Research on the Application of GSR and ECG in the Usability Testing of an Aggregation Reading App

Sha Liu*, Bao-Yue Zhang, Cong Liu

College of Engineering
China Agricultural University
17 Qinghua Donglu Haidian District
Beijing 100083, People's Republic of China
E-mails: Lius02@cau.edu.cn, 1024520610@qq.com, 1036851037@qq.com

Corresponding author

Received: January 27, 2016 **Accepted: September 10, 2016**

Published: September 30, 2016

Abstract: *Usability testing is a very important step in improving App design and development. The traditional usability testing methods are based on users' expressions and behaviors, which hardly show users' emotional experience and cognitive load in real time. The introduction of an electrophysiological technique can make up for the deficiency of the traditional usability testing methods. In this study, a usability testing was carried out with the old and the new version of an App software. The behavior and the subjective evaluation of the participants were recorded, and their GSR and ECG signals were collected. Then, 14 physiological characteristics, such as GSR-Mean, LF, HF, LF/HF, etc., were extracted from the GSR and ECG signals. These characteristics were analyzed, and a significance test of difference of the two versions was made. This research indicated that there is a certain application value of GSR and HRV in usability testing and evaluation of an App product. But the meanings of the physiological characteristics must be explained in combination with the behavior and subjective evaluation of users. The result can prove that physiological characteristics have obvious advantages in real-time monitoring users' emotional changes, which can be helpful to find the usability problems of the product.*

Keywords: *GSR, HRV, ECG, Usability testing.*

Introduction

Usability is one of the core competitive powers of App software, so usability testing is a necessary step of the process of App design and development. Today, with the more and more increasing concern about users' experience with software, it is very important to select the appropriate usability testing methods which can make a timely, accurate and effective evaluation of the product and find the points for improvement. Currently, the main methods of usability testing include user testing, heuristic evaluation, contextual inquiry, cognitive walkthrough, FG, scenarios interviews, surveys, Goals, operators, methods, and selection rules (GOMS) model, etc., and among them user testing is the most commonly used research method.

However, either quantitative or qualitative, the common methods depend on users' active report to collect data and related information, so the collected information will be affected by the users' understanding and description ability, and it is subjective, unstable and difficult to verify. Moreover, it is hard to collect the real-time information in the test. Lastly, users may also deliberately conceal or adjust their true feelings for a certain purpose or unconsciously subject to the guidance of researchers [6, 22].

Electrophysiology and related equipment have developed rapidly in recent years, and the researches on users' physiological indexes have provided a new direction for quantitative and real-time research of users' subjective experience and the evaluation of product's usability. Now, the physiological indexes which have been used in user research mainly include Electromyography (EMG), Electrocardiograph (ECG), Galvanic skin response (GSR) or Electrodermal activity (EDA), blood pressure (BP), Electroencephalography (EEG), Respiration rate, etc. Among them, GSR and ECG are easier to collect and can effectively present the level of physiological arousal, so they are suitable to use in researches on user experience [10].

GSR is the most widely used physiological index at present [2]. Skin conductance is linearly related to the arousal level [9, 14]. Some studies show skin conductance levels appear to rise in the following cases: 1) when the users are attracted by the tasks; for example, when the difficulty of a game is increased or users feel it is funny [17]; 2) poor usability or more problems encountered with the system alarm [18, 23, 29]. But, on the other hand, studies also suggest that users will appear to have raised skin conductance levels when they send and receive emails through the website with the better usability [3, 15, 19]. The above researches show that the increase in skin conductance is surely related to users' arousal level, but if we want to judge whether usability is good or bad, it is still necessary to make an analysis of the context and the actual situation.

ECG is a typical representative of peripheral physiological indexes [15]. ECG has many indexes, and among them Heart rate (HR) and Heart rate variability (HRV) are the most commonly used indexes in user research. HR can reflect emotional and cognitive activities, and it is more sensitive to cognitive demands, time constraints, uncertainty and attention levels [1, 24]. HRV is the change of heart rate, which is an effective index of mental load and emotional state [2, 7, 11]. As an index of user testing HR is: firstly, a sensitive indicator of pressure, which can be accelerated under pressure [20]; secondly, an indicator of emotional arousal, as positive or negative strong emotions can also lead to HR acceleration. In [29] authors figured out that HR would drop after 3 minutes when they used a website with better usability, while HR would keep a relatively higher level when they used a website with worse usability. On the other hand, HRV is also a very important index of a human's emotional state. Generally speaking, the high frequency (*HF*, 0.15~0.4 Hz) component reflects the function of the vagus nerve, and the low frequency (*LF*, 0.04~0.15 Hz) component reflects the function of the sympathetic nerve. The ratio of the low frequency component and the high frequency component (*LF/HF*) reflects the balance between the sympathetic and the vagus nerves. So *LF* can be seen as a signal of anxiety and nervousness, and it will increase with the increasing of the task's difficulty [16, 26]. Authors in [13] proposed by a research on a webpage game usability that the higher the power of Mid-Frequency (0.07~0.15 Hz) were, the more relaxed the person would feel.

The above researches show that GSR and ECG are the earlier electrophysiological indexes which are used in users' experience research and have shown a relatively stable result [21, 25, 27, 30]. However, there are still some different opinions on how to explain the correspondence between the physiological indexes and usability [8]. It is worth to further verify and summarize the relevant laws. In addition, the research on the application of GSR and ECG in App usability testing is relatively less now, and it is worth further research and exploration.

In the present study, usability testing with an aggregation reading App is carried out, and GSR and ECG are collected in the process (Fig. 1). Then, the relative indexes, such as GSR, HR, HRV etc., are extracted and compared with the results of the usability testing and the users' subjective evaluation. Finally, the relationship between electrophysiological indexes and users' experience is analyzed and the usability of the two versions of the App is evaluated according to the above studies.

Materials and methods

Usability testing of an aggregation reading App

Two versions of an aggregation reading App software, "a little information", are chosen as the study objects whose main functions are to be tested. In the process of testing, users' actions and test results are recorded by a camera and record screen software; their subjective evaluation is investigated through a questionnaire, and their GSR and ECG data are collected by the electrophysiology synchronization instrument and BioLab system, which are a product of an American company, MindWare.

The object of usability testing

The Table 1 shows the main interfaces of the two versions of the App, "a little information", which can provide users' personalized and social news reading services. The main functions of the App include:

a) To subscribe to any channel of interest;
b) To read, collect, review and share the information and hot spots of the subscription channel;
c) In addition, the function of offline download, notification push, night mode are provided by the App.

Based on these functions, usability testing sets up 7 tasks, which include: left slide menu, discovery, offline download, collection, sharing, editing channel and setting. Then the users' behaviors and performances as well as subjective evaluation are compared with the physiological signals collected in the testing. Finally, the differences between the old and the new version are analyzed and the relationship between the main indexes of GSR and ECG and users' experience is studied.

The process of usability testing

47 (25 males, 22 female) college students were recruited to participate in the test. They were 18-26 years old, with normal or corrected to normal vision, no color blindness and heart diseases. They were all using smart phones with Android system, but they had not experimented with the "a little information" App.

The participants were divided into 2 groups which had a similar sex ratio. One group was tested with the new version of the App, while the other group was tested with the old version.

In order to reduce the influence of temperature and electromagnetic radiation on electrophysiological signals, the test was carried out in a laboratory environment (22 ± 3 °C) and no strong electromagnetic radiation. Taking into account the different arousal level of the participants at different times, all the tests were carried out in the evening (6:00~9:00 pm).

At the beginning of the test, the participants were introduced to the test's purpose, the procedure and key points to pay attention to during the test. Then, the electrode sheets were placed and the BioLab system was set up to confirm that the signal waveforms were normal, and signal collection began. As Fig. 2 shows: 1) The acquisition positions of skin electrical signals were on the thenar and hypothenar of the participant's subdominant hand (Fig. 2a);

2) The acquisition positions of the ECG signal were based on the CM5 bipolar chest lead. The positive (LL) was located in the junction of the left anterior axillary line and the 5th rib, the negative (RA) was located in the 1/3 in the right clavicle fossa, and the grounding wire (RL) was in a symmetric position to LL on the right side of the body (Fig. 2b).

Table 1. The two versions of "a little information" App

	Left slip menu	**Discovery news**	**Editing channels**
New version			
Old version			

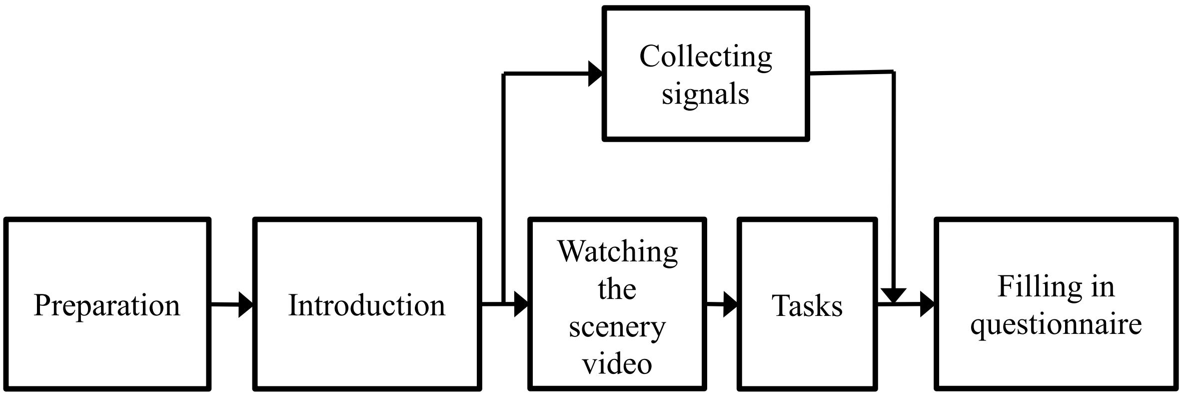
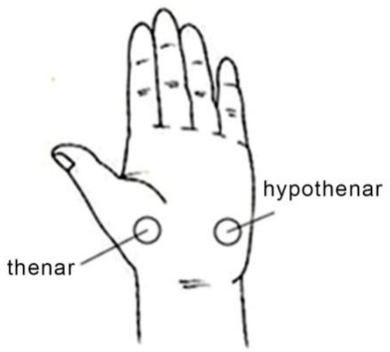
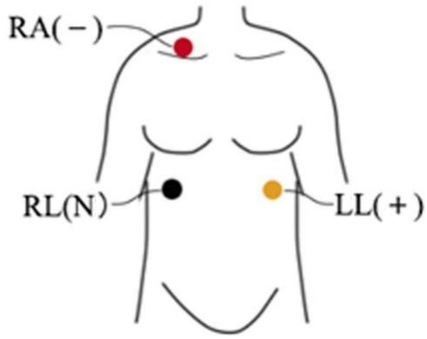


Fig. 1 The process of usability testing





a) The measuring position of GSR signals

b) The measuring position of ECG signals

Fig. 2 The measuring positions of GSR and ECG signals

The test began. Firstly, the participants watched a 3 minutes video of natural scenery to make them relax and calm. Then the participants completed the 7 tasks under the guidance of the staff, while at the same time their GSR and ECG signals were collected by the BioLab system, and a screen recording software recorded their behaviors during the test. In the course of the test, the participants operated the mobile phones with their dominant hand, while seated in a naturally relaxed position.

After the participants finished all the steps, they filled in a questionnaire with a satisfaction scale (5 points, 20 items). Finally, we carried out a brief interview with the participants to ask them about their difficulties, subjective feelings and evaluation of the App.

Results and discussion

Analysis of the results of traditional usability testing

In traditional usability testing, users' behavior and operation performances and the subjective evaluation after the test are the important bases to evaluate the product. The indexes of users' behavior and operation performances mainly include the completion rate of tasks, the mistakes, the time of completing the tasks, the times of guidance and the times of participants seeking help, etc., which mainly reflect the effectiveness and efficiency of the product. Further, the scores from the questionnaire or the scale and the users' subjective evaluation can reflect the users' subjective experience and feelings to a certain extent and have a certain reference value. Our test mainly recorded and analyzed the 4 common usability indexes: the completion rate of tasks, the mistakes, the time of completing each task, and user's satisfaction.

The completion rate of tasks: to figure the number of the subtasks that the participants did not effectively complete, and then calculate the average task completion rate for all the participants' operations.

Mistakes: to figure the sum of all the mistakes in all the tasks, and then calculate the average value of the mistakes made in the old and the new version, respectively.

The time of completing each task: to calculate the average amount of time the participants needed to complete each task in the old and the new version, respectively.

User's satisfaction: to figure out the average scores of the old and the new version, respectively, given by the participants in the final questionnaire.

Then, the significant difference of the average values of the old and the new version were figured out. The average value of each index and the significant difference test are shown in Table 2.

a. The samples are analyzed by a homogeneity test of variances (F-Test). The completion rate of tasks, mistakes and user's satisfaction are according to normal distribution and homogeneity of variance. So, they are analyzed by a t-Test and it is proved that there are significant differences between the new and the old version.
b. Only the times of completing tasks *b*, *d*, *e*, and *f* are according to normal distribution and homogeneity of variance, and they are analyzed by a t-Test, while only tasks *d*, *e*, *f* show a significant difference between the new and the old version ($\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.05$). The times of completing tasks *a*, *c*, *g* are analyzed by a t-Test and only task *c* shows a significant difference between the new and the old version ($\alpha = 0.05$).
c. The questions in the final questionnaire are classified. The results of the questions are summarized as 6 sub-indexes (complexity, easy to use, easy to understand, attractive, easy to learn, and comfortable) and the mean values of each index are calculated. As Table 2c shows, the new version is easier to learn, understand and use, and users also think it looks more beautiful and attractive.

By the analysis above, the completion rate of tasks and user's satisfaction with the new version are higher than with the old version, and the mistakes with the new version are fewer than with the old version. Moreover, for tasks *b*, *c*, *d*, and *f*, there are significant differences between the new and the old version, and the time of completing a task with the new version is shorter than with the old version. Therefore, according to the indexes of traditional usability testing, the new version has a higher usability and provides better user's experience.

Table 2. Statistics of the usability indexes

Table 2a. The completion rate of tasks, mistakes and satisfaction

	Completion rate of tasks, (%)	**Mistakes**	**User's satisfaction**
Old version	91.09	8.15	74.5
New version	94.20	6.40	83.8
Significant difference	Y ($\alpha = 0.1$)	Y ($\alpha = 0.05$)	Y ($\alpha = 0.05$)

Table 2b. Statistics of the time of completing each task

	Time of completing task(s)						
	Task *a*	**Task *b***	**Task *c***	**Task *d***	**Task *e***	**Task *f***	**Task *g***
Old version	92.15	26.32	68.40	68.40	52.12	83.00	45.12
New version	88.18	32.35	62.45	62.45	49.45	45.32	39.08
Significant difference	N	N	Y ($\alpha = 0.05$)	Y ($\alpha = 0.1$)	Y ($\alpha = 0.05$)	Y ($\alpha = 0.05$)	N

Table 2c. Statistics of the sub-indexes of the subjective evaluation of the users by classified problems of the final questionnaire

	Complexity	**Easy to use**	**Easy to understand**	**Attractive**	**Easy to learn**	**Comfortable**
Old version	4.0	4.1	3.7	3.5	4.6	3.9
New version	4.3	4.6	4.4	4.0	4.8	4.1
Significant difference	N	Y (α = 0.05)	Y (α = 0.01)	Y (α = 0.05)	Y (α = 0.1)	N

The processing of ECG and GSR signal and characteristics extraction

After finishing all the tests, every participant had 8 groups of GRS data and EMG data. By observing the signal waveforms, the 60 seconds to 140 seconds physiological signals were selected, processed and used as the calm-state average GSR (X_{calm}) and mean R-R interval ($MeanNN_{calm}$).

The processing of GSR signal and characteristics extraction

Firstly, the 7 groups of GSR of each participant are standardized. The 7 GRS data matrixes respectively subtract the calm-state average GSR, and the new data matrixes are the standardized GSR data (X_{calm}). That is:

$$X_0 = X_{task} - X_{calm}, \tag{1}$$

where X_0 is the standardized GSR data, X_{task} is the GSR data collected while the participants performed the tasks, and X_{calm} is the calm-state average GSR. Five common time-domain characteristics are extracted from the standardized GSR data of the 7 tasks, according to the method promoted by the scholars in Augsburg University. All characteristics are shown in Table 3.

Table 3. The time-domain characteristics of GSR

Characteristic	**The meaning of characteristic**
GSR-Mean	The average level of GSR
GSR-STD	The standard deviation of GSR
GSR-Max	The maximum value of GSR
GSR-Min	The minimum value of GSR
GSR-Range	GSR-Max – GSR-Min

The HRV characteristics extraction from ECG signals

In the ECG waveform, R wave is the first wave of the QRS wave group, which is usually the most obvious characteristic wave with the highest amplitude, so it is easy to detect. In order to detect the R wave, the R wave peak point (the point of the maximum amplitude) is regarded as the position of the R wave. Then, the length of each R-R interval, which is the duration of the cardiac cycle, is figured out [5, 12, 28]. In this study, a MATLAB program has been used to detect the position of the R wave through the threshold detection method and R-R interval is calculated by Excel software. Then, according to the R-R interval, the main HRV time-domain characteristics and frequency-domain characteristics are extracted, as shown in Table 4 and Table 5.

Table 4. The time-domain characteristics of HRV

	Characteristic	Meaning	Formula
Normalized *Mean NN*	Normalized mean value of R-R interval	The average level of R-R interval after removing the individual differences	$Mean\ NN = \frac{1}{N}\sum_{i=1}^{N} RR_i - Mean\ NN_{calm}$ *$Mean\ NN_{calm}$ is the average value of R-R interval in calm-state.
SDNN	Standard deviation of R-R interval	The whole change of R-R interval	$SDNN = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(RR_i - \overline{RR})^2}$ $\overline{RR} = \frac{1}{N}\sum_{i=1}^{N} RR_i$
rMSSD	Mean variance of the difference between adjacent R-R intervals	The fast change of R-R intervals	$rMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2}$
PNN50	The Percentage of the difference between adjacent R-R intervals is greater than 50 ms	The sudden change of R-R interval, which can sensitively reflect the activity of the vagus nerve	$PNN50 = n/(N-1) \times 100\%$ *n is the number of the absolute value of the difference between adjacent R-R intervals greater than 50 ms.

The frequency-domain characteristics of HRV are extracted by HRV Power spectrum density (PSD) [4]. In the study, the method of cubic spline interpolation is elected to interpolate to R-R interval sequence and then to resample it. The frequency of resampling is 16 Hz, and, ultimately, the R-R interval sequence with an equal interval can be figured out. Then, the BT spectrum estimation method is used to make the estimation of PSD. Fig. 3 shows a part of the power spectrum of the participants. The power spectrum reflects the distribution of energy (power) in different frequency bands, and the area of frequency band under the power spectrum curve can be used as well. In the PSD, respectively integrating the intervals of [0.04~0.15 Hz] and [0.15~0.4 Hz], *LF* and *HF* can be figured out.

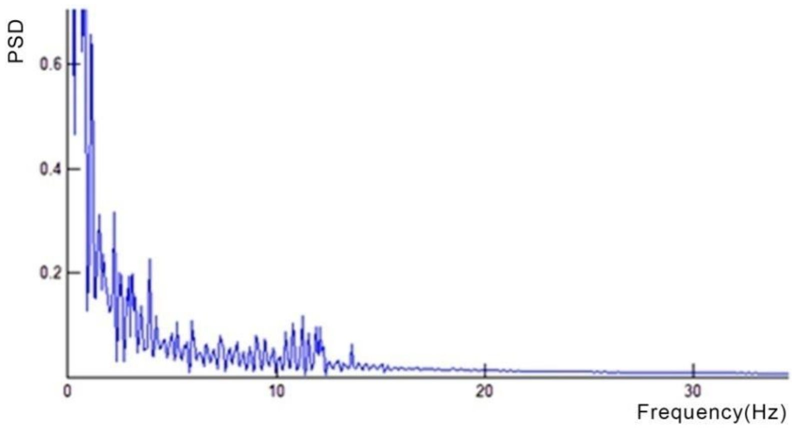


Fig. 3 HRV power spectrum density (Part)

Finally, *LF*, *HF*, *LF/HF*, *LFnorm* and *HFnorm*, the five frequency-domain characteristic of HRV which mostly reflect the changes of autonomic nervous activity can be extracted from the estimated PSD.

Table 5. The frequency-domain characteristics of HRV

	Characteristic	Meaning	Formula
LF	Low frequency power	It mainly reflects the level of activity of the sympathetic nerve	
HF	High frequency power	It mainly reflects the level of activity of the vagus nerve	
LF/HF	*LF/HF*	It mainly reflects the balance between the sympathetic and the vagus nerves	*LF/HF*
LFnorm	Standardized low frequency power	It reflects the changes of autonomic nerve regulation, avoiding the effect of total power.	$LFnorm = \frac{LF}{TP-VLF} \approx \frac{LF}{LF+HF}$ **TP* is the total power under 0.4 Hz. **VLF* is the Ultra low frequency power (0.0033~0.04 Hz).
HFnorm	Standardized high frequency power		$HFnorm = \frac{HF}{TP-VLF} \approx \frac{HF}{LF+HF}$ **TP* is the total power under 0.4 Hz. **VLF* is the Ultra low frequency power (0.0033~0.04 Hz).

Statistical analysis of HRV and GSR

Among the participants of the test, there are 35 participants whose GSR signals are normal and effective. 17 of them are tested with the old version and 18 with the new version. There are 41 participants whose ECG signals are normal and effective. 20 of them are tested with the old version and 21 with the new version.

After the processing of the GSR and ECG signals collected in the tests, the total 14 physiological characteristics were extracted, including the time-domain characteristics (GSR-Mean, GSR-STD, GSR-Max and GSR-Min), HRV time-domain characteristics (Normalized *Mean NN*, *SDNN*, *rMSSD*, *PNN50*) and HRV frequency-domain characteristics (*LF*, *HF*, *LF/HF*, *LFnorm* and *HFnorm*). The mean value of the physiological characteristics with the old and the new version and the results of significant difference between the two versions are shown in Tables 6-8.

Table 6. The statistical results of GSR time-domain characteristics

	GSR-Mean	**GSR-STD**	**GSR-Max**	**GSR-Min**	**GSR-Range**
Old version	0.3149	0.07601	0.4897	0.1873	0.3023
New version	0.4050	0.07725	0.5858	0.2781	0.3077
Significant difference	Y ($\alpha = 0.1$)	N	Y ($\alpha = 0.1$)	Y ($\alpha = 0.05$)	N

Table 7. The statistical results of HRV time-domain characteristics

	Normalized *Mean NN*	***SDNN***	***rMSSD***	***PNN50*（%）**
Old version	–0.02074	0.03832	0.03874	18.74
New version	–0.01581	0.04378	0.04521	24.18
Significant difference	N	Y ($\alpha = 0.01$)	Y ($\alpha = 0.01$)	Y ($\alpha = 0.01$)

Table 8. The statistical results of HRV frequency-domain characteristics

	LF	***HF***	***LF/HF***	***LFnorm***	***HFnorm***
Old version	0.4279	0.3016	1.2042	0.5128	0.4872
New version	0.8532	0.6006	1.0998	0.4961	0.5039
Significant difference	Y ($\alpha = 0.1$)	Y ($\alpha = 0.1$)	N	N	N

Statistical results determine that the eight characteristics of GSR-Mean, GSR-Max, GSR-Min, *SDNN*, *rMSSD* and *PNN50*, *LF* and *HF*, showed a significant difference between the old and the new version, and the levels with the new version were higher than those with the old version. As for the other six characteristics, GSR-STD, GSR-Range, standardized *Mean NN*, *LF/HF*, *LFnorm* and *HFnorm*, there was no significant difference between the old and the new version.

Based on the traditional usability indexes, the new version had a higher usability than the old version. Then, by the statistical results of physiological characteristics, the following can be seen:

a. GSR time-domain characteristics: The GSR-Mean, GSR-Max and GSR-Min with the new version were all higher than those with the old version while in GSR-STD and GSR-Range there was no significant difference between the two versions. It proved that the users' sympathetic nerve activity and their emotional arousal were higher when they used the new version. It may be concluded that users would have a higher GSR levels when they use the version with the better usability or the more attractive one. The above can be explained with the fact that the new version is more easy to use and has more beautiful visual effects, which stimulate users' positive feelings of pleasure and can result in a higher level of GSR.

b. HRV time-domain characteristics: The *SDNN*, *rMSSD* and *PNN50* with the new version were higher than those with the old version, while the standardized *Mean NN* between the old and the new version had no significant difference. These showed that the instantaneous HR change and the rapid change of HR were strongly significant when users used the version with the better usability or with the more beautiful visual effects, but in the average HR there was not a significant difference between the two versions. Thus, it can also be speculated that the new version with the better usability can help users quickly focus their attention and can improve the level of arousal to complete the tasks.

c. HRV frequency-domain characteristics: *LF* and *HF* with the new version were higher than those with the old version, while in *LF/HF*, *LFnorm* and *HFnorm* there was no significant difference between the old and the new version. These showed that both sympathetic activity and vagal activity were more active when users used the new version, but the two activities kept their balance.

In a word, the statistical results of HRV time-domain characteristics and frequency-domain characteristics all proved that users had a higher HRV when using the new version, which reflected the more intense emotional experience and the higher level of emotional arousal.

Conclusion

Usability is one of the core competitive powers of App software, so usability testing is a very important step of the process of App design and development. The traditional usability testing methods are based on users' expressions and behaviors, which hardly indicate users' emotional experience and cognitive load in real-time. The introduction of electrophysiological technique can make up for the deficiency of the traditional usability testing methods.

In this study, a usability testing was carried out with the old and the new version of an aggregation reading App named "a little information". The behaviors and subjective evaluation of the participants in the test were recorded and their GSR and ECG signals were collected by Biolab system. Then, 14 physiological characteristics: GSR-Mean, GSR-STD, GSR-Max, GSR-Min, *Mean NN*, *SDNN*, *rMSSD*, *PNN50*, *LF*, *HF*, *LF/HF*, *LFnorm* and *HFnorm* were extracted from the GSR and ECG signals. These characteristics were analyzed and a significance test of difference of the two versions was made.

The results of the research show that users' GSR level and HRV are all higher when they use the new version. On the other hand, it is proved that the new version has better usability and more beautiful visual effects. As the earlier studies had shown, the higher levels of GSR and HRV are related to high arousal or higher pressure loads. Since the cognitive load of the new version is not higher than the old version by the traditional usability evaluation, (Table 2c), it can be speculated that the higher GSR and HRV levels in the test may be related to the positive emotional experience and the higher emotional arousal. The higher levers of *SDNN*, *rMSSD* and *PNN50* (%) may also prove that the new version can help users quickly focus their attention and improve the level of arousal to complete the tasks.

Generally speaking, better usability often leads more easily to a pleasant, relaxed and positive emotion, but users' cognitive load is often lower and mental effort is often less. The higher GSR and HRV levels reflect the higher emotional arousal, which can be caused by either the strong senses of the strong positive emotions such as pleasure, accomplishment etc., or the strong negative senses, such as disgust, anger, frustration etc. In the present research, the higher level of GSR and HRV can be explained with the fact that the new version can better attract users' attention and make them more efficient in completing the task.

Therefore, although the physiological characteristics have obvious advantages in real-time monitoring users' emotional changes, arousal level and cognitive load, we still should analyze and explain the meanings of physiological characteristics in combination with the behavior and subjective evaluation of users. The change of GSR and HRV in real-time can also be used as a basis to discover the problems in the product.

Acknowledgements

This research is supported by the National Key R & D program "R & D of Multifunction Cane and Beet Harvest Technology and Equipment" (2016YFD0701200), Sub-project "Research on Key Technologies of Efficient Harvesting and Intelligent Control for Cane and Beet Efficient Harvesting Equipments" (2016YFD0701201).

References

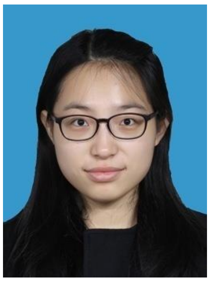
1. Allanson J., S. H. Fairclough (2004). A Research Agenda for Physiological Computing, Interacting with Computers, 857-878.
2. Boucsein W. (2012). Electrodermalactivity, New York: Springer.
3. Bruneau D., M. Sassem, J. D. McCarthy (2002). The Eyes Never Lie: The Use of Eyetracking Data in HCI Research, Proceedings of the CHI'02 Workshop on Physiological Computing, New York: ACM Press.
4. Buqing W., W. Weidong (2007). Research Progress of HRV Analysis Method, Beijing Biomedical Engineering, 26(5), 551-554.
5. Christov I., I. Jekova, V. Krasteva, I. Dotsinsky, T. Stoyanov (2009). Rhythm Analysis by Heartbeat Classification in the Electrocardiogram, International Journal Bioautomation, 13(2), 84-96.
6. Czerwinski M., E. Horvitz, E. Cutrell (2001). Subjective Duration Assessment: An Implicit Probe for Software Usability, Proceedings of the IHM-HCI 2001 Conference, Lille, France, Vol. 2, 167-170.
7. Dirican A. C., M. Göktürk (2011). Psychophysiological Measures of Human Cognitive States Applied in Human Computer Interaction, Procedia Computer Science, 3, 1361-1367.
8. Foglia P., C. A. Prete, M. Zanda (2008). Relating GSR Signals to Traditional Usability Metrics: Case Study with an Anthropomorphic Web Assistant, Proceedings of the IEEE Instrumentation and Measurement Technology Conference, 1814-1818.
9. Ganglbauer E., J. Schrammel, S. Deutsch, M. Tscheligi (2009). Applying Psychophsiological Methods for Measuring User Experience: Possibilities, Challenges and Feasibility, Workshop on User Experience Evaluation Methods in Product Development, UXEM'09, Uppsala, Sweden.
10. Ge Y., Y. Chen, Y. Liu, W. Li, X. Sun (2014). Electrophysiology Measures Applied in User Experience Studies, Advances in Psychological Science, 22(6), 959-967.
11. Hercegfi K. (2011). Heart Rate Variability Monitoring during Human-computer Interaction, Acta Polytechnica Hungarica, 8(5), 205-224.
12. Jekova I., V. Tsibulko, I. Iliev (2014). ECG Database Applicable for Development and Testing of Pace Detection Algorithms, International Journal Bioautomation, 18(4), 377-388.
13. Károly H. (2011). Improved Temporal Resolution Heart Rate Variability Monitoring – Pilot Results of Non-laboratory Experiments Targeting Future Assessment of Human-Computer Interaction, International Journal of Occupational Safety and Ergonomics, 17(2), 105-117.
14. Lang P. J., M. K. Greenwald, M. M. Bradley, A. O. Hamm (1993). Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reaction, Psychophysiology, 30(3), 261-273.
15. Lean Y, F. Shan (2012). Brief Review on Physiological and Biochemical Evaluations of Human Mental Workload, Human Factors and Ergonomics in Manufacturing & Service Industries, 22(3), 177-187.
16. Lin T., A. Imamiya, X. Mao (2008). Using Multiple Data Sources to Get Closer Insights into User Cost and Task Performance, Interacting with Computers, 20(3), 364-374.
17. Mandryk R. L., K. M. Inkpen, T. W. Calvert (2006). Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies, Behaviour & Information Technology, 25(2), 141-158.
18. Marek T., W. Karwowski, V. Rice (Eds.) (2011). Advances in Understanding Human Performance: Neuroergonomics, Human Factors, New York, CRC Press.
19. McCarthy J. J. (2002). A Thematic Guide to Optimality Theory, Research Surveys in Linguistics, Cambridge: Cambridge University Press.

20. Meehan M., S. Razzaque, B. Insko, M. Whitton, F. P. Brooks Jr. (2005). Review of Four Studies on the Use of Physiological Reaction as a Measure of Presence on Stressful Virtual Environments, Applied Psychophysiology and Biofeedback, 30(3), 239-258.
21. Nacke L. E., S. Stellmach, C. A. Lindley (2010). Electroencephalographic Assessment of Player Experience: A Pilot Study in Affective Ludology, Simulation Gaming, 42(5), 632-655.
22. Nielsen J., J. Levy (1994). Measuring Usability: Preference vs. Performance, Communications of the ACM, 37(4), 66-75.
23. Pifister H. R., S. Wollstadter, C. Peter (2011). Affective Responses to System Messages in Human-computer-interation: Effect of Modality and Message Type, Interacting with Computers, 23(4), 372-383.
24. Pollatos O., B. M. Herbert, E. Matthias, R. Schandry (2007). Heart Rate Response after Emotional Picture Presentation is modulated by Interoceptive Awareness, Int J Psychophysiol, 63(1), 117-124.
25. Riseberg J., J. Klein, R. Fernandez, R.W. Picard (1998). Frustrating the User on Purpose: Using Biosignals in a Pilot Study to Detect the User's Emotional State, Proceeding of the CHI'98 – Conference Summary on Human Factors in Computing Systems, 227-228
26. Rowe D. W., J. Sibert, D. Irwin (1998). Heart Rate Variability: Indicator of User State as an Aid to Human-computer Interaction, Proceedings of the Conference on Human Factors in Computing Systems, ACM: New York, 480-487.
27. Scheirer J., R. Fernandez, J. Klein, R.W. Picard (2002). Frustrating the User on Purpose: A Step toward Building an Affective Computer, Interacting with Computers, 14(2), 93-118.
28. Tanev S. (2013). Ventricular Beat Detection and Classification in Long Term ECG Recordings, International Journal Bioautomation, 16(4), 273-290.
29. Ward R. D., P. H. Marsden (2003). Physiological Responses to Different WEB Page Designs, International Journal of Human-Computer Studies, 59(1-2), 199-212.
30. Wilson G. M., M. A. Sasse (2004). From Doing to Being: Getting Closer to the User Experience, Interacting with Computers, 16, 697-705.

Sha Liu, Ph.D.
E-mail: Lius02@cau.edu.cn

Sha Liu has a Ph.D. degree in Design. She is an Associate Professor at the College of Engineering, China Agricultural University in Beijing, China. Her main research fields are design psychology and human factor.

Bao-Yue Zhang
E-mail: 1024520610@qq.com



Bao-Yue Zhang is a graduate student at the College of Engineering, China Agricultural University in Beijing, China. Her main research fields are design psychology and computer-aided industrial design.

Cong Liu, M.Sc.
E-mail: 1036851037@qq.com

Cong Liu has a Master's degree in Mechanical Design and Theory. He graduated from the College of Engineering, China Agricultural University in Beijing, China in 2014. His main research fields are interactive design and user experience.