Immersive Virtual Environments for Investigating Building Emergency Evacuation Behaviors: A Feasibility Study

Hao Zou [a], Nan Li [a] and Lijun Cao [a]

[a]Department of Construction Management, Tsinghua University, China
E-mail: zouh14@mails.tsinghua.edu.cn, nanli@tsinghua.edu.cn, caolj12@mails.tsinghua.edu.cn

Abstract –

Evacuee behaviors have significant influence on the efficiency of emergency evacuation in buildings. While crowd evacuee behaviors have been extensively examined with various simulation models, the understanding of individual evacuee behaviors has to date mostly relied on sociological models, which are usually subjective and descriptive. Using 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. One challenge in conducting such experiments is to ensure their ecological validity, so that the decisions and actions made by the experiment subjects are what they would actually make in reality. As a first step to achieve sufficient ecological validity, this study examines the feasibility of using a combination of subjective and objective measures, including an emotion scale and a physiological indicator, to assess the emotional responses of subjects in IVE-based evacuation experiments. Two IVEs were developed in this study, both representing a fire emergency scenario in an apartment but having different levels of realism and hence different levels of ecological validity. Subjects were asked to perform an evacuation task in both IVEs, and their emotional responses were monitored and analyzed throughout the experiments. Statistical difference was observed between the subjects' emotional responses when exposed to the two IVEs. The results suggested it was feasible to assess the ecological validity by assessing the emotional response of the subjects in IVE-based evacuation experiments.

Keywords –

Emergency evacuation; Virtual reality; Evacuation experiment; Sense of presence; Feasibility study

1 Introduction

When people evacuate in buildings during emergencies such as fire and explosions, their individual and collective behaviors have significant impact on the overall efficiency of the evacuation process, which in turn impacts the chances of occurrence of fatalities and injuries. The crowd behaviors in evacuation has been extensively examined in prior research, resulting in a variety of simulation models such as those based on social forces, multi-agent modeling and cellular automata. These models, however, have largely relied on various assumptions and over simplification about behavioral patterns of individual particles and agents, which represent individual evacuees in reality, limiting to a large extent fine-grained simulation and prediction of evacuation processes and assessment of evacuation management measures. Despite their importance in evacuation studies, the behaviors of individual evacuees have to date not received the attention they deserve, as evidenced by the absence of computable and verifiable behavioral models. Prior research on individual evacuee's behaviors mostly employs one of the following approaches, including evacuation drills, incident investigation and questionnaire-based psychological experiments. However, these approaches all bear certain limitations, such as high costs, lack of details of emergency scenarios, and difficulty in isolating controlled variables, which limit the collection of adequate behavioral data that are needed for in-depth behavioral analysis of individual evacuees.

Recent development of virtual reality (VR) technologies has provided a promising solution to these limitations, allowing for the creation of low-cost, low-risk, and controllable immersive virtual environments (IVEs) for virtual evacuation experiments. The advantages of IVEs have been well demonstrated in various behavioral psychology research [1], including studies on evacuation behaviors [2]. A number of IVE-based evacuation systems have been proposed in the literature, such as Vegas [3] and interFIRE VR [4], and

used for firefighting training or evacuation simulation. IVE-based behavioral experiments, however, are usually challenged by its ecological validity, namely the extent to which experiment subjects' perception and responses can be generalized to real-life settings [5]. 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 [6]. 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 [7], driving behavior analysis [8], and interpersonal behavior analysis [9]. 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.

Prior research has pointed out that the sense of presence of the subjects is a fundamental factor that determines the ecological validity of IVE-based studies [10]. There is also evidence that shows noticeable correlation between the sense of presence of experiment subjects and their emotions when immersed in virtual environments [11]. Therefore, assessing the emotional response of the subjects in IVE-based experiments provides a possible way of assessing the subjects' sense of presence and hence the overall ecological validity of the experiments and ultimately the level of ecological validity of IVE-based evacuation experiments. This paper examines the feasibility of using a combination of subjective and objective measures, including an emotion scale and a physiological indicator, to assess emotional responses, which are comprised of emotional valence and emotional arousal, of the subjects in IVE-based evacuation experiments. Two IVEs were developed in this study. They both represented a fire emergency scenario in an apartment, but they had different levels of realism, and were expected to cause different emotional responses. The combined emotional assessment was applied to distinguish the emotional responses of experiment subjects elicited by the two IVEs when performing virtual evacuation tasks. The remainder of the paper is organized as follows: section 2 reviews relevant literature, followed by section 3 that introduces a combined method for emotional assessment. Section 4 presents the development of two building fire emergency IVEs with two different levels of realism. The design and implementation of a simplified IVE-based evacuation experiment is explained in section 5 and the results are presented in section 6. Section 7 concludes the paper.

2 Literature Review

Emotional response 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. Self-report emotion scale is a typical subjective measure for emotional assessment in psychological studies. It allows subjects to assess the type and magnitude of their own emotions by filling a deliberately designed survey. Several self-report emotion scales have been proposed and tested in prior research [12]. While they are convenient to use, emotion scales are challenged due to its subjectivity, as different subjects may have different comprehension about the evaluation standard [13], and their lack of sensitivity to subtle emotional fluctuations [14]. On the other hand, certain types of physiological reactions, such as brain activations and facial muscle activity, have found to be correlated with and sensitive to human emotions, and have hence been adopted in prior research for assessing the sense of presence of subjects in IVE-based studies [15]. For instance, autonomic nervous system (ANS) activity is an autonomic activity that is not controlled by subjective factors, and it has been proven a direct reflection of subjects' emotional response [16]. There is a number of types of physiological indicators, such as cardiovascular, respiratory and electrodermal measures, that can be used to measure ANS activities. Kreibig [17] summarized the use of these physiological indicators and their varying patterns under different emotions. It was reported 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 with certain types of physiological data. For instance, skin conductance (SC) [18] and minute ventilation (VE) [19] were 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 to respectively reflect the activity of two parts of ANS including sympathetic and parasympathetic nervous system [20]. Despite these findings, there still lacks a widely adopted approach to analyzing physiological data for ANS activity assessment and ultimately emotion responses assessment of subjects in IVE-based experiments.

In short, various approaches have been proposed in literature for assessing emotional valence or emotional arousal. However, there lacks an integrated approach to assess both elements of emotions applicable to typical emotional responses observed in virtual evacuation experiment settings, where further research is needed.

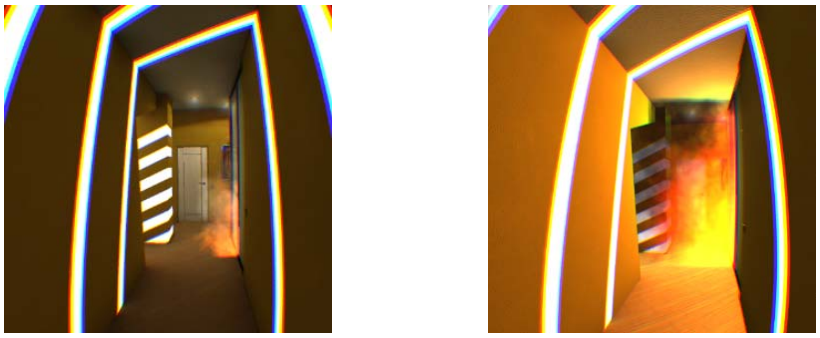
3 A Combined Method for Assessing Emotional Responses

As discussed above, self-report emotion scale is the most widely used measure of emotional responses, especially for assessing emotional valance. However, it bears certain limitations including high subjectivity, and low sensitivity to emotional changes. Physiological indicators, on the other hand, have proven relevance with emotional arousal, and are used in prior research as an alternative measure of emotional responses. This study proposes to use a combination of subjective measure (emotion scale) and objective measure (physiological indicator) to assess the emotional responses 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 accurately capture subtle changes in the magnitude of emotional responses.

Specifically, the emotion scale used in this study is the Positive Affect and Negative Affect Scale (PANAS) scale, which is widely adopted in the field of psychology. Designed by Watson and Clark in 1980s [21], PANAS scale is composed of twenty adjectives, including ten adjectives describing positive emotions and the other ten adjectives describing negative emotions. Subjects are 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 their present emotional state. The total scores of all positive and all negative emotions respectively represent the extent of the subjects' overall positive and negative emotional state. A Chinese translation proposed in [22] was presented to experiment subjects in this study along with the original PANAS survey to aid precise comprehension of the adjectives. The physiological indicator used in this study is the SC, which is used to describe the conductivity of skin. The SC can be characterized by skin conductance level (SCL), which is a slowly varying tonic activity, and skin conductance response (SCR), which is a fast varying phasic activity, and the SC score is equal to SC tonic score plus SC phasic score. The SCL and SCR were measured using a multichannel physiological recorder in the experiments.

4 Development of Two Building Fire Emergency IVEs

Two IVEs were developed in this study to validate the feasibility of using the proposed method for ecological validity assessment. Both IVEs simulated a fire emergency scenario in a one-bedroom apartment. Basic geometries of the apartment were modeled in Revit, and exported to 3ds Max for rendering, by adding photorealistic materials, texture, lighting and furniture. The rendered model was imported into Unity 3D, in which the components of fire hazards, including flame, smoke and glow effect, were created using the embedded particle system. A first-person controller was also created in Unity 3D that enabled user-IVE interactions such as navigation and door opening. Stereo combustion sound effect and fire alarms were also added using audio source and listener system in Unity 3D. An Oculus Rift DK2 head-mounted display (HMD) and a Bose QC20 noise cancelling headphone were used to immerse subjects into the virtual environment, and a Microsoft Xbox joystick was used by the subjects to navigate in the virtual environment. The only difference of the two IVEs was their level of realism: for all components of the fire hazards, one IVE (hereafter referred to as IVE 1) contained flame effect only and the other one (hereafter referred to as IVE 2) contained flame, smoke, glow effect as well as the sound of combustion and fire alarm. Screenshots of both IVEs are shown in Figure 1. The computer used for IVE creation and evacuation experiments was Dell Precision T7800 workstation with Microsoft Windows 7 operating system, Intel Xeon E5-2603 processors, NVIDIA Quadro K620 graphics card and 16GB of installed memory.



(a) (b)

Figure 1. IVEs with low (a) and high (b) level of realism

5 Virtual Evacuation Experiment

A total of 35 subjects were recruited to participate in the experiments. The subjects were all engineering students from Tsinghua University, including 25 male and 10 female between 18 and 28 years old. The experiments were conducted in a sound-attenuated room. Every subject went through the following procedure: at the beginning of the experiment, the subject wore SC data acquisition sensors by binding two electrodes around his/her index finger and middle finger, and filled in a PANAS survey. The subject then took a three-minute rest during which his/her baseline physiological data were recorded. Next, the subject put on the HMD, and took a training in the IVE, which contained no fire hazards, to learn how to navigate around. After another three-minute rest, the subject was asked to complete an evacuation task in IVE 1 and then IVE 2, with a three-minute rest between the two evacuation tasks. The subject's physiological conditions, interactions with IVE, and first-person view from the HMD were monitored and recorded throughout the experiment using ErgoLAB platform (Figure 2). After each task, the subject was asked to fill up a PANAS survey. Since the investigation of the subject's behaviors was not the objective of this feasibility study, the evacuation task was purposely designed to be simple: the subject would find him/herself in a bedroom at the beginning of the experiment, with a health warning message about use of IVE displayed for a few seconds. After the message disappeared, he/she would be instructed to evacuate from the apartment, by leaving the bedroom, going through a short burning hallway, and exiting from a door across a living room. Figure 3 shows some of the subjects performing the assigned tasks (training, evacuation, filling up the survey) during the experiments. If a subject felt any sickness in the IVE, he/she could ask to terminate the experiment at any time.

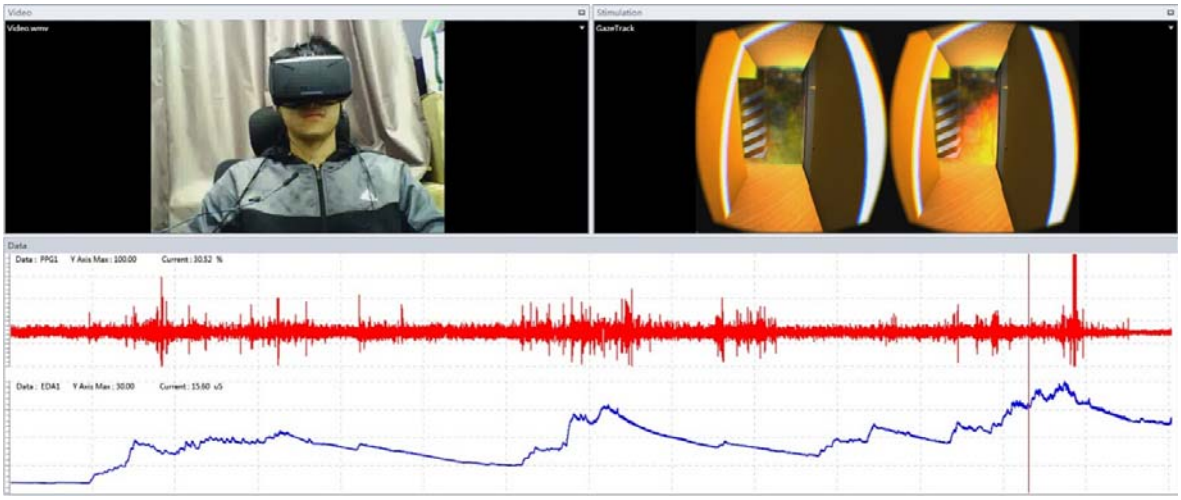


Figure 2. Interface of the experiment data collection platform

Figure 3. Subjects performing the assigned tasks

6 Findings and Discussions

Out of the 35 subjects, 33 subjects completed the experiments. Two subjects chose to quit due to motion sickness in the IVEs. One data set was incomplete due to technological problem of the physiological sensors. A total of 32 complete data sets were collected and used for analysis. The difference between subjects' respective emotional responses to IVE 1 and IVE 2 were compared, in order to examine the feasibility of using the proposed approach to assess the level of sense of presence. T-tests were conducted for negative affective scores and positive affective scores of PANAS, and SC mean scores. The findings are presented in this section.

6.1 Analysis of PANAS Scores

A pairwise comparison was conducted between three sets of PANAS scores, which were collected at the baseline (training) stage and after the two evacuation tasks. The results are summarized in Table 1. At a 95% confidence level, t-test results showed that the negative scores of IVE 2 was statistically larger compared with the negative scores of the baseline ($t = 1.796$, $p = 0.041 < 0.05$) and IVE 1 ($t = 5.065$, $p = 0.000 < 0.05$). It suggested that the higher level of realism in IVE 2 did cause subjects to experience sensibly higher negative emotions, and that such difference in emotional responses could be captured using the PANAS emotion scale. It needs to be pointed out, however, that there was noticeable decrease in the negative scores from the baseline stage to IVE 1 ($t = -2.204$, $p = 0.982 > 0.95$). It was possibly because that the subjects were not familiar with the feeling of being immersed in a virtual environment during the baseline period and therefore felt uncertain and dubious, which faded away to a larger extent than that of negative emotions elicited by IVE 1 when the subjects performed the evacuation task. For positive scores, t-test results suggested that there were no significant changes ($t = 0.909$ and $p = 0.185 > 0.05$ between baseline and IVE 1; $t = 1.155$ and $p = 0.128 > 0.05$ between baseline and IVE 2; $t = 0.740$ and $p = 0.232 > 0.05$ between IVE 1 and IVE 2), suggesting that the emotional responses to virtual fire emergency scenarios were mainly negative while positive emotions mostly remained the same. Furthermore, a detailed comparison between the average scores of each type of negative emotions collected during the baseline stage and IVE 2, as summarized in Table 2, revealed that scared (1.394 to 2.152), nervous (2.061 to 2.545) and afraid (1.485 to 2.000) emotions were the main negative emotions elicited by the fire emergency IVEs. These types of emotions were in accordance with the nature of fire emergency scenarios, which demonstrated the feasibility of using PANAS emotion scales to assess emotional valence and changes in their magnitude.

Table 1. Comparison between three sets of PANAS scores

	Baseline and IVE 1		Baseline and IVE 2		IVE 1 and IVE 2	
	Positive	Negative	Positive	Negative	Positive	Negative
t-score	0.909	-2.204	1.155	1.796	0.740	5.065
p-value	0.185	0.982	0.128	0.041	0.232	0.000

Table 2. Comparison between average scores of different types of negative emotions

	Upset	Distressed	Guilty	Scared	Hostile	Irritable	Ashamed	Nervous	Jittery	Afraid
Baseline	1.727	1.879	1.242	1.394	1.152	1.394	1.758	2.061	1.515	1.485
IVE 2	1.727	1.939	1.333	2.152	1.333	1.303	1.152	2.545	1.485	2.000
Δ	0	0.061	0.091	0.758	0.182	-0.091	-0.606	0.485	-0.030	0.515

6.2 Analysis of SC Scores

The SC scores were calculated by adding average SC tonic score and average SC phasic score. A typical SC curve of a subject throughout the entire experiment is shown in Figure 4. A pairwise comparison was conducted between the SC scores collected at the baseline stage and during the two evacuation tasks. The results are summarized in Table 3. At a 95% confidence level, t-test results showed that there was significant pairwise difference between the SC scores of the baseline, IVE 1 and IVE 2. Specifically, the SC scores of the subjects first increased by 120.84% on average from the baseline to IVE 1, and further increased by another 14.93% from IVE 1 to IVE 2. The comparison of the SC scores is shown in Table 4.

Moreover, it was observed that the SC score at the beginning of each evacuation task was slightly different than each other and both were higher than the SC score at the baseline stage. It showed that the subjects' emotion

did not return to the original state after the three-minute rest. To take into consideration such residual impact of the preceding training or evaluation task, the relative score of SC, namely the increase of SC score over the period of each evaluation task, was calculated to compare between the two IVEs. The relative SC scores of the two IVEs were denoted as ΔIVE 1 and ΔIVE 2, respectively, and summarized in Table 5. It can be seen the table that ΔIVE 2 (5.351) was 23.52% higher than ΔIVE 1 (4.332), which indicated that the experiment in IVE 2 resulted in more significant emotional response than the experiment in IVE 1.

In short, the analysis of SC scores showed, in accordance with the assessment results by the emotion scale, that the subjects experienced higher emotional responses and hence higher sense of presence in a more realistic IVE. Moreover, unlike the PANAS scores, the SC scores captured statistically noticeable difference in the emotional arousal between the baseline and IVE 1. The results indicated that the SC score was an effective indicator, with higher sensitivity than the emotion scale, for assessing the emotional arousal in IVEs.

Table 3. Comparison between three sets of SC scores

	Baseline and IVE 1	Baseline and IVE 2	IVE 1 and IVE 2
t-score	7.307	7.503	6.929
p-value	0.000	0.000	0.000

Table 4. Increase of SC scores in IVE 1 and IVE 2 compared to their preceding stages

	Baseline	IVE 1	IVE 2
SC mean (μS)	7.366	16.267	18.696
Percentage increase	/	120.84%	14.93%

Table 5. Comparison between two sets of relative SC scores

	Rest 1	IVE 1	Rest 2	IVE 2
SC mean (μS)	11.936	16.267	13.345	18.696
Relative SC scores (μS)	/	4.331	/	5.351

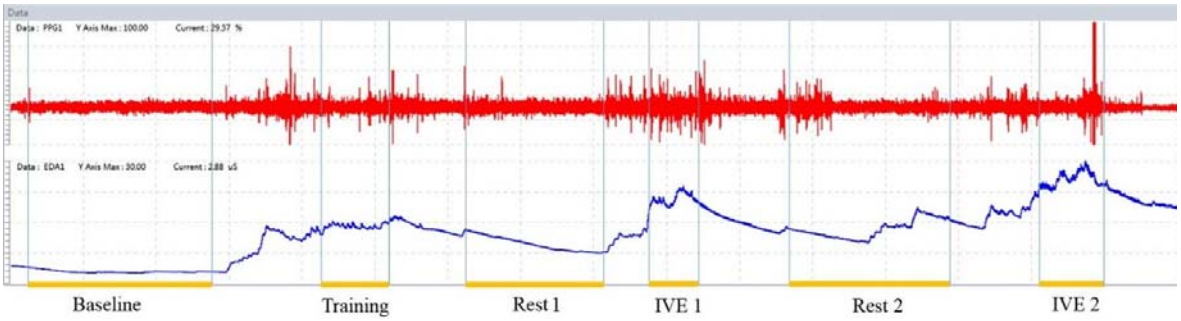


Figure 4. A typical SC curve in the experiments

7 Conclusions and Future Work

This study examines the feasibility of using a combination of subjective measure (an emotion scale) and objective measure (physiological indicators) for both emotional valance and emotional arousal to assess subjects' sense of presence, a critical factor that determines the overall level of ecological validity of evacuation experiments conducted in IVEs. Two IVEs were developed in this study, both representing a fire emergency scenario in an apartment but having different levels of realism. Experiment subjects were asked to perform an evacuation task in both IVEs, and their emotional responses were monitored throughout the experiments using the subjective and objective measures. The results showed that the PANAS emotion scale and the SC score were effective measures of emotional responses of the subjects to fire emergency IVEs. Assessment reported by both measures were generally in accordance with each other. The emotion scale was able to distinguish specific types of emotion responses elicited by the IVEs, especially scared, nervous and afraid, whereas the physiological indicators were more sensitive

to subtle changes in the magnitude of emotional responses. When used together, the subjective and objective measures provided reasonable assessment of the sense of presence that subjects experience in emergency evacuation IVEs.

To advance this line of research, the authors plan to examine the effectiveness of other types of physiological indicators in assessing the sense of presence in the IVEs, and establish the relationship between these indicator and the emotional responses based on the recognition of physiological signal patterns. Moreover, the long-term objective of this study is to build IVEs with high sense of presence so that they can be adopted in evacuation behavioral experiments with high ecological validity for investigating the sensing, decision making and behaviors of individual evacuees during emergencies. Therefore, further work is needed to determine the threshold of the sense of presence needed to ensure sufficient ecological validity of virtual evacuation experiments, and to explore new approaches and technologies to develop IVEs that can satisfy the threshold of the sense of presence.

Acknowledgements

This material is based upon work supported by Tsinghua University Initiative Scientific Research Program under Grant No. 2014z21050. The authors are thankful for the support of Tsinghua University. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of Tsinghua University. 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

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

[2] Mc Daid A. Study into the use of signage on exit choice behaviour in a virtual environment. *Applied Mechanics & Materials*, 796:1-11, 2015.

[3] Xi M. and Smith S.P. Simulating cooperative fire evacuation training in a virtual environment using gaming technology. In *Virtual Reality (VR)*, IEEE, 2014.

[4] Putorti Jr A. and McElroy J. Full-scale house fire experiment for InterFIRE VR. *National Institute of Standards and Technology Report of Test FR4009*, US Dept.of Commerce, 2000.

[5] Brewer M.B., Reis H. and Judd C. Research design and issues of validity. *Handbook of research methods in social and personality psychology*, 3-16, 2000.

[6] Anderson C.A. and Bushman B.J. External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology*, 1(1):19, 1997.

[7] Heydarian A., Carneiro J.P., Gerber D., Becerik-Gerber B., Hayes T. and Wood W. Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations. *Automation in Construction*, 54:116-26, 2015.

[8] Shechtman O., Classen S., Awadzi K. and Mann W. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. *Traffic injury prevention*, 10(4):379-85, 2009.

[9] Iachini T., Coello Y., Frassinetti F., Senese V.P., Galante F. and Ruggiero G. Peripersonal and Interpersonal Space in Virtual and Real Environments: Effects of Gender and Age. *Journal of Environmental Psychology*, 45:154–164, 2016.

[10] Slater M., Lotto B., Arnold M.M. and Sánchez-Vives M.V. How we experience immersive virtual environments: the concept of presence and its measurement. *Anuario de Psicología*, 40:193-210, 2009.

[11] Riva G., Mantovani F., Capideville C.S., Preziosa A., Morganti F. and Villani D. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1):45-56, 2007.

[12] Luigi M., Massimiliano M., Aniello P., Gennaro R. and Virginia P.R. On the Validity of Immersive Virtual Reality as Tool for Multisensory Evaluation of Urban Spaces. *Energy Procedia*, 78:471-6, 2015.

[13] Mardaga S., Laloyaux O. and Hansenne M. Personality traits modulate skin conductance response to emotional pictures: An investigation with Cloninger's model of personality. *Personality and individual differences*, 40(8):1603-14, 2006.

[14] Wilson G.M. and Sasse M.A. Do users always know what's good for them? Utilising physiological responses to assess media quality. In *People and Computers XIV—Usability or Else! Springer*, pages 327-339, 2000.

[15] Rodríguez A., Rey B., Clemente M., Wrzesien M. and Alcañiz M. Assessing brain activations associated with emotional regulation during virtual reality mood induction procedures. *Expert Systems with Applications*, 42(3):1699-709, 2015.

[16] Stephens C.L., Christie I.C. and Friedman B.H. Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis.

Biological Psychology, 84(3):463-473, 2010.

[17] Kreibig S.D. Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3):494-421, 2010.

[18] Kallinen K. Emotion related psychophysiological responses to listening music with eyesopen versus eyes-closed: electrodermal (EDA), electrocardiac (ECG), and electromyographic (EMG) measures. In *Proceedings of Music Perception & Cognition*, pages 299-301, 2004.

[19] Gomez P., Shafy S. and Danuser B. Respiration, metabolic balance, and attention in affective picture processing. *Biological Psychology*, 78(2):138-149, 2008.

[20] Appelhans B.M. and Luecken L.J. Heart rate variability as an index of regulated emotional responding. *Review of general psychology*. 10(3):229-240, 2006.

[21] Watson D., Clark L.A. and Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6): 1063-1070, 1988.

[22] Zuoji Z. Behavioral medicine inventory manual. *Chinese Journal of Behavioral Medicine Science*, 10(10):19-24, 2001.