Evaluation of the user experience of “astronaut training system”: An Immersive, VR-Based, Motion-Training System

Kang Yue1,a, Danli Wang1,2,b*, Xinpan Yang1, Haichen Hu1, Yuqing Liu3 and Xiuqing Zhu3
1Beijing Key Lab of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, Beijing, China
2State Key Laboratory of Computer, Institute of Software, Chinese Academy of Sciences, Beijing, China
3National Key Laboratory of Human Factors Engineering, China Astronaut Research and Training Center; Beijing, China
aeinhep@gmail.com, bdanli@iscas.ac.cn

ABSTRACT
To date, as the different application fields, most VR-based training systems have been different. Therefore, we should take the characteristics of application field into consideration and adopt different evaluation methods when evaluate the user experience of these training systems. In this paper, we propose a method to evaluate the user experience of virtual astronauts training system. Also, we design an experiment based on the proposed method. The proposed method takes learning performance as one of the evaluation dimensions, also combines with other evaluation dimensions such as: presence, immersion, pleasure, satisfaction and fatigue to evaluation user experience of the System. We collect subjective and objective data, the subjective data are mainly from questionnaire designed based on the evaluation dimensions and user interview conducted before and after the experiment. While the objective data are consisted of Electrocardiogram (ECG), reaction time, numbers of reaction error and the video data recorded during the experiment. For the analysis of data, we calculate the integrated score of each evaluation dimension by using factor analysis. In order to improve the credibility of the assessment, we use the ECG signal and reaction test data before and after experiment to validate the changes of fatigue during the experiment, and the typical behavioral features extracted from the experiment video to explain the result of subjective questionnaire. Experimental results show that the System has a better user experience and learning performance, but slight visual fatigue exists after experiment.

Keywords: virtual reality; user experience; astronaut training system; factor analysis; ECG

1. INTRODUCTION
In recent years, astronaut training methods based on VR technology have been improved rapidly. Traditional astronaut training methods contain classroom training, multimedia presentations, physical mock-up and swimming pool-based training and parabolic flights. Compared with these traditional methods, virtual astronaut training methods are more efficient, especially when dealing with the task that is difficult and expensive to reproduce. Until now, researchers have developed various virtual astronaut training system based on virtual reality technology to accomplish a better user experience (UX). But, the literature review shows little effort has been made to evaluate the UX of these virtual astronaut training systems (VATS).

UX is generally understood as inherently dynamic, given the ever-changing internal and emotional state of a person and
differences in the circumstances during and after an interaction with a system\textsuperscript{3, 4}. Since the term UX was first introduced by Donald Norman in the 1990s to describe overall aspects of a person’s experience with a system\textsuperscript{5, 6}, it has become a key concept of Human-Computer Interaction (HCI). For evaluating UX of an application, most studies are conducted using methods like questionnaires, interviews, heuristic evaluation and video analysis.

To evaluate the UX of a VR-based application, presence and immersion are always considered as the core of the evaluation. Previous research has suggested that presence in a virtual environment (VE) is important for several reasons\textsuperscript{8}. Therefore, an immersive virtual environment may be a useful system for training and skill acquisition, where to train or gain the skill in the real world may be too expensive or dangerous.

Presence is generally defined as a user’s subjective sensation of “being there” in a scene depicted by a medium\textsuperscript{9}. Specifically speaking, it is always thought as a state of consciousness, the psychological sense of being in the virtual environment\textsuperscript{10}. In current theoretical models, the sense of presence is seen as the outcome or a direct function of immersion. The more inclusive, extensive, surrounding, and vivid the virtual environment (VE) is, or the more similar the transformations in the VE are to those in the real world, the higher the presence\textsuperscript{11}. For the reasons mentioned above, presence and immersion are usually measured by subjective methods i.e. questionnaire, self-report or interview. Also, a few studies using device feature such as screen update rate\textsuperscript{12}, view angle or screen size\textsuperscript{13} to assess immersion tendencies. Except for presence and immersion, there are many other dimensions used for UX evaluation for a specific type of application such as emotion and task performance. As the fact that UX is a state of consciousness, emotion is a nature characteristics of UX. Especially when evaluating the UX of a VR-based game, emotions such as challenge, exciting, achievement are the main aspects of UX evaluation and even more important than other aspects such as pleasantness. For a training system, the performance of subjects may be an important indicator to measure how well the training effects are. Rajesh Aggarwal\textsuperscript{14} measure the performance by features as time taken, path length and error times during the task. Gerard Jounghyun Kim\textsuperscript{2}, Lin Yao et al.\textsuperscript{7} use task completion rate which is the proportion of participants who successfully complete the task as a measurement of task performance to evaluate the UX. There are many other indicators used in evaluating UX of an application, have used metrics called “In-game Navigation”\textsuperscript{15}, which calculated the distance covered in the virtual environment as well as average camera rotation speed and jump frequency. These metrics describe the activity of the user in the virtual scene, which imply the users’ feelings.

In this paper, we propose a reasonable method to evaluate the UX of the virtual astronaut training system. The proposed method evaluates the UX from different dimensions such as presence, immersion, emotion, fatigue and learning performance, and validates the evaluation result by analysis of objective data including ECG, reaction time and behavior features.

2. EXPERIMENT

2.1 Participants

A total of 33 subjects took part in the experiment. All participants were from collage who are majoring in computer science or human-computer interaction technology and take part in the experiment with receiving compensation. Twelve subjects were excluded due to technical problems with system hardware. Therefore, data of 21 subjects were analyzed, with an age distribution between 21 and 25 years, a mean (m) age of 23.5 years, and all subjects are male. The stereoscopic vision of the subjects are acceptable if the distinguishable disparity is equal to or smaller than 100".

2.2 VR-based astronauts training system (VATS)

A VATS, VR-based astronauts training system, is a 3D VR interactive learning system that is designed for astronaut candidates to experience the environment without gravity and obtain knowledge about the structure of the space station.
A fixed-based seat, head mounted display (HMD), posture tracking device and space mouse comprised the experimental setting. A space station with 5 capsules are designed to simulate the feeling of a real space environment. Posture tracking device is used for tasks such as grasping objects or space walking, and the space mouse is used for fast move.

2.3 Measurements
For acquisition and recording of participants' physiology the BIOPAC MP 150 system with AcqKnowledge data acquisition and analysis software are used. The channel measured is ECG and the acquisition frequency is 1000 Hz. We use reaction time recorder to test subjects' fatigue before and after experiment. The questionnaire we designed include 60 questions and is evaluated using 5-point Likert scale (ranging from 1 which means “strongly disagree” to 5 which means “strongly agree”). Eighteen questions which measure the changes of fatigue need to be filled before and after experiment, while others are filled after the experiment.

2.3 procedure
The experiment was conducted in a laboratory, with suitable temperature and humidity. The participants were in the natural condition when they interacted with the virtual environment (Figure 2). Before experiment starts, each subject is asked to report visual fatigue and physical fatigue in focusing on a questionnaire with 18 items. Then participants are instructed to complete reaction test and a training session was carried out to introduce the task he will execute. Meanwhile, equipment includes HMD, posture tracking and MP 150 are worn by subjects with the help of auxiliaries. The task includes a series of actions such as grasping, pressing, space walking and fast move using space mouse. The first three actions are designed to let the subjects experience act in a zero gravity environment, and we select data glove as the interactive method. The last action focuses on helping subjects learn the structure of space station and therefore we select space mouse interaction as convenient. The flow chart of whole experiment is shown in fig. 1.

Fig. 1. Flow chart of whole experiment
The task mentioned above needs to be performed twice. In the first pass, every action is trained for a fixed length of time and we call it “training phase”. The second pass is called “examine phase” in which subjects are asked to complete task as fast as possible without errors. The following with the reaction time test and questionnaire. ECG data is collected all though the experiment.
3. DATA ANALYSIS AND RESULT

3.1 The subjective data

As each dimension contains more than one question, the final score of the dimension should be calculated by a reasonable algorithm with these questions. In this paper, we select factor analysis based-on principal component analysis (PCA) to determine the main factors of each dimension (except fatigue) and compute their weight. Before analysis, the Kaiser-MeyerOlkin (KMO>0.6) and Bartlett’s test (sig. < 0.001) are calculated to determine whether the subjective data are suitable for factor analysis. Table 1 shows the KMO and Bartlett's test result of each dimension (except fatigue). Besides, we exclude fatigue because it is collected before and after experiment compared with other dimensions which are collected after experiment only and we use T-test to determine whether the fatigue increases significantly after experiment.

Table 1 KMO and Bartlett’s test results of each dimension (except fatigue)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Presence</th>
<th>Immersion</th>
<th>Emotion</th>
<th>Learning performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>KMO</td>
<td>0.736</td>
<td>0.61</td>
<td>0.67</td>
<td>0.62</td>
</tr>
<tr>
<td>Sig. of Bartlett's test</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 2 Main factors of each dimensions (except fatigue)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Presence</th>
<th>Immersion</th>
<th>Emotion</th>
<th>Learning performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>3.11</td>
<td>2.54</td>
<td>3.97</td>
<td>5.31</td>
</tr>
<tr>
<td>Factor 2</td>
<td>1.08</td>
<td>1.34</td>
<td>1.54</td>
<td>1.7</td>
</tr>
<tr>
<td>Factor 3</td>
<td>2.69</td>
<td>1.15</td>
<td>1.11</td>
<td>1.15</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>51.85%</td>
<td>42.32%</td>
<td>44.14%</td>
<td>44.22%</td>
</tr>
<tr>
<td>Variance</td>
<td>69.83%</td>
<td>64.6%</td>
<td>73.5%</td>
<td>67.93%</td>
</tr>
<tr>
<td>Cumulative explain</td>
<td>3.07</td>
<td>3.68</td>
<td>3.39</td>
<td>3.41</td>
</tr>
<tr>
<td>Factor score</td>
<td>4.07</td>
<td>3.75</td>
<td>3.32</td>
<td>3.52</td>
</tr>
<tr>
<td>Comprehensive score</td>
<td>3.24</td>
<td>3.7</td>
<td>3.28</td>
<td>3.43</td>
</tr>
</tbody>
</table>

Table 2 shows the main factors of each dimensions except fatigue. As mentioned above, we designed a presence questionnaire included 3 aspects such as realness, involvement and spatial presence, while the result of factor analysis gives only 2 aspects. The reason for this is that the number of subject may be not quite enough after analysis of data. And we notice that when we set the threshold of eigenvalue to 0.8, the factors calculated by SPSS match the previous research as well as the questions of each factor. Similarly, the results of immersion are inconsistent with our hypothesis motioned above and become the same when threshold is stetted to 0.8. Dimensions such as emotion and learning performance is quite consistent with our hypothesis and the results are shown in table 2.

We calculate comprehensive score using Component Score Coefficient produced by SPSS. The equation is shown as following.

\[ F_p = \sum_{j=1}^{n} \beta_{pj} x_j \]  

\[ cS = \sum_{p=1}^{n} Var_p \ast F_p \]  

\[ \sum_{p=1}^{n} Var_p \]
The $\beta_p$ is the weight value of $p$th factor and $i$th variable in Component Score Coefficient matrix and $x_i$ is the $i$th variable of a sample. The total sample number is $m$. In equation (2), $r_p$ is the $p$th factor of a dimension and $\text{Var}_p$ correspond to variance that $r_p$ can explain. The factor score and comprehensive score (CS) is shown as follow in table 2.

Figure 2 Subjective fatigue score

Figure 2 illustrates that most symptoms of fatigue increase significantly. None have exceeded 2.5 which is smaller than 3. This means the fatigue increase after watching the display but not serious.

3.2 HR from ECG

The ECG signals for each subject are processed with a 50 Hz notch filter and band-pass filter firstly. Then it is split into segments of 30s length and the segments are averaged. After averaging, the HR is extracted using AcqKnowledge analysis software. Because most tasks in the experiment have lots of body movements except the task of fast move, the HR extracted from ECG contain noise as well. We select ECG during task of fast move to compare HR changes between the first and the last 3 minutes to investigate the HR changes. As fast move need no body movement except fingers to control space mouse, we can minimal interference.

Figure 3 HR changes between first and last 2.5 minutes during fast move task

From figure 3, we can clearly see HR of last 2.5 minutes increases compared with the first 2.5 minutes which produces essentially agreement with the result of fatigue questionnaire.  

3.3 Reaction indicators

Actually, reaction consists of reaction time and numbers of reaction error which reflect the subjects’ fatigue. We investigate whether the time or numbers of error has significant changes based on data before and after experiment using T test.
Figure 5 shows that reaction time and errors times increase slightly after experiment ($p_{\text{reaction time}} = 0.14$, $p_{\text{error times}} = 0.27$). The results imply that fatigue has not increased though the experiment and is consist with the result of questionnaire and HR.

### 3.4 Features extracted from video

Behaviors of subjects during experiment can reflect the subjects’ feelings or UX of system directly. A smile on one’s face always means a good mood, touching the HMD implies the VR headset is not properly worn, sweating and nausea means there is a serious visual fatigue symptom has appeared. In this paper, we define behaviors such as smiling, touching the HMD, Sweating and nausea to reveal the UX of the system by analyzing the number of times that they have occurred during experiment.

Statistics show that almost 50% of subjects touched their HMD at least once. This means the HMD is uncomfortable for subjects and reduced the quality of immersion. About 76.19% of subjects smiled at least once during the experiment which means the emotion of subjects is positive as well as a good mood. 9.52% of subjects had sweating and nausea symptom. This means that although the visual fatigue has increased little in general, there still a serious visual fatigue occurred for few people.

### 4. CONCLUSION

According to the results motioned above, we can see that the system has a higher score in dimensions such as presence, immersion, emotion and learning performance (3.24, 3.7, 3.28 and 3.43) which means the UX of system is satisfactory. The subjective score of fatigue changes little implies subjects did not feel obvious symptoms of fatigue and results from analysis of HR and reaction time also support this conclusion. But few subjects (2 people) felt serious visual fatigue symptom such as sweating and nausea which means the system still cause damage to crown who are sensitive to visual fatigue.

### 5. ACKNOWLEDGEMENTS

We gratefully acknowledge financial support from the Major State Basic Research Development Program of China under Grant No.2013CB328805, the National Natural Science Foundation of China under Grant No. 61272325 and No. 60970090 and the Foundation Scientific Research Plan of National Defense (B1720132001).

### 6. REFERENCES


