2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation

An Event-based User Experience Evaluation Method for Virtual Reality Applications

Siqi Ma, Xiaoping Che, Qi Yu, Chenfeng Yue
School of Software Engineering, Beijing Jiaotong University, Beijing, China
17121712@bjtu.edu.cn, xpche@bjtu.edu.cn, 16121736@bjtu.edu.cn, hill753@163.com

***Abstract*—With the development and advancement of Virtual Reality (VR) technology in recent years, hardware is no longer the main factor limiting VR into the public life, and the demand for user experience (UX) continues to increase. The development of VR is beginning to reveal a content-oriented trend. Whether the content is attractive enough will greatly affects users' experience and determine the service life of VR applications. Consequently, searching for interaction events that would affect VR UX is critical to improve user stickiness. At present, the researches on VR mainly focus on the improvement of hardware and software, and pay less attention to the content event design. Moreover, there is no unified and clear standard for VR UX evaluation. This paper attempts to find out the relationship between user traits, VR interaction events and user experience through realistic experiments. This paper firstly define four types of VR interaction events, and design a questionnaire for collecting the tester's traits and their subjective evaluation. During the experiment, objective physiological data of the testers and their participation process were recorded. The 80 testers were divided into two groups to experience two types of VR games with constant time. The statistical method and improved Prism algorithm were used to find out the correlation among user traits, type of game interaction events and user experience. The experiment results could provide references for VR designers and developers, and at the same time provide the preliminary study to standardized VR user experience evaluation.**

***Index Terms*—Virtual Reality, User Experience, User Traits, VR Interaction, Prism Algorithm**

I. Introduction

According to market research in recent years, with the gradual maturity of virtual reality (VR) technology, the market scale has further expanded. Goldman Sachs predicts that the VR/AR market revenue will reach 80 billion US dollars in 2025. At the same time, the decreasing cost of VR products and the increase in usability have further promoted the popularity of VR. While VR hardware has made considerable strides, unique attention was given to applications and the associated decisions that they support [1]. Just as the high-quality game content that led to the consumption in games [2], the factor that can really stimulate users' consumption in the VR industry is high-quality content. With the content consumption habits of consumers gradually formed, the revenue of VR application is gradually increasing. However, the development of VR content still faces a major bottleneck. According to the statistics of Yingwei.com, until April 2018, the number of VR games on Chinese main VR content platforms Steam, Oculus PC, Oculus Mobile and PlayStation VR platform are about 2400, 1000, 1300 and 230, accordingly. It is still a small number compared with the large amount of mobile and console games.

With the recent release of several VR Helmet-Mounted Displays (HMD), there has been growing interest in the design and development of VR applications [3]. Although VR offers exciting possibilities for new content and interaction designs, designers are facing with tough challenges due to the immature VR application design. Designers can use sophisticated interaction patterns and conventions when designing mature environments such as desktop or mobile platform. For example, when using desktop software, the user is familiar with having a button at the corner of the window to close the game. While on the smartphone, the user is familiar with the hamburger menu and pinch zoom. Since 3D environments within VR often afford substantially different modes of interaction than any other software interfaces, the conventions of touch screen and desktop software cannot directly map to VR interactions. Due to this reason, the interactions in VR are essentially needed to be validated through users. Therefore, UX research is critical to the success of new VR software interaction design.

At present, UX researches of VR mainly focuses on the following aspects: immersion and motion sickness [4] [5] [6], VR for cultural transmission or as a teaching aid [7] [8] [9], VR for diseases (especially mental diseases) treatment [10] [11] [12], the locomotion in VR [3] [13] [14], currently only a single user can be allowed in most VR environments, which is easy to cause loneliness. There are also studies trying to find possible social ways in VR [15] [9]. However, existing researches on VR UX have not fully explored the correlation between users' own traits and interaction events.

The purpose of this paper is to find out the variant sets of the VR game that is positively related to the UX, which is composed of the user's own traits and the categories of interaction events in the game. It aims to accurately recommend the appropriate VR content to the target user group and improve the interaction in the game according to the preferences of targeting users. We selected 'Slingshot' and 'Longbow' in 'The Lab' developed by game company VAVLE as the test VR applications. 'The Lab' is the top three VR game on Steam list since its launch in 2016, and gains a lot of praise from users. The user's own traits and subjective evaluation used in experiments are obtained through questionnaires. And four typical types of interaction events (Touch, Grab, Operation, Move) have been defined. During the experiment, tester's participation during the whole test process are recorded. In

978-1-7281-4034-6/19/$31.00 ©2019 IEEE

DOI 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00164

order to get more accurate factor that affect the UX, we improve the Prism rule induction algorithm.

The main contributions of this paper are:

- A data acquisition method is designed to collect users' traits and their subjective and objective experience.
- An improved Prism algorithm is proposed for dealing with the compound data samples.
- The deductive correlations among users' traits, interaction events and VR UX are concluded and represented by the form of rule sets.

The second part of this paper presents a brief overview of the existing research on VR UX and methods for evaluating UX. The third part shows the design of questionnaire and definition of interaction events. The fourth part describes the processing of the questionnaire data, heart rate data and the improvement of the Prism algorithm in detail. The fifth part introduces the experimental environment and analysis of the experimental results. Finally, the last part summarizes all the work and looks into the future.

II. Related Work

As mentioned in Section I, UX research is critical to the success of new VR software interaction design. However, existing researches on VR UX have not fully explored which users' own traits and categories of interaction events in games are positively correlated with UX. At the same time, a precise evaluation of UX combined both subjective methods and objective methods is still a challenging task.

A. UX Reasearch of VR

Research on VR has studied users' experience of exiting, instruction, transition, sound and locomotion in VR. [5] explore the momentary experience of exiting VR and transitioning back to the real-world through explication interviews and present six designs for easing or heightening the exit experience. [16] examined effects of instruction methods on UX in VR serious games through a user study with 15 college aged users. It tried to find the type of instructions which provided better UX. [17] studying the impact of transitions on UX of presence in VR using a questionnaire-based user experiment. [6] used comparative experiment and found that the moderate and strategic use of onomatopoeia can indeed help direct user attention, offer object affordance and thereby enhance UX and even the sense of presence and immersion. [3] present a study examining the effects of four frequently used first person controller-based locomotion techniques on player experience in a VR game. Results show that free teleport locomotion provides the best player experience while simultaneously eliciting the lowest discomfort of the compared techniques.

However, the existing VR UX research of interaction in VR is not enough. According to [18], Affordances which defines what actions are possible and how something can be interacted with by a user is one of the key terms relevant to interaction design in VR. [19] identifying the affordance features and find the role of affordance to enhance UX in VR games. [20] examines the VR hand interaction experience of the third-person compared to the first person, and then uses the questionnaire to verify whether this interaction provides a better experience for users. It defines the interaction of third-person virtual reality TPVR from three types: moving inside the virtual environment, selecting virtual objects and using a menu interface that includes a variety of functions and behaviors. In [20], the UX of the three kinds of interactions was studied, but only the questionnaire was used and no objective method was used for verification.

B. Subjective and Objective Research Methods of UX

Existing methods for assessing VR UX include subjective assessment methods, objective assessment methods and a combination of the two methods.

The subjective assessment method of UX is mainly self-assessment of users, and the direct experience of various factors of the product can be assessed through oral assessment or through questionnaires [21]. The questionnaire can effectively assess the subjective experience of users on different scales of products [22]. [23] designed a survey instrument Virtual Experience Test (VET) which used questions scored from 1 (strongly disagree) to 5 (strongly agree) to measure holistic virtual environment experiences, but the experiment was studied in only one game environment. [24] revised the existing simulator sickness questionnaire to develop a VR sickness questionnaire, which is used as the measurement index in a VR environment. [7] evaluate usability and UX in a three-dimensional VR environment prototype of a beach place. It evaluated the testers interaction in virtual world through their behavior, and responses to questionnaires of presence and simulator sickness.

The objectively assessment method of UX is mainly to evaluate the user's physiological signals, such as brain waves, heart rate, eye movements, etc. [25]. The heartbeat activity expresses the changes of user emotions. Monitoring heart rate is considered to be an effective way of measuring users emotion state [26]. They developed a new method to test VR game UX: Electroencephalography (EEG) signals and the brain functional connectivity (FC). The results showed the FC has a significant difference for two VR games of different usability, and a significant difference were observed in Gamma band, which indicated EEG can be a good tool to analyze UX of VR game.

A large number of studies focus on the combination of subjective and objective assessment methods. [27] focus on the questionnaire method because most of components in UX of immersive virtual environments (IVE) can be measured through questionnaires. However, they still believe that the best way to measure the UX in IVEs is to gather and compare results from the appropriate subjective methods with the appropriate objective methods. As an example of the type of study implemented in this paper, testers were asked to answer the Igroup Presence Questionnaire (IPQ) after using the simulation. This allows objective visual performance data to be compared to subjective performance assessment tools.

Combining the questionnaire with the heart rate can be a more comprehensive and convincing assessment of UX [28].

According to literature research, VR experience is related to user traits and interaction behavior, but existing research is insufficient. Therefore, this paper chooses to study the relationship between user traits and interaction events and UX in VR. At the same time, in view of the fact that the existing research does not use objective methods to verify the final conclusion. The electrocardiograph can be a good tool to objectively verify the UX.

III. PRELIMINARY

In order to explore the correlation between UX, traits of the VR user and different types of interaction events in the game, this paper designed a questionnaire containing the user's own traits, basic physiological data and scores. The interaction events in the game are defined and counted with the successes or failures actions. This chapter describes the design of questionnaires and the definition of interaction events.

A. Questionnaire Design

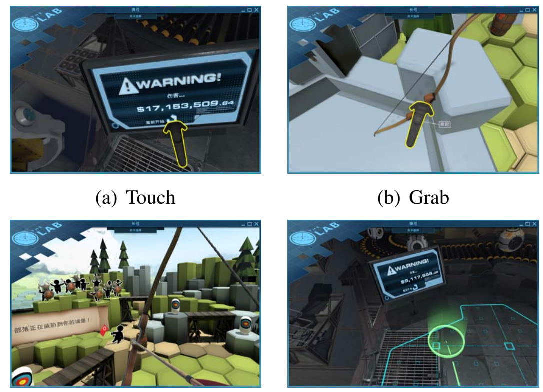
Studies have shown that gender differences may affect the VR experience [29]. [4] imply that UX in VR depend on individual traits. Therefore, this article considers the UX related to the user's own traits. The questionnaire includes gender, age, personality traits, exercise frequency, game experience, and basic heart rate and willingness to continue playing the VR game. Among them, personality traits are divided into extroversion and introversion. The game experience are divided into rich and poor. The exercise frequency are divided into frequent and occasional. The basal heart rate is used as a heart rate reference value to process heart rate data. In order to evaluate the user's subjective experience of VR, the questionnaire uses the user's score on the game process. Since users' preferences for game may vary from stage to stage, the score is divided into three stages according to the game time: the first, middle and last stage. According to the 5-point Likert scale [30], the score range is 1-5, 0 represents 'dislike', and 5 represents 'like'. Table I shows the questionnaire containing the users' traits and the user's subjective experience.

TABLE I
USER EXPERIENCE QUESTIONNAIRE

ITEMS	VALUE1	VALUE2
Age	16-50	
Gender	male	female
Game Experience	rich	poor
Exercise Experience	rich	poor
Character Traits	extroversion	introversion
Basic heart rate	value	
Willing to Continue Playing	yes	no
Score of the first stage	1-5	
Score of the middle stage	1-5	
Score of the last stage	1-5	

B. Interaction Event Definition

According to [20], we speculate that the probability of success and failure of VR interaction event would affect the UX of users. In this paper, according to the interaction with VR environment, we define four types of interaction events that are shown in Fig.1:



(a) Touch (b) Grab

(c) Operation (d) Move

Fig. 1. Four types of interaction events

- Touch: touching elements of the VR environment, including touching the menu and items.
- Grab: grabbing elements of the VR environment, including grabbing bows, arrows and items.
- Operation: performing a specific action to achieve the goal. Such as hitting a target with a projectile or an arrow.
- Move: transporting the user across a distance.

IV. METHODS

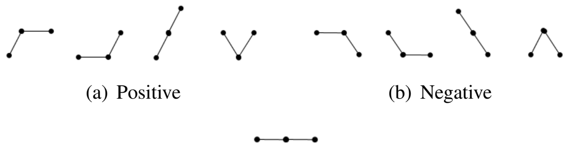
This section deals with the questionnaire data, interaction events and heart rate data mentioned above. Then introduce the Prism algorithm and the improvement.

A. Data Processing

According to [4], user traits may affect their experience. This paper selects the gender, game experience (including mobile games, console games), sports frequency and personality traits in the questionnaire as the features. Age was not considered as a feature because testers' age was concentrated between 16 and 25. The value of the features are digitized into different sets separately.

According to the three stages rating system we defined, different score trends would appear. In our approach, rising trends are regarded as positive UX, declining trends are regarded as negative UX, and steady trend is regarded as neutral UX. For example, if the scores of three stages are 3, 4 and 5, then it is a positive sample. The relationship between score trends and definition are concluded in Fig.2:

In order to accurately obtain the time of each interaction event and the heart rate change, we have synchronized their occurrence time. During the experiment, professional instruments were used to measure the user's real-time heart rate. The types of interaction events and the success or failure of the events were marked at the corresponding heart rate at the time of the events. And the success probability of each type of interaction events for each user was also calculated by the



(a) Positive (b) Negative

(c) Neutral

Fig. 2. UX definition based on score trends

ratio of the success number of interaction event and the total number of interaction events. The success rates of the four types of interaction events mentioned above were converted into integer, as shown in the Table II.

TABLE II
INTERACTION EVENT FEATURE VALUE

Category of Interaction event	Success Rate	Numerical Value
Touch	high	1
	low	0
Grab	high	1
	low	0
Operation	high	1
	low	0
Move	high	1
	low	0

Finally, the users' traits and the success rate of four types of interaction events were taken as the features, and the user's rating trends were taken as the positive category and negative category, then the rule set related to UX was obtained through Prism algorithm.

In order to verify these rule sets objectively, ECG signals of users will be captured during the experiment. Since the baseline of electrocardiogram of each user was different [26], the baseline of each user was adjusted before the experiment. Continuous RR intervals during the game were measured by a pulse oximeter and were used to obtain accurate heart rate. Based on the heart rate value, calculate the heart rate change value before and after the interaction event, which reflects the user's physiological experience when the interaction event occurs, as shown in Fig.3.

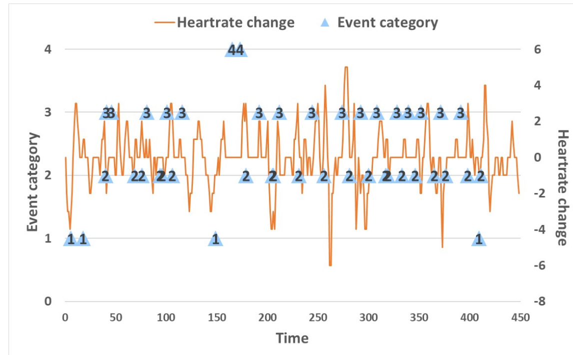


Fig. 3. Interaction Event and Real-time heart rate Change (On the left interaction event number 1 corresponds to touch events, 2 corresponds to grab events, 3 corresponds to Operation events, 4 corresponds to move events.)

B. Algorithm

The purpose of this paper is to find the relationship between user traits, VR interaction events and their UX. Current Rule induction classification algorithms can be divided into two categories according to different types of knowledge representation: decision tree algorithm and modular rule induction algorithm.

However, the algorithms based on decision tree have some disadvantages. The output of the decision tree is difficult to understand, and even after pruning it is difficult to form an intuitive understanding of the problem to be solved. Besides, decision trees are prone to overfitting. Therefore, Prism algorithm that can summarize relevant rule sets from data sets is selected. Although it is based on ID3, it uses different induction strategies to guide modular rules, thus avoiding many problems related to decision trees.

The algorithm defines the training sample set D as the cartesian product of attribute space X and category range space $dom(y)$, i.e,

$$D = X \times dom(y) \tag{1}$$

When the property is a discrete value,

$$X = dom(a_1) \times dom(a_2) \times \cdots \times dom(a_n) \tag{2}$$

$dom(a_i)(i = 1, 2, \cdots, n)$ represents the range space of the ith attribute.

On the other hand, all the samples in the training set can be represented as m tuples,

$$D = (< x_1, y_1 >, \cdots, < x_m, y_m >) \tag{3}$$

$x_j(j = 1, 2, \cdots, m)$ is the attribute vector, $y_i(i = 1, 2, \cdots, m)$ is the value in the category range space.

Given the training set described above, the rule form concluded by the rule learning system is

$$\text{IF } f_1 \wedge f_2 \wedge \cdots \wedge f_l \text{ THEN } Class = y_j \tag{4}$$

Where, the antecedent of the rule is the logical conjunction of a series of features, features $f_k(k = 1, 2, \cdots, l)$ is used to test whether the attribute value of the instance to be classified meets this condition. The conclusion of the rule gives the result of classification. The Prism algorithm considers the two-dimensional data table as a discrete decision system, in which the attribute value is taken as a message in the information theory.

The task of the induction algorithm is to find attribute-value pairs, such as $a_i = v$(v is a value of the attribute), so that it contributes the most information to a specific category y_j, that is, to make $I(Class = y_j | a_i = v)$ the largest, where

$$I = \log_2\left(\frac{p(Class = y_j | a_i = v)}{p(Class = y_j)}\right) \tag{5}$$

$p(Class = y_j)$ for all the attribute-value matching are the same. Therefore, finding the attribute-value pair $a_i = v$ that

Algorithm 1 Learning classification rules from labelled data instances

1: **for** $i = 1 \rightarrow dom(y)$ **do**
2: $D \leftarrow$ Dataset;
3: **while** D *does not contain only instances of* y_j **do**
4: **for all** $attribute\ a_i \in D$ **do**
5: **if** $attributes\ a_i\ is\ categorical$ **then**
6: Calculate the conditional probability, $p(Class = y_j|a_i = v)$ for all possible attribute-value $(a_i = v)$ from attribute a;
7: **else if** $atrribute\ a_i\ is\ numerical$ **then**
8: sort D according to v values;
9: **for** $v\ value\ of\ a_i$ **do**
10: calculate $p(Class = y_j|a_i \leq v)$ and $p(Class = y_j|a_i > v)$
11: **end for**
12: **end if**
13: **end for**
14: Select the $(a_i = v)$, $(a_i > v)$, or$(a_i \leq v)$ with the maximum conditional probability as a rule term;
15: $D \leftarrow S$, create a subset S from D containing all the instances covered by selected rule term at line 14;
16: **end while**
17: The induced rule R is a conjunction of all selected $(a_i = v)$, $(a_i > v)$, or$(a_i \leq v)$ at line 14;
18: Remove all instances covered by rule R from original Dataset;
19: **repeat**
20: lines 2 to 18;
21: **until** $all\ instances\ of\ class\ y_j\ have\ been\ removed$;
22: Reset input Dataset to its initial state;
23: **end for**
24: **return** induced rules;

maximizes $p(Class = y_j|a_i = v)$. And the details of the algorithm is shown in Algorithm 1.

Before the algorithm, due to the existence of continuous values and missing values in the questionnaire and heart beat data collected, this paper also improved the part of read function. Replace the missing feature value with the mode in the same feature and converts all feature value into discrete integers.

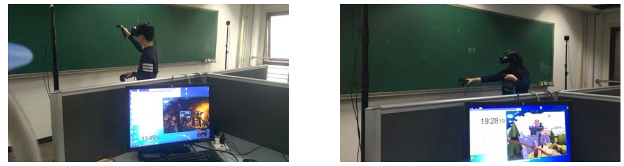
V. Experiment

This section first describes the experiment environment and experiment results, then uses the heart rate to verify the results, and compares the difference between two games.

A. Environment

80 participants without VR experience are selected for the experiment. They were split evenly into two groups to experience the two VR games 'Slingshot' and 'Longbow' in 'The Lab', as shown in Fig.4.

Before the start of game, testers needs to fill out a questionnaire which includes all the information defined in Table I.



(a) Slingshot (b) Longbow

Fig. 4. Experiment Environment

During the experiment, the user always wore a pulse oximeter, which used to measure and record the real-time heart rate. A video recorder was used to record the process of each user's game experience. The experience time of each tester was no less than 8 minutes, during which there was no external guidance or interference, ensuring the objectivity of test results. After the game, users filled in the last two columns of the questionnaire (scored the first, middle and last stages of the game).

B. Results

Due to the instability of video collection and heartbeat collection equipment (for example, heartbeat collection equipment may be disconnected due to the intense activity of testers, thus heartbeat collection will be invalid), invalid data will be deleted. In order to accurately calculate the success rate of each type of interaction events and the corresponding heart rate change values, we manually marked the four types of interaction events of each tester and the success/failure of each event, following the form of Fig.3.

According to the data processing method above, the success rate of four types of interaction events and the average heart rate change of every kind of events was calculated. After data preprocessing, the improved Prism algorithm was used to learn the relationship between user traits, interaction event categories and UX (positive or negative). Due to the small number of samples in the experiment, in order to avoid the occurrence of overfitting, this paper used the method of random sampling to find better rule sets.

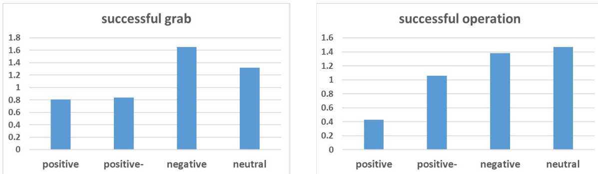
This paper select the following rule set of the two games for analysis and validation.

1) Slingshot:

- The rule sets that are positively related to the UX are listed below.
  a) $positive \leftarrow (gameex, 1) \cap (gender, 2)$
  b) $positive \leftarrow (sportex, 1) \cap (grab, 1)$
  c) $positive \leftarrow (personality, 1) \cap (operation, 1)$

It can be seen from rule a) that women with rich game experience have positive UX of 'Slingshot'. Since this rule does not involve the interaction events, verification of this rule by heart rate was not discussed.

In the figures related to positive rules, 'positive' means positive samples covered by the rule, 'positive-' means positive samples not covered by the rule, 'negative' means negative samples not covered by the rule, 'neutral' means neutral samples not covered by the rule.



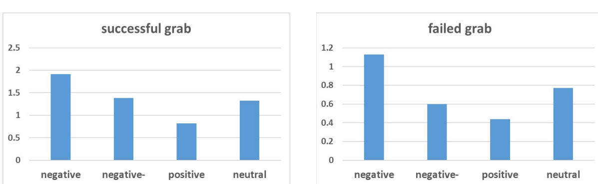
(a) Average heart rate change of different group in successful grab events (b) Average heart rate change of different group in successful operation events

Fig. 5. Slingshot Positive UX Results

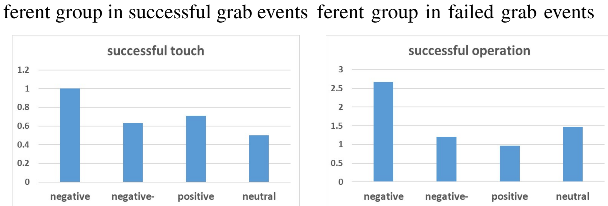
Rule b) shows that people who have rich exercise experience with high success rate of grab events are likely to have positive UX. According to Fig.5(a), the average heart rate change of 'positive' was significantly lower than that of 'negative' and 'neutral' and slightly lower than that of 'positive-'. Since a steady heart rate status generally corresponds to positive UX [31], it partially proves our conclusion. Fulfilling grab events design will increase the UX of people who have rich exercise experience.

Rule c) shows that extroverted users with high success rate of operation events are likely to have positive UX. According to Fig.5(b), the average heart rate change of 'positive' was significantly lower than that of 'positive-', 'negative' and 'neutral'. According to the same theory above, the physiological signal changes prove our conclusion. Achievable operation events design will have positive impact on extroverted users.

- The rule sets that are negatively related to the UX are listed below.
  a) $negative \leftarrow (sportex, 0) \cap (grab, 1)$
  b) $negative \leftarrow (grab, 0) \cap (sportex, 1) \cap (touch, 1)$
  c) $negative \leftarrow (gameex, 1) \cap (operation, 1)$



(a) Average heart rate change of different group in successful grab events (b) Average heart rate change of different group in failed grab events



(c) Average heart rate change of different group in successful touch events (d) Average heart rate change of different group in successful operation events

Fig. 6. Slingshot Negative UX Results

In the figures related to negative rules, 'negative' means negative samples covered by the rule, 'negative-' means negative samples not covered by the rule, 'positive' means positive samples not covered by the rule, 'neutral' means neutral samples not covered by the rule.

Rule a) shows that people who have little sport exercise with high success rate of grab events are likely to have negative UX. According to Fig.6(a), the average heart rate change of 'negative' was significantly higher than that of 'negative-', 'positive' and 'neutral'. Since high volatility in heart rate signal is generally related to negative UX [31], the results also prove our deduction. Inactive users would quickly lose their interest with undemanding grab events.

Rule b) shows that people who regularly have sport exercise with low success rate of grab events and high success rate of touch events are likely to have negative UX. According to Fig.6(b) and Fig.6(c), the average heart rate change in both grab and touch events of 'negative' was significantly higher than that of 'negative-', 'positive' and 'neutral', therefore the results still prove our conclusion. For active users, achievable grab events and less touch events will satisfy their needs.

Rule c) shows that people who have rich game experience with high success rate of operation events are likely to have negative UX. According to Fig.6(d), the average heart rate change of 'negative' was significantly higher than that of 'negative-', 'positive' and 'neutral'. Similarly, it proves the accuracy of the rule set. For experienced users, only challenging operation events can attract them.

2) Longbow:

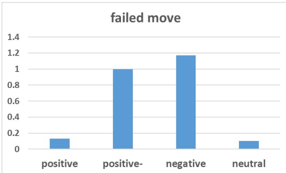
- The rule sets that are positively related to the UX are listed below.
  a) $positive \leftarrow (move, 0) \cap (personality, 1)$
  b) $positive \leftarrow (personality, 0) \cap (gender, 2) \cap (touch, 1)$
  c) $positive \leftarrow (personality, 0) \cap (gameex, 1) \cap (grab, 1)$

Rule a) shows that extroverted people with low success rate of move events are likely to have positive UX. According to Fig.7(a), the average heart rate change of 'positive' was significantly lower than that of 'positive-' and 'negative' and slightly higher than that of 'neutral'. Since we didn't take neutral UX into account, the high difficulty move events will attract extroverted users.

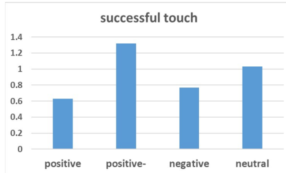
Rule b) shows that introverted female with high success rate of touch events are likely to have positive UX. According to Fig.7(b), the average heart rate change of 'positive' was significantly lower than that of 'positive-' and 'neutral' and slightly lower than that of 'negative', therefore the introverted female are easily satisfied with low difficulty touch events.

Rule c) shows that introverted people with rich game experience and high success rate of grab events are likely to have positive UX. According to Fig.7(c), the average heart rate change of 'positive' was significantly lower than that of 'positive-' and 'neutral' and slightly lower than that of 'negative', therefore the introverted experienced users are more used to low difficulty grab events.

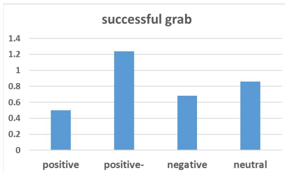
- The rule sets that are negatively related to the UX are listed below.
  a) $negative \leftarrow (gameex, 0) \cap (personality, 1)$
  b) $negative \leftarrow (gameex, 0) \cap (gender, 1)$



(a) Average heart rate change of different group in failed move events



(b) Average heart rate change of different group in successful touch events



(c) Average heart rate change of different group in successful grab events

Fig. 7. Longbow Positive UX Results

It can be seen from rule a) that extroverted people with little game experience have negative UX of 'Longbow'. And rule b) shows that male with little game experience have negative UX of this game as well. Since these two rules does not involve interaction events, verification of the rules by ECG data is not discussed here.

C. Discussion

According to the experiment results, different to the common sense. The configuration of fundamental interaction events 'Grab', 'Touch' and 'Move' has significant impact on VR UX. Fuzzy and vague design of these events would lead to unsuccessful actions, thus has negative impact on VR UX. The usually concerned goal-based interaction event 'Operation' has secondly influence on VR UX. Meanwhile, it should be designed with dynamic difficulty adjustment for different user group, in order to increase the average positive impact of VR UX.

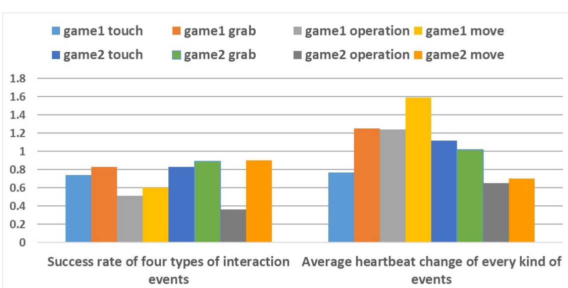
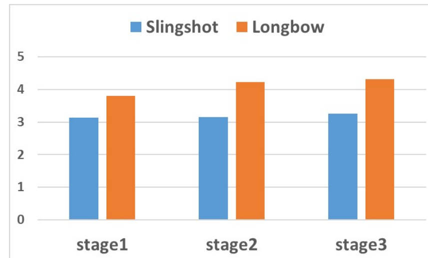


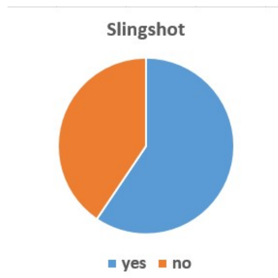
Fig. 8. Success rate and average heart rate change of interaction events

Additionally, the following aspects can be used to prove the arguments above. In the design part, 'Longbow' has better configuration of 'Grab', 'Touch' and 'Move' events than 'Slingshot'. As can be seen from the Fig.8, the action success rate of 'Grab', 'Touch' and 'Move' of 'Longbow' are higher than 'Slingshot'. And the total average heart rate change of interaction events in 'Longbow' is smaller than that in 'Slingshot', which confirmed the positive impact of UX.

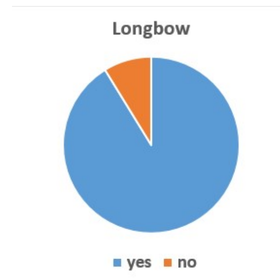
On the other hand, the scores of the first, middle and last stages of all testers for the two games are shown in Fig.9(a). As the figure shows, the scores of the three stages of 'Longbow' are all higher than 'Slingshot which indicates that 'Longbow' bring users better UX than 'Slingshot.



(a) Average score of different stages in two games



(b) Willingness to continue playing Slingshot



(c) Willingness to continue playing Longbow

Fig. 9. Comparison of the Two Games

Last but not least, as shown in Fig.9(b) and Fig.9(c), in the survey of 'willingness to continue playing the games', even with high difficulty of 'Operation' event, the percentage of testers that would like to continue playing 'Longbow' is still significantly higher than 'Slingshot'. This result also proves the importance of VR events design in a different perspective.

VI. Conclusion

This paper designed a questionnaire containing user traits and user subjective evaluation. During the experiment, the VR experience data of 80 testers were recorded, including questionnaire information, real-time heart rate and game record. This paper summarizes and classifies the interaction events in two VR games. Four types of interaction events were defined and correlated with the heart rate change. The mortality of different types of events of each tester is accurately calculated. Due to the existence of continuous values and missing values in the data collected, this paper improved part of Prism algorithm, and converts all data into discrete integer values after replacing the missing feature values. Finally, the user traits in the questionnaire and the success rate of interaction events were used as the characteristics, and the user rating trends were used as the consequence categories to obtain the rule sets related to UX. In order to verified the accuracy of rule set, the mean change of heart rate of users covered and not covered by rules when experiencing interaction events were calculated separately. The experiment result showed that the rule sets related to UX can be verified by heart rate change, indicating that the rule sets have reference value for the VR designers and developers.

ACKNOWLEDGEMENT

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2018JBM073 and in part by the National Natural Science Foundation of China under Grant 61502028.

REFERENCES

[1] L. P. Berg and J. M. Vance, "Industry use of virtual reality in product design and manufacturing: a survey," *Virtual Reality*, vol. 21, no. 1, pp. 1–17, 2017.

[2] J. Hamari, K. Alha, S. Järvelä, J. M. Kivikangas, J. Koivisto, and J. Paavilainen, "Why do players buy in-game content? an empirical study on concrete purchase motivations," *Computers in Human Behavior*, vol. 68, pp. 538–546, 2017.

[3] J. Frommel, S. Sonntag, and M. Weber, "Effects of controller-based locomotion on player experience in a virtual reality exploration game," in *International Conference on the Foundations of Digital Games*, 2017.

[4] D. Shin, "Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience?" *Computers in Human Behavior*, vol. 78, pp. 64–73, 2018.

[5] J. Knibbe, J. Schjerlund, M. Petraeus, and K. Hornbæk, "The dream is collapsing: The experience of exiting VR," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*, 2018, p. 483.

[6] J. Oh and G. J. Kim, "Effect of accompanying onomatopoeia with sound feedback toward presence and user experience in virtual reality," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST 2018, Tokyo, Japan, 11.28 - 12.01, 2018*, 2018, pp. 81:1–81:2.

[7] Y. Trindade, F. Rebelo, and P. Noriega, "Tourism and virtual reality: User experience evaluation of a virtual environment prototype," in *Design, User Experience, and Usability: Users, Contexts and Case Studies - 7th International Conference, DUXU 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part III*, 2018, pp. 730–742.

[8] C. Kelling, O. Kauhanen, H. Väätäjä, J. Karhu, M. Turunen, and V. Lindqvist, "Implications of audio and narration in the user experience design of virtual reality," in *Proceedings of the 22nd International Academic Mindtrek Conference, Mindtrek 2018, Tampere, Finland, October 10-11, 2018*, 2018, pp. 258–261.

[9] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. E. Latoschik, "Social augmentations in multi-user virtual reality: A virtual museum experience," in *IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2017 Adjunct, Nantes, France, October 9-13, 2017*, 2017, pp. 42–43.

[10] B. Senno and P. Barcha, "Customizing user experience with adaptive virtual reality," in *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion, Tokyo, Japan, March 07-11, 2018*, 2018, pp. 42:1–42:2.

[11] L. Chittaro, R. Sioni, C. Crescentini, and F. Fabbro, "Mortality salience in virtual reality experiences and its effects on users' attitudes towards risk," *Int. J. Hum.-Comput. Stud.*, vol. 101, pp. 10–22, 2017.

[12] T. Andersen, G. Anisimovaite, A. Christiansen, M. Hussein, C. Lund, T. Nielsen, E. Rafferty, N. C. Nilsson, R. Nordahl, and S. Serafin, "A preliminary study of users' experiences of meditation in virtual reality," in *2017 IEEE Virtual Reality, VR 2017, Los Angeles, CA, USA, March 18-22, 2017*, 2017, pp. 343–344.

[13] L. E. Warren and D. A. Bowman, "User experience with semi-natural locomotion techniques in virtual reality: the case of the virtuix omni," in *Proceedings of the 5th Symposium on Spatial User Interaction, SUI 2017, Brighton, United Kingdom, October 16 - 17, 2017*, 2017, p. 163.

[14] D. Calandra, M. Billi, F. Lamberti, A. Sanna, and R. Borchiellini, "Arm swinging vs treadmill: A comparison between two techniques for locomotion in virtual reality," in *Eurographics 2018 - Short Papers, Delft, The Netherlands, April 16-20, 2018.*, 2018, pp. 53–56.

[15] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio, "Sharevr: Enabling co-located experiences for virtual reality between HMD and non-hmd users," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017.*, 2017, pp. 4021–4033.

[16] L. Bozgeyikli, A. Raij, S. Katkoori, and R. Alqasemi, "Effects of instruction methods on user experience in virtual reality serious games," in *Virtual, Augmented and Mixed Reality - 9th International Conference, VAMR 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings*, 2017, pp. 215–226.

[17] M. Liang, N. Bryan-Kinns, A. S. Hassard, and Z. Ma, "The impact of transitions on user experience in virtual reality," in *Virtual Reality*, 2017.

[18] J. Jerald, *The VR Book: Human-Centered Design for Virtual Reality*, 2015.

[19] J. Lee, J. Bang, and H. Suh, "Identifying affordance features in virtual reality: How do virtual reality games reinforce user experience?" in *International Conference on Augmented Cognition*, 2018.

[20] M. Kim, J. Lee, C. Kim, and J. Kim, "Tpvr: User interaction of third person virtual reality for new presence and experience," *Symmetry*, vol. 10, no. 4, p. 109, 2018.

[21] J. Harms, D. Seitz, C. Wimmer, K. Kappel, and T. Grechenig, "Low-cost gamification of online surveys: Improving the user experience through achievement badges," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY 2015, London, United Kingdom, October 5-7, 2015*, 2015, pp. 109–113.

[22] M. Schrepp, M. P. Cota, R. Gonçalves, A. Hinderks, and J. Thomaschewski, "Adaption of user experience questionnaires for different user groups," *Universal Access in the Information Society*, vol. 16, no. 3, pp. 629–640, 2017.

[23] D. B. Chertoff, B. Goldiez, and L. V. Jr, "Virtual experience test: A virtual environment evaluation questionnaire," in *Virtual Reality Conference*, 2013.

[24] H. K. Kim, J. Park, Y. Choi, and M. Choe, "Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment," *Applied Ergonomics*, vol. 69, pp. 66–73, 2018.

[25] J. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 2005.

[26] L. Yao, Y. Liu, W. Li, L. Zhou, Y. Ge, J. Chai, and X. Sun, "Using physiological measures to evaluate user experience of mobile applications," in *Engineering Psychology and Cognitive Ergonomics - 11th International Conference, EPCE 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings*, 2014, pp. 301–310.

[27] K. Tcha-Tokey, E. Loup-Escande, O. Christmann, and S. Richir, "A questionnaire to measure the user experience in immersive virtual environments," in *International Virtual Reality Conference*, 2016.

[28] Q. Yu, X. Che, S. Ma, S. Pan, Y. Yang, W. Xing, and X. Wang, "A hybrid user experience evaluation method for mobile games," *IEEE Access*, vol. 6, pp. 49 067–49 079, 2018.

[29] T. D. Parsons, P. Larson, K. Kratz, M. Thiebaux, B. Bluestein, J. G. Buckwalter, and A. A. Rizzo, "Sex differences in mental rotation and spatial rotation in a virtual environment," *Neuropsychologia*, vol. 42, no. 4, pp. 555–562, 2004.

[30] R. Likert, "A technique for the measurement of attitudes," *Archieves of Psychology*, vol. 22 140, pp. 1–55, 1932.

[31] E. Yuda, K. Yamamoto, Y. Yoshida, and J. Hayano, "Discrimination of emotional type by heartbeat signal information," in *IEEE 7th Global Conference on Consumer Electronics, GCCE 2018, Nara, Japan, October 9-12, 2018*, 2018, pp. 531–532.