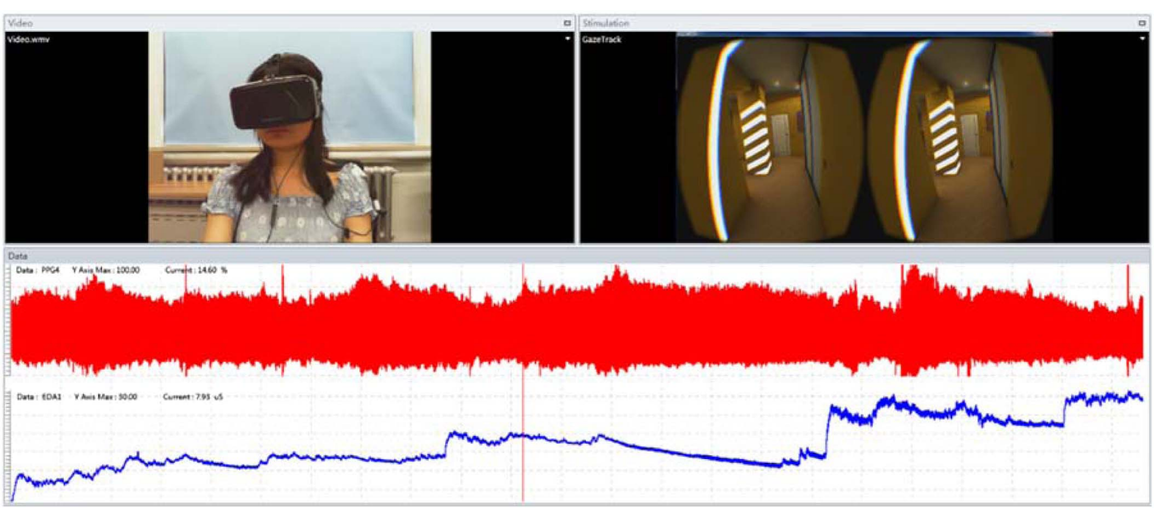

Fig. 2. Interface of the experiment data collection platform (image by Hao Zou)

The message would disappear after a few seconds, and the subject would be instructed to evacuate from the apartment as soon as possible. To do that, the subject would have to open a door and leave the bedroom, walk through a short hallway, bypass a living room, and exit from a front door. In IVE-b, the apartment was in a normal situation, while in IVE-l and IVE-h part of the hallway started burning and the subject would have to keep a certain distance from the fire during the evacuation. After each task was completed, the subject filled out a PANAS survey to report his/her current emotion and rested until his/her physiological conditions restored to the baseline status.

Findings

The entire experiment procedure took about 30 minutes for each subject to complete. All of the subjects completed the procedure, and a total of 40 data sets were collected for analysis. The experiment results were analyzed to evaluate the feasibility of using the proposed approach to assess the emotional response of subjects in IVE-based evacuation experiments. The evaluation was conducted by validating two hypotheses: H1, that the proposed approach can detect subjects' emotions triggered by IVE; and H2, that the proposed approach can differentiate subjects' emotions when they are immersed in IVEs with different levels of realism. Paired t-tests at 95% confidence level were conducted to validate the hypotheses, and the results are summarized in Tables 1 and 2. Detailed analysis of the results and validation of the hypotheses are presented in the rest of this section.

Analysis of Positive Scores of PANAS

The assessment of ten types of positive emotion included in the PANAS scale was analyzed first. A comparison of the total score of all positive emotions reported by the subjects showed that the overall positive emotion increased by 6.45% when the level of realism of the IVE increased. As t-test results in Table 1 show, the subjects experienced statistically stronger positive emotions in IVE-h compared to IVE-l ($t = -3.198$, $p = 0.003$), rejecting the null hypothesis that subjects' positive emotions reported by the emotion scale were the same across the two IVEs. These results suggested that H2 was valid in terms of emotion valence. The average scores of the ten types of emotion under both IVE-l and IVE-h are depicted in Fig. 3. As can be seen in the figure, for four types of emotion—excited, strong, alert, and determined—there was an obvious discrepancy between the average scores in IVE-l and those in IVE-h. The t-test results shown in Table 1 further confirm that the average scores of these types of emotion were statistically higher in IVE-h than in IVE-l. Excited, strong, alert, and determined are all related to the degree of mental concentration, and their magnitude was expected to increase when the subjects experienced deeper immersion as they conducted the virtual evacuation task. Therefore the results suggest that the PANAS scale can detect the types of positive emotion that are sensitive to the level of realism of the IVEs, which adds to the evidence for the validation of H2. However, the PANAS scores could barely identify any statistical difference between the overall positive emotion in IVE-b and that in IVE-l ($t = 1.994$, $p = 0.053$), suggesting that the sensitivity of the emotion scale was limited to certain extent.

Analysis of Negative Scores of PANAS

Similarly, the assessment of ten types of negative emotion included in the PANAS scale was analyzed. A comparison of the total score of all negative emotions reported by the subjects showed that the overall negative emotion increased by 27.80% when the level of realism of the IVE increased. As t-test results in Table 1 show, the subjects experienced significantly stronger negative emotions in IVE-h compared with IVE-l ($t = -5.080$, $p = 0.000$), rejecting the null hypothesis that subjects' negative emotions reported by the emotion scale were the same across the two IVEs, which again validated H2 in terms of emotion valence. The average scores of the ten types of emotion under both IVE-l and IVE-h are depicted in Fig. 4. As shown in the figure, the subjects reported that they felt more scared, afraid, distressed, nervous, jittery, irritable, upset, and guilty in IVE-h than in IVE-l. The t-test results shown in Table 1

Table 1. Paired t-Test between Three Sets of Assessment Using Subjective Measures

Emotions		Baseline and IVE-l		Baseline and IVE-h		IVE-l and IVE-h	
		t	p	t	p	t	p
Positive	Interested	3.873	0.000	2.896	0.000	−0.240	0.812
	Excited	0.476	0.637	−1.669	0.103	−2.644	0.012
	Strong	2.211	0.033	−1.138	0.262	−4.416	0.000
	Enthusiastic	2.379	0.022	3.286	0.002	1.290	0.205
	Proud	1.820	0.076	1.571	0.124	−0.339	0.736
	Alert	1.525	0.135	−1.674	0.102	−3.317	0.002
	Active	1.854	0.071	0.842	0.405	−0.781	0.440
	Inspired	0.725	0.534	1.325	0.473	0.703	0.486
	Determined	1.571	0.124	−0.561	0.578	−2.655	0.011
	Attentive	0.758	0.453	−0.598	0.553	−1.651	0.107
	Total	3.085	0.004	0.777	0.442	−3.198	0.003
Negative	Scared	1.045	0.303	−2.553	0.015	−4.714	0.000
	Afraid	1.497	0.142	−2.310	0.026	−5.414	0.000
	Distressed	1.578	0.123	−1.793	0.081	−3.894	0.000
	Nervous	3.518	0.001	0.443	0.660	−4.051	0.000
	Jittery	0.829	0.412	−1.568	0.125	−2.806	0.008
	Irritable	1.597	0.118	−0.183	0.855	−2.360	0.023
	Ashamed	5.309	0.000	5.439	0.000	0.000	1.000
	Hostile	−0.771	0.446	−1.183	0.244	−1.000	0.323
	Upset	−0.502	0.618	−2.464	0.018	−2.393	0.022
	Guilty	1.525	0.135	−0.723	0.474	−2.467	0.018
	Total	2.810	0.008	−1.353	0.184	−5.080	0.000

Table 2. Paired t-Test between Three Sets of Assessment with Objective Measures

Physiological indicators		Baseline and IVE-l		Baseline and IVE-h		IVE-l and IVE-h	
		t	p	t	p	t	p
GSR	SC	−11.024	0.000	−12.157	0.000	−6.901	0.000
	SCT	−10.208	0.000	−10.975	0.000	−4.973	0.000
	SCP	−1.057	0.297	−1.392	0.172	−0.687	0.496
HRV	AVNN	2.654	0.011	3.503	0.001	2.301	0.027
	SDNN	−5.401	0.000	−5.010	0.000	0.462	0.646
	RMSSD	−3.165	0.003	−1.888	0.066	−1.273	0.211
	LF	3.735	0.001	3.942	0.000	1.592	0.119
	HF	−4.940	0.000	−5.861	0.000	−0.844	0.404
	LF/HF	6.504	0.000	7.028	0.000	2.531	0.016

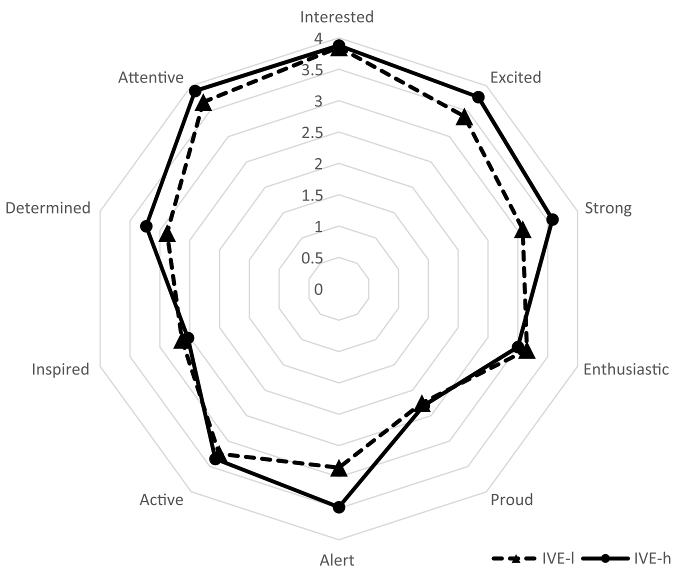

Fig. 3. Average scores of the 10 types of positive emotion

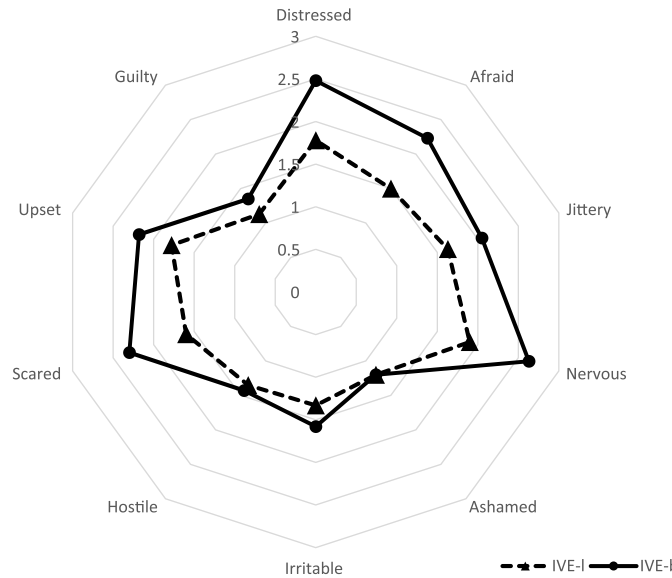

Fig. 4. Average scores of the ten types of negative emotion

further confirm that the average scores of these types of emotion were statistically higher in IVE-h in IVE-l. These emotions were likely related to the mental disorder elicited by emergencies, and their magnitude was likely expected to increase in response to deeper immersion in the virtual evacuation task. Therefore the results suggest that the PANAS scale can detect the types of negative emotion that are sensitive to the level of realism of the IVEs, which is consistent with the finding about positive emotions and provides further evidence of the validity of H2. Similar to the analysis of positive PANAS scores, the PANAS scores could not identify any statistical difference between the overall negative emotion in IVE-b and that in IVE-l ($t = 0.735$, $p = 0.467$), suggesting that the sensitivity of the emotion scale was limited to certain extent.

Analysis of GSR

The GSR was measured with the SC score as well as its breakdown, including the SCT score and the SCP score. The experiment results were analyzed to identify the effective GSR measures that could contribute to the proposed emotion-assessment approach by effectively recognizing whether the subjects were immersed in IVE and differentiating the GSR conditions of the subjects immersed in IVE-l as opposed to IVE-h.

The average SC, SCT, and SCP scores of all subjects during the baseline period and when the subjects were immersed in IVE-l and IVE-h are summarized in Table 3. As can be seen in the table, all of the scores increased from the baseline to IVE-l and to IVE-h. The pairwise t-test was conducted to further assess the statistical significance of the differences between the scores. Based on t-test results shown in Table 2, it can be concluded that the SC and SCT scores were significantly higher in IVE-h and IVE-l compared with the baseline, and that the SC and SCT scores were significantly higher in IVE-h compared with IVE-l. In other words, by using the SC and SCT scores, which reflect the overall and slowly varying skin conductivity, respectively, to measure the GSR, the proposed approach could not only detect subjects' emotions triggered by the IVEs, which validates H1, but could also differentiate subjects' emotions when they were immersed in IVE-l and IVE-h that had different levels of realism, which validates H2. The t-test results also show that there was no statistically significant pairwise difference between SCP scores in the baseline and in the two IVEs. This means that the SCP, which reflects the instantaneous fast activity of GSR

Table 3. Average Scores of Three Sets of GSR

Physiological indicators of GSR	Baseline	IVE-l	IVE-h
SC (μS)	4.65	8.41	9.14
SCT (μS)	4.04	7.52	8.18
SCP (μS)	0.61	0.88	0.96

and which can be impacted by various nonemotion-related factors such as finger movement, is not suitable for the proposed approach.

Analysis of HRV

The HRV reflects the variation of the time interval between heartbeats of a subject. The HRV can be measured from either the time domain or the frequency domain. Six different HRV measures used in previous studies were calculated based on the physiological data collected in the experiment. These measures were AVNN, SDNN, and RMSSD of the time domain and LF, HF, and LF/HF of the frequency domain. The results were analyzed to identify the effective HRV measures that could recognize whether the subjects were immersed in IVE and effectively differentiate the HRV conditions of the subjects immersed in IVE-l as opposed to IVE-h. Such effective HRV measures can be integrated into the proposed emotion assessment approach.

The average scores of all six measures for all subjects during the baseline period and when the subjects were immersed in IVE-l and IVE-h are summarized in Table 4. As can be seen in the table, RMSSD, and HF increased from the baseline to IVE-l and to IVE-h, whereas AVNN, LF, and LF/HF decreased. These trends were generally consistent with the trends reported in previous studies (Fabes et al. 1993; Fuller 1992; Gorman and Sloan 2000; Murakami and Ohira 2007; Sgoifo et al. 2003). A pairwise t-test was conducted to further assess the statistical significance of the difference between the scores. Based on t-test results shown in Table 2, it can be concluded that the scores of all six measures were significantly higher in IVE-h and IVE-l compared with the baseline and that the AVNN and LF/HF scores were significantly higher in IVE-h compared with IVE-l. In other words, by using the AVNN and LF/HF scores to measure the HRV, the proposed approach could not only detect subjects' emotions triggered by the IVEs, which validates H1, but could also differentiate subjects' emotions when they were immersed in IVE-l and IVE-h that had different levels of realism, which validates H2. The other four measures could only partially fulfill the requirement, lacking the capability to differentiate the HRV conditions of the subjects when they were immersed in IVEs with different levels of realism.

Discussions and Limitations

As the results show, the proposed approach composed of PANAS, GSR, and HRV could detect subjects' emotional response when they were immersed in IVEs with different levels of realism. One underlying assumption, based on which the major conclusions of this study are drawn, is that when subjects experienced IVE-h, their emotional responses were likely to be stronger than when they experienced IVE-l. There are two reasons why this could be a reasonable assumption. First, the PANAS scale is a mature emotion scale that has been repeatedly tested and validated in psychology. Subjective assessment using the PANAS scale clearly suggested that subjects experienced stronger emotional responses when they were immersed in IVE-h. Second, there was a close connection between the results of subjective assessment and those of objective assessment, suggesting good internal validity of the experiment results. Specifically, analysis of the six measures (PANAS positive emotions, PANAS negative emotions, SC, SCT, AVNN, and LF/HF) of the individual subjects revealed that for 38 (or 95%) of the 40 subjects, at least four measures agreed with each other in terms of how they changed in response to the different levels of realism between IVE-l and IVE-h. The internal validity was also enhanced by efforts made to conduct unbiased selection of the subjects and their random assignment to one of the two experiment orders, and to maintain constant experiment settings, apparatus, and procedure for all the subjects.

However, it needs to be pointed out that when subjects are immersed in relatively more realistic IVEs, they do not necessarily experience stronger emotions. For instance, in a world that is overloaded with highly graphically detailed but unrealistic artistic media, the level of realism is not the only determinant of the emotions of the subjects. Stronger emotions might actually be elicited by discomfort caused by a virtual environment being too unreal, as has been demonstrated in the gaming and movie industry. While the IVEs created in this study were intended to be made real, except for the hazard elements including flame, smoke and fire alarm effects that were purposefully differentiated between the two IVEs and made unreal in IVE-l, it was still possible for some subjects to have experienced discomfort. Hence interpreting the results would require certain caution. Another limitation of this study is that the experiment subjects were all university students, and therefore the results are tied to this particular demographic group. Because the demographics of the subjects might impact their prior experience with and adaptability to IVEs and their resulting emotional responses, further experiments would be needed to generalize the findings of this study to other demographic groups.

Table 4. Average Scores of Six Sets of HRV

Physiological indicators of HRV	Baseline	IVE-l	IVE-h
AVNN (ms)	813.15	784.81	769.94
SDNN (ms)	47.27	70.55	68.68
RMSSD (ms)	42.95	57.71	83.68
LF (ms^2)	428.48	334.31	302.68
HF (ms^2)	97.23	222.51	233.48
LF/HF	6.75	2.31	1.68

Conclusions

This study proposed an approach for assessing emotional response of subjects in virtual evacuation experiments as an indicator of the subjects' sense of presence, a critical factor that determines the overall level of ecological validity of evacuation experiments conducted in IVEs. The approach included subjective and objective measures for assessing emotional valence and emotional arousal, respectively. The widely used PANAS emotion scale was used as the subject measure. Three measures of the GSR and six measures of the HRV were tested in this study, and the SC, SCT, AVNN, and LF/HF were found to be effective objective measures. Three IVEs were developed in this study which presented an apartment under normal or fire emergency situations with different levels of realism. Experiment subjects were asked to perform an evacuation task in all IVEs and their emotional responses were monitored throughout the experiments using the subjective and objective measures. The results show that the proposed approach could not only detect changes in subjects' emotions triggered by IVEs, but could also differentiate subjects' emotions when they were immersed in IVEs with different levels of realism. In particular, the PANAS survey revealed that the magnitude of four types of positive emotion, including excited, strong, alert, and determined, and eight types of negative emotion, including scared, afraid, distressed, nervous, jittery, irritable, upset, and guilty, increased significantly when the level of realism of the IVE increased. Moreover, the SC, SCT, AVNN, and LF/HF scores were reflective of whether the subjects were immersed in the IVE and whether the level of realism of the IVE was low or high. The findings suggest that it is feasible to use the proposed approach to assess the sense of presence experienced by subjects conducting virtual evacuation tasks, and ultimately the level of ecological validity of IVE-based evacuation studies.

Acknowledgments

This material is based upon work supported by the National Natural Science Foundation of China (NSFC) under Grant No. 71603145 and the Humanities and Social Sciences Foundation of the Ministry of Education (MOE) of China under Grant No. 16YJC630052. The authors thank the NSFC and MOE for their support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSFC and MOE. The authors would also like to thank Kingfar International, Inc. for providing the ERGOLAB platform that was used in this study, and Shengxing Ban, Lizhi Bao, and Po Tian for their technical support in the acquisition of physiological data and development of IVEs.

References

3ds Max [Computer software]. Autodesk, San Rafael, CA.

Anderson, C. A., and Bushman, B. J. (1997). "External validity of 'trivial' experiments: The case of laboratory aggression." *Rev. Gen. Psychol.*, 1(1), 19–41.

Andrée, K., Nilsson, D., and Eriksson, J. (2016). "Evacuation experiments in a virtual reality high-rise building: Exit choice and waiting time for evacuation elevators." *Fire Mater.*, 40(4), 554–567.

Appelhans, B. M., and Luecken, L. J. (2006). "Heart rate variability as an index of regulated emotional responding." *Rev. Gen. Psychol.*, 10(3), 229–240.

Azhar, S. (2011). "Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry." *Leadersh. Manage. Eng.*, 10.1061/(ASCE)LM.1943-5630.0000127, 241–252.

Baumgartner, T., Valko, L., Esslen, M., and Jäncke, L. (2006). "Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and psychophysiology study." *CyberPsychol. Behav.*, 9(1), 30–45.

Brewer, M. B., Reis, H., and Judd, C. (2000). "Research design and issues of validity." *Handbook of research methods in social and personality psychology*, Cambridge University Press, New York, 3–16.

Charlotte, H. (2005). *Self-organization and evolution of social systems*, Cambridge University Press, Cambridge, U.K.

Cheng, X., Zhang, H., Xie, Q., Zhou, Y., Zhang, H., and Zhang, C. (2009). "Study of announced evacuation drill from a retail store." *Build. Environ.*, 44(5), 864–870.

Chittaro, L. (2012). "Passengers' safety in aircraft evacuations: Employing serious games to educate and persuade." *Persuasive technology: Design for health and safety*, Springer, Berlin, 215–226.

Clemente, M., Rodríguez, A., Rey, B., and Alcañiz, M. (2014). "Assessment of the influence of navigation control and screen size on the sense of presence in virtual reality using EEG." *Expert Syst. Appl.*, 41(4), 1584–1592.

De Leo, G., Diggs, L. A., Radici, E., and Mastaglio, T. W. (2014). "Measuring sense of presence and user characteristics to predict effective training in an online simulated virtual environment." *Simul. Healthcare*, 9(1), 1–6.

Duarte, E., Rebelo, F., Teles, J., and Wogalter, M. S. (2014). "Behavioral compliance for dynamic versus static signs in an immersive virtual environment." *Appl. Ergon.*, 45(5), 1367–1375.

Fabes, R. A., Eisenberg, N., and Eisenbud, L. (1993). "Behavioral and physiological correlates of children's reactions to others in distress." *Dev. Psychol.*, 29(4), 655–663.

Feinberg, W. E., and Johnson, N. R. (1995). "FIRESCAP: A computer simulation model of reaction to a fire alarm." *J. Math. Sociol.*, 20(2–3), 247–269.

Fuller, B. (1992). "The effects of stress-anxiety and coping styles on heart rate variability." *Int. J. Psychophysiol.*, 12(1), 81–86.

Gamberini, L., Chittaro, L., Spagnolli, A., and Carlesso, C. (2015). "Psychological response to an emergency in virtual reality: Effects of victim ethnicity and emergency type on helping behavior and navigation." *Comput. Human Behav.*, 48, 104–113.

Gorman, J. M., and Sloan, R. P. (2000). "Heart rate variability in depressive and anxiety disorders." *Am. Heart J.*, 140(4), S77–S83.

Heiman, G. W. (1995). *Research methods in psychology*, Houghton Mifflin, Boston.

Heydarian, A., Carneiro, J. P., Gerber, D., Becerik-Gerber, B., Hayes, T., and Wood, W. (2015). "Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations." *Autom. Constr.*, 54, 116–126.

Iachini, T., Coello, Y., Frassinetti, F., Senese, V. P., Galante, F., and Ruggiero, G. (2016). "Peripersonal and interpersonal space in virtual and real environments: Effects of gender and age." *J. Environ. Psychol.*, 45, 154–164.

Ioannou, I., Avery, A., Zhou, Y., Szudek, J., Kennedy, G., and O'Leary, S. (2014). "The effect of fidelity: How expert behavior changes in a virtual reality environment." *Laryngoscope*, 124(9), 2144–2150.

Kallinen, K. (2004). "Emotion related psychophysiological responses to listening music with eyes-open versus eyes-closed: Electrodermal (EDA), electrocardiac (ECG), and electromyographic (EMG) measures." *Proc., Music Perception and Cognition*, Causal Productions, Sydney, Australia, 299–301.

Kinateder, M., et al. (2014a). "Social influence on route choice in a virtual reality tunnel fire." *Transp. Res. Part F. Traffic Psychol. Behav.*, 26, 116–125.

Kinateder, M., Müller, M., Jost, M., Mühlberger, A., and Pauli, P. (2014b). "Social influence in a virtual tunnel fire-influence of conflicting information on evacuation behavior." *Appl. Ergon.*, 45(6), 1649–1659.

Kobes, M., Helsloot, I., de Vries, B., Post, J. G., Oberijé, N., and Groenewegen, K. (2010). "Way finding during fire evacuation: An analysis of unannounced fire drills in a hotel at night." *Build. Environ.*, 45(3), 537–548.

Kreibig, S. D. (2010). "Autonomic nervous system activity in emotion: A review." *Biol. Psychol.*, 84(3), 394–421.

Lee, J., Cha, M., Choi, B., and Kim, T. (2010). "A team-based firefighter training platform using the virtual environment." *Proc., 9th ACM SIGGRAPH Conf. on Virtual-Reality Continuum and Its Applications in Industry*, ACM, New York, 299–302.

Lo, S., Fang, Z., Lin, P., and Zhi, G. (2004). "An evacuation model: The SGEM package." *Fire Saf. J.*, 39(3), 169–190.

Luigi, M., Massimiliano, M., Aniello, P., Gennaro, R., and Virginia, P. R. (2015). "On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces." *Energy Procedia*, 78, 471–476.

Mardaga, S., Laloyaux, O., and Hansenne, M. (2006). "Personality traits modulate skin conductance response to emotional pictures: An investigation with Cloninger's model of personality." *Personality Individual Differences*, 40(8), 1603–1614.

Murakami, H., and Ohira, H. (2007). "Influence of attention manipulation on emotion and autonomic responses." *Perceptual Motor Skills*, 105(1), 299–308.

Musharraf, M., Smith, J., Khan, F., Veitch, B., and MacKinnon, S. (2016). "Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian network approach." *Reliab. Eng. Syst. Saf.*, 152, 28–37.

Ozel, F. (1987). "The computer model 'BGRAF': A cognitive approach to emergency egress simulation." Ph.D. dissertation, Univ. of Michigan, Ann Arbor, MI.

Putorti, A., Jr., and McElroy, J. (2000). "Full-scale house fire experiment for InterFIRE VR." *Rep. of Test FR4009*, U.S. Dept. of Commerce, National Institute of Standards and Technology, Washington, DC.

Riva, G., et al. (2007). "Affective interactions using virtual reality: The link between presence and emotions." *CyberPsychol. Behav.*, 10(1), 45–56.

Rodríguez, A., Rey, B., Clemente, M., Wrzesien, M., and Alcañiz, M. (2015). "Assessing brain activations associated with emotional regulation during virtual reality mood induction procedures." *Expert Syst. Appl.*, 42(3), 1699–1709.

Ronchi, E., et al. (2015). "Evacuation travel paths in virtual reality experiments for tunnel safety analysis." *Fire Saf. J.*, 71, 257–267.

Ronchi, E., et al. (2016). "A virtual reality experiment on flashing lights at emergency exit portals for road tunnel evacuation." *Fire Technol.*, 52(3), 623–647.

Rüppel, U., and Schatz, K. (2011). "Designing a BIM-based serious game for fire safety evacuation simulations." *Adv. Eng. Inf.*, 25(4), 600–611.

Saloma, C., Perez, G. J., Tapang, G., Lim, M., and Palmes-Saloma, C. (2003). "Self-organized queuing and scale-free behavior in real escape panic." *Proc. Natl. Acad. Sci. U.S.A.*, 100(21), 11947–11952.

Santos, G., and Aguirre, B. E. (2004). "A critical review of emergency evacuation simulation models." *Proc., Conf. on Building Occupant Movement during Fire Emergencies*, NIST, Gaithersburg, MD.

Schuemie, M. J., van der Straaten, P., Krijn, M., and van der Mast, C. A. P. G. (2001). "Research on presence in virtual reality: A survey." *CyberPsychol. Behav.*, 4(2), 183–201.

Sgoifo, A., et al. (2003). "Cardiac autonomic reactivity and salivary cortisol in men and women exposed to social stressors: Relationship with individual ethological profile." *Neurosci. Biobehav. Rev.*, 27(1), 179–188.

Sharma, S., Jerripothula, S., Mackey, S., and Soumare, O. (2014). "Immersive virtual reality environment of a subway evacuation on a cloud for disaster preparedness and response training." *2014 IEEE Symp. on Proc., Computational Intelligence for Human-Like Intelligence*, IEEE, Piscataway, NJ, 1–6.

Shechtman, O., Classen, S., Awadzi, K., and Mann, W. (2009). "Comparison of driving errors between on-the-road and simulated driving assessment: A validation study." *Traffic Inj. Prev.*, 10(4), 379–385.

Shiwakoti, N., and Sarvi, M. (2013). "Understanding pedestrian crowd panic: A review on model organisms approach." *J. Transp. Geogr.*, 26, 12–17.

Shiwakoti, N., Sarvi, M., Rose, G., and Burd, M. (2009). "Enhancing the safety of pedestrians during emergency egress: Can we learn from biological entities?" *Transp. Res. Rec.*, 2137, 31–37.

Slater, M., Lotto, B., Arnold, M. M., and Sánchez-Vives, M. V. (2009). "How we experience immersive virtual environments: The concept of presence and its measurement." *Anuario De Psicología*, 40, 193–210.

Smith, S. P., and Trenholme, D. (2009). "Rapid prototyping a virtual fire drill environment using computer game technology." *Fire Saf. J.*, 44(4), 559–569.

Soria, S., Josens, R., and Parisi, D. (2012). "Experimental evidence of the 'faster is slower' effect in the evacuation of ants." *Saf. Sci.*, 50(7), 1584–1588.

Stephens, C. L., Christie, I. C., and Friedman, B. H. (2010). "Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis." *Biol. Psychol.*, 84(3), 463–473.

Steuer, J. (1992). "Defining virtual reality: Dimensions determining telepresence." *J. Commun.*, 42(4), 73–93.

Susi, T., Johannesson, M., and Backlund, P. (2007). "Serious games: An overview." *Technical Rep. HS-IKI-TR-07-001*, School of Humanities and Informatics, Univ. of Skövde, Sweden.

Thompson, P. A., and Marchant, E. W. (1995). "Testing and application of the computer model 'SIMULEX.'" *Fire Saf. J.*, 24(2), 149–166.

Tichon, J., and Burgess-Limerick, R. (2011). "A review of virtual reality as a medium for safety related training in mining." *J. Health Saf. Res. Pract.*, 3(1), 33–40.

Tobin, G. A. (1997). *Natural hazards: Explanation and integration*, Guilford Press, New York.

Unity 3D [Computer software]. Unity Technologies, San Francisco.

Urbina, E., and Wolshon, B. (2003). "National review of hurricane evacuation plans and policies: A comparison and contrast of state practices." *Transp. Res. Part A. Policy Pract.*, 37(3), 257–275.

Wang, B., Li, H., Rezgui, Y., Bradley, A., and Ong, H. N. (2014). "BIM based virtual environment for fire emergency evacuation." *Sci. World J.*, 2014, 589016.

Watson, D., Clark, L. A., and Tellegen, A. (1988). "Development and validation of brief measures of positive and negative affect: The PANAS scales." *J. Personality Soc. Psychol.*, 54(6), 1063–1070.

Wiederhold, B. K., and Wiederhold, M. D. (2008). "Virtual reality for posttraumatic stress disorder and stress inoculation training." *J. Cyber Ther. Rehabil.*, 1(1), 23–35.

Wilson, G. M., and Sasse, M. A. (2000). "Do users always know what's good for them? Utilising physiological responses to assess media quality." *People and computers XIV—Usability or else!* Springer, Berlin, 327–339.

Witmer, B. G., and Singer, M. J. (1998). "Measuring presence in virtual environments: A presence questionnaire." *Presence, Teleoperators Virtual Environ.*, 7(3), 225–240.

Xi, M., and Smith, S. P. (2014). "Simulating cooperative fire evacuation training in a virtual environment using gaming technology." *Proc., 2014 IEEE Virtual Reality*, Piscataway, NJ, 139–140.

Yu, Y., El Kamel, A., and Gong, G. (2013). "Modeling intelligent vehicle agent in virtual reality traffic simulation system." *Proc., 2013 2nd Int. Conf. on Systems and Computer Science*, IEEE, Piscataway, NJ, 274–279.

Zhang, Z. (2001). "Behavioral medicine inventory manual." *Chin. J. Behav. Med. Sci.*, 10(10), 19–24.

Zhao, C., Lo, S. M., Zhang, S., and Liu, M. (2009). "A post-fire survey on the pre-evacuation human behavior." *Fire Technol.*, 45(1), 71–95.

Zheng, X., Zhong, T., and Liu, M. (2009). "Modeling crowd evacuation of a building based on seven methodological approaches." *Build. Environ.*, 44(3), 437–445.

Zou, H., Li, N., and Cao, L. (2016). "Immersive virtual environments for investigating building emergency evacuation behaviors: A feasibility study." *Proc., Int. Symp. on Automation and Robotics in Construction*, IAARC, Bratislava, SK.