Application of Frontal EEG Asymmetry to User Experience Research

Jing Chai[1,2], Yan Ge[1,*], Yanfang Liu[3], Wen Li[3], Lei Zhou[3], Lin Yao[3], and Xianghong Sun[1]

[1] Key Laboratory of Behavioral Science, Institute of Psychology, CAS
{chaij,gey,sunxh}@psych.ac.cn
[2] University of Chinese Academy of Sciences
[3] User Experience Lab, China Mobile Research Institute, Beijing, China
{liuyanfang,liwen,zhoulei,yaolin}@chinamobile.com

Abstract. The electrophysiology technique now provides an alternative way to evaluate users' emotional states in real time, but how to confirm the valence of emotions using these techniques is still a concern to researchers. Frontal alpha asymmetry (FAA) is often used as an index of pleasantness or liking in neuromarketing, but results in related fields are not consistent. In this study, we investigated the emotional states of users interacting with mobile phone applications (APPs) using FAA. Twenty participants participated in this experiment. They were asked to complete several tasks in a scene of everyday life using three APPs of the same type. EEG data and subjective evaluations were recorded during the experiment. The FAA results showed a positive trend when using an APP that provided an excellent user experience. The mechanism of emotional change during interacting with mobile applications and the implications of this research are also discussed in this study.

Keywords: user experience, emotional state, FAA, EEG.

1 Introduction

According to ISO 9241-210 [1], user experience (UX) is defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system, or service." This concept represents the integrated feelings users get while using a product, including the way they understand a product (understanding experience) and their emotional response to it (emotional experience) along with the pleasure experienced from sensory perceptions (esthetic experience) [2]. The emotional experience is frequently measured by subjective evaluations in usability testing. The self-report method can be influenced by social desirability. This evaluation was often performed after participants completed the tasks, making it difficult to measure changes in emotional states in real time [3]. The electrophysiology technique provides an alternative solution for this problem. Skin conductivity, heart rate, blood pressure, etc. are

* Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang Dist., Beijing, China, 100101.

D. Harris (Ed.): EPCE 2014, LNAI 8532, pp. 234–243, 2014.

effective indices for measuring arousal levels during human–machine interactions [4-6], but it is difficult to distinguish the valence of emotions during the interactions using these indices.

The electroencephalography (EEG) has been used in UX fields to measure emotional experiences and cognitive workloads, with α and β as the main indicators [7-12]. Li, et al. recorded EEG data when users were visiting a distance learning simulation website. The results showed a correlation between EEG-α rhythm (8-13 Hz) and emotional state. The amplitude, amplitude's deviation, and power of α wave decreased when users were reading simple and interesting content rather than boring content [8]. Stickel, et al. also used EEG recording in usability testing and analyzed several specific frequency bands including β (12.5–28 Hz) and α (8–12.5 Hz). This research revealed that learnability can be estimated by brainwave patterns: the α waves were dominant when the software was easy to learn; the β waves were dominant when the software was hard to learn [12]. EEG spectral analysis can also be used to assess the effects of pre-training on learning software. Masaki, et al. found that the β/α power decreased with the software use experience, in which they took β/α value as an index of mental workload. If the β/α value was greater than 1.0, it indicated a state of higher mental workload; otherwise, a value lower than 1.0 indicated a state of lower mental workload [9].

Although early studies found that the EEG could be a common indicator of cognitive workloads and emotional states, how to identify the valence of emotions when users interacted with a product remains an open question. The correlation between the hemispheric asymmetries in pre-frontal activity and approach withdraw related motivation is extensively used in basic research [13]. The frontal alpha asymmetry (FAA) is a potential and valuable index in this research field when used to index frontal brain activity when processing different states of emotion. In Davison's model [14], the left pre-frontal cortex (PFC) is involved in the processing of positive affects, whereas the right PFC is involved in the processing of negative affects. This index has been widely used in neuromarketing [10, 11]. Ohme, et al. found that FAA can capture the difference between two slightly different versions of TV ads in a few seconds [11]. It has also been used to explore preferences in product design [15-17].The FAA provides a new solution to the measurement problem of human–machine interaction.

The main purpose of this research is to investigate the feasibility of using neurophysiology index to evaluate emotional experiences in UX, with FAA as the main index of affects valence. Three different mobile applications were tested in the experiment, and all the behavior, subjective, and physiological data were collected. We assume a consistency between self-report users of experiences and frontal asymmetric activities, which will make a distinction between positive and negative emotion.

2 Methods

2.1 Participants

Twenty participants were recruited from an open access, online part-time job platform in China (http://zhan.renren.com/jobcome).They aged from 21 to 29 years old (mean: 23.8 ± 2.484), including ten females and ten males. Their experience of using this type of APP and emotional sensitivity were controlled.

2.2 Materials

Mobile applications. Three APPs, designated DZD, YEW, and HMB, were chosen as experimental materials according to their usability testing results in pilot studies. These APPs have similar functions, but different usability ratings. They provided a lot of information about restaurants, hotels, and entertainment venues in a city. Users could search, order, or evaluate a restaurant using APPs like these. The latest versions of all APPs were installed on a smart phone using an Android operating system in advance of the experiment.

Questionnaires. Three self-report questionnaires were used as subjective measurement in this study.

The Positive and Negative Affect Scale (PANAS). This scale was used to measure the current mood of participants. According to Watson and Tellegen's two-factor model, positive affect (PA) and negative affect (NA) are two basic and mutually independent dimensions in the structure of emotion [18]. PA refers to people's feelings of enthusiasm, activeness, and pleasure. A high PA value represents a state of powerfulness, concentration, and joviality. The opposite, NA relates to distress and unpleasantness, including anger, contempt, disgust, guiltiness, fear, and tension. A low NA value represents peace and calmness. The PANAS consists of 20 affective words, 10 for each dimension. The participants were to indicate to what extent their state matched each word on a scale from 1 ("very slightly or not at all") to 5 ("very much") [19].

User Experience Questionnaire (UEQ). UEQ is a widely used and most common tool for software quality and usability assessment. It can be used to assess the comprehensive impression of user experience in a convenient and quick way. It consists of six relatively independent factors, including attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty [20]. This scale includes 26 items, each composed of an adjective and its antonym. Users evaluated their preference between each pair of words using a 7-point scale.

System Usability Scale (SUS). SUS is used for usability assessment of software systems, products, or websites, especially for competitive analysis [21]. SUS includes 10 declarative sentences. The participants rated each sentence from 1 (not at all) to 5 (very much) according to how much they agreed with those sentences. Note that the SUS score only represents the overall usability of a system, and different attributes of the system such as effectiveness and efficiency are not measured in the SUS.

2.3 Procedures

When the participants arrived at the laboratory room, the moderator introduced them to the procedures and tasks of the experiment in detail and then prepared them for EEG recording. Next, the participants were asked to complete the PANAS for the first time, which would be used as the baseline of affective state for the whole experiment.

In the formal study, the participants used three APPs respectively to perform a series of tasks. The order of using APPs was counterbalanced in a Latin square sequence. For each APP, the order of tasks was consistent. First, the participant sat quietly and statically for three minutes. The EEG data of this stage was recorded as a baseline. Then, a daily situation was setup as follows, "Suppose a friend of yours has recently come to visit this city. Please book a restaurant and a hotel room for this friend according to the experimenter's instructions." The participant searched for related information and made choices using an APP. When the participant finished a task, she/he reported their choice orally and answered several questions asked by the experimenter, e.g., the reasons for making such choices and the difficulty of the task. If the participant felt the task was extremely difficult to complete after several minutes, she/he could give up. There were no strict time constraints for each task. When a task was executed for over ten minutes, the participant received a warning. The completion rate, completion time, and the difficulty of the task reported by the participants were collected as the behavioral data. After the tasks for one APP were finished, the participants were asked to complete two questionnaires, the PANAS and UEQ. The EEG data was collected for the entire session. The entire assessment for one APP lasted for about 30 min. After a brief break, participants continued on to the second APP. The situations and tasks for the three APPs were similar, except the meeting locations for each APP were different. When the tasks for all three APPs were completed, participants compared the usability of the APPs using the SUS. The entire experiment lasted approximately two hours.

2.4 EEG Recording and Analysis

The EEG data was continuously recorded from 32 scalp sites using tin electrodes mounted in an elastic cap arranged according to the 10–20 international placement system (Neuroscan Inc.). The online reference was right mastoid (A2). The electrode sites on the participant's mastoids and forehead were cleaned with alcohol cotton balls gently. The impedances of the EEG electrodes were below 5 kΩ. The EEG data were amplified with a bandpass filter of 0.05–100 Hz and digitized at 500 Hz. The recordings obtained from the prefrontal and frontal regions of the cortex (Fp1, Fp2, F3, F4, F7, and F8) were saved.

The EEG data was processed using Neuroscan 4.5 software. All the EEG data were DC corrected and re-referenced to linked mastoids offline. The filter was set to 30 Hz low pass and 0.1 Hz high pass. Then, the EEG data were epoched into periods of 512 points (i.e., 1024 ms). The power of alpha band (8–12 Hz) and beta band (12–29.3 Hz) in each of the recorded electrodes was calculated for further analysis.

The frontal alpha asymmetry (FAA) index was calculated as the difference between right-hemispheric data minus left-hemispheric data (ln (right alpha power) - ln (left alpha power)) according to previous studies [22, 23]. Due to the negative correlation between alpha power and brain activation, the positive score of the FAA index implies the dominance of left PFC and the negative score of the FAA index implies the dominance of right PFC.

3 Results

All the data were processed by SPSS 16.0. The EEG data exceeding three standard deviations were considered as extreme values and removed from further analyses.

3.1 Behavioral Results

The behavioral data included the task completion rate, the task completion time, and the task difficulty, which were recorded by the experiment assistant during the experiment. A single-factor repeated measure of variance analysis was taken to analyze the difference among the three APPs. There are significant differences among the three APPs in the completion rate ($F = 44.333$, $p < 0.01$), completion time ($F = 12.314$, $p < 0.01$), and task difficulty ($F = 112.405$, $p < 0.01$). Further analysis showed that the completion rate of HMB was significantly lower than that of YEW ($p < 0.01$) and DZD ($p < 0.01$); the completion time of HMB was significantly longer than that of YEW ($p < 0.01$) and DZD ($p < 0.01$), and the self-report difficulty of HMB was significantly higher than that of YEW ($p < 0.01$) and DZD ($p < 0.01$). No significant difference was found between YEW and DZD in completion time ($p = 0.152$) or completion rate ($p = 0.163$). But YEW is significantly more difficult than DZD ($p < 0.01$). The descriptive statistics of the behavioral data are given in Table 1.

Table 1. Descriptive statistics of the behavioral data

	YEW (M ± SD)	DZD (M ± SD)	HMB (M ± SD)	*F*
Completion rate	0.950 ± 0.215	1.000 ± 0.000	0.500 ± 0.498	44.333**
Completion time (100s)	1.722 ± 1.038	1.437 ± 1.019	2.764 ± 1.612	12.314**
Task difficulty	2.025 ± 0.898	1.575 ± 0.794	3.775 ± 1.078	112.405**

* p < 0.05, ** p < 0.01

3.2 Subjective Evaluation

UEQ and SUS. Results from repeated measures of variance analysis showed a significant main effect for the three APPs ($F = 94.162$, $p < 0.01$) in UEQ. Results also revealed a significant main effect for the three APPs ($F = 113.274$, $p < 0.01$) in SUS. Pairwise comparisons showed that the SUS score of HMB was lower than that of YEW ($p < 0.01$) or DZD ($p < 0.01$). No significance was found between YEW and DZD in the SUS score ($p = 0.07$).

Table 2. Descriptive statistics of UEQ and SUS

	YEW (M ± SD)	DZD (M ± SD)	HMB (M ± SD)	*F*
UEQ	1.141 ± 1.064	1.465 ± 0.737	−1.072 ± 0.946	94.162**
SUS	82.375 ± 16.130	76.375 ± 15.780	20.125 ± 11.711	113.274**

* p < 0.05, ** p < 0.01

PANAS. The positive affect (PA) and the negative affect (NA) were analyzed respectively by repeated measures of variance analysis. Results showed significant differences among the three APPs in both PA ($F = 44.457$, $p < 0.01$) and NA ($F = 17.341$, $p < 0.01$). Pairwise comparisons showed that the PA value when using HMB was lower than when using YEW ($p < 0.01$) or DZD ($p < 0.01$), the NA value when using HMB was significantly higher than when using YEW ($p < 0.01$) or DZD ($p < 0.01$). No significance was found between YEW and DZD neither in PA value ($p = 0.702$) nor NA value ($p = 0.07$). The descriptive statistics of PANAS are given in Figure 1.

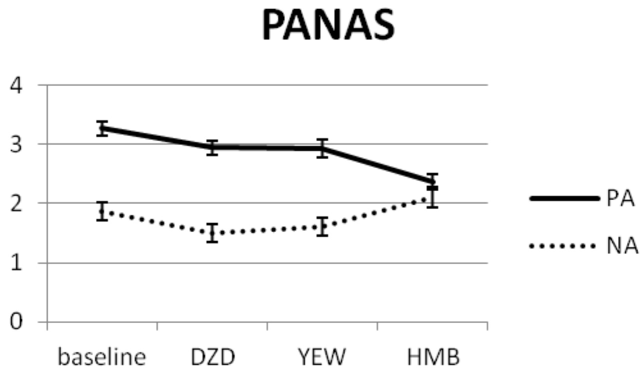

Fig. 1. Descriptive statistics of PANAS. Standard error bars are included

3.3 FAA Analyses

We only analyzed the data of the first two tasks for each APP. The results of FAA are shown in Figure 2. Unfortunately, results from repeated measures of variance analysis showed no significance in FAA index among the three APPs ($F = 1.417$, $p = 0.261$). But we could see a positive trend in DZD in comparison to the other two APPS.

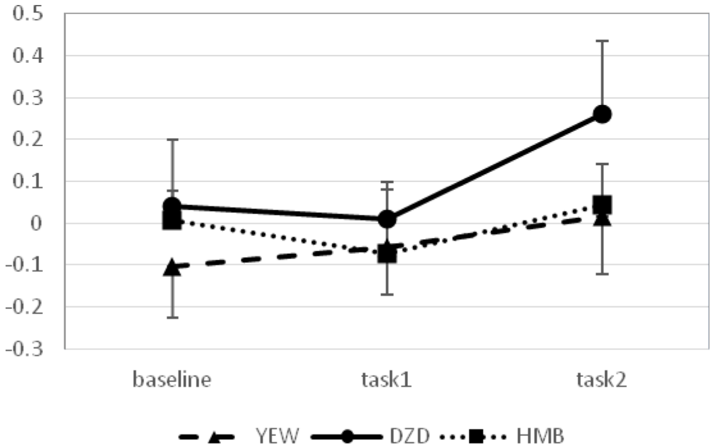

Fig. 2. Descriptive statistics of FAA index. Standard error bars are included.

3.4 Correlations

The correlations between EEG data and subjective data were also analyzed. Results of Pearson correlation coefficients revealed that the PA value is negatively correlated with both alpha power ($r = -0.310$, $p < 0.05$) and alpha power's change from baseline (i.e., task–baseline, $r = -0.309$, $p = 0.026$), and the NA change from baseline (task–baseline) is negatively correlated with both beta power ($r = -0.298$, $p < 0.05$) and beta power's change from baseline ($r = -0.318$, $p < 0.05$). The correlation between the FAA change (task–baseline) and the task's difficulty are marginally significant ($r = -0.264$, $p = 0.053$). Also in task 1, the results revealed an inverse correlation between the FAA index and the task completion time ($r = -0.315$, $p < 0.05$).

4 Discussion and Conclusions

This experiment investigated the feasibility of using EEG to distinguish positive and negative emotional states in user experiences. The user experience was significantly different for these three APPS. First, the results of the behavioral data showed YEW and DZD were better than HMB in task completion rate, completion time, and self-report task difficulty. Second, YEW and DZD were better than HMB in usability; their scores on UEQ and SUS were higher than the scores for HMB. Finally, the emotional state after using each APP was also different. The PA after using HMB was lower than for the other two APPS, and the trend of NA was reversed. These results revealed that participants felt more positive emotions and less negative emotions when using DZD and YEW than when using HMB.

No significant difference was found in FAA among the three APPs, but there were some trends that we can discuss. The FAA was positive when using DZD, revealing an approach trend with this APP. This result was consistent with behavioral and subjective evaluation results. The FAAs were around zero when using YEW and HMB and revealed no preference in using these two APPs, even the results of behavior and subjective evaluation showed YEW was better than HMB. It is worth noting that the task difficulty of YEW was greater than for DZD based on the self-report results. The participants felt different moods when using these two APPs, but they could not consciously report this difference. So, one possible explanation is that the EEG results were probably to distinguish the tiny differences in emotional states. The marginally significant correlation between FAA change and task difficulty supported this explanation.

The results from the correlation analysis also indicated a significant relationship between EEG data and subjective evaluation. Previous studies have proven that brainwave activity relates to changes in mental or physical states, that is, the dominance of fast rhythmic activity (beta/gamma) indicates states of high arousal (e.g., reasoning, problem solving) and the dominance of slow rhythmic activity (alpha) indicates a relaxed state [12]. On one hand, a high PA value represents a state of powerfulness, concentration, and joviality [19]. The PA score was negatively correlated with alpha power and the change of alpha power in this experiment. It implies that less concentration accompanies high alpha power. On the other hand, low NA represents a state of peace and calmness [19]. A negative correlation between the change of NA and beta power was found, but the internal mechanism still needs more discussion.

There are some limitations in this research. The emotional arousal levels for the three apps are not high enough. This could be a reason that no significant difference was found in FAA among three APPs. The materials used in marketing and design were complicated, which often arouses people's emotional response in physical or visual ways [10, 11, 16]. However, the interfaces of the three APPs used here were much simple than those in commercial ads. The tasks, such as finding a restaurant or planning a route to a specific location, were too easy to cause a change in emotions. The differences in usability were easy for users to perceive and self-report, but the change of affect caused by the usability levels of the APPs was extremely slight to be captured by the FAA. Other software, which could arouse strong fluctuations, such as game APPs, can be explored in future research.

In sum, using EEG indicators to evaluate the emotional state of users is a feasible tool in human computer interaction. FAA could be used as an index of approach, although we did not reach significant results in this study. The stability of FAA and its scope of application need to be explored in future research.

Acknowledgement. This work was supported by User Experience Lab of China Mobile Research Institute, Science and Technology (S&T) basic work (2009FY110100) and NSF China (31100750, 91124003).

Reference

1. FDIS, I.: 9241-210: 2009 Ergonomics of human system interaction-Part 210: Human-centered design for interactive systems (2009)
2. Hekkert, P.: Design aesthetics: principles of pleasure in design. Psychology Science 48, 157 (2006)
3. Vermeeren, A.P., Law, E.L.-C., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K.: User experience evaluation methods: current state and development needs. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp. 521–530. ACM (2010)
4. Riseberg, J., Klein, J., Fernandez, R., Picard, R.W.: Frustrating the user on purpose: using biosignals in a pilot study to detect the user's emotional state. In: CHI 1998 Conference Summary on Human Factors in Computing Systems, pp. 227–228. ACM (1998)
5. Scheirer, J., Fernandez, R., Klein, J., Picard, R.W.: Frustrating the user on purpose: a step toward building an affective computer. Interacting with Computers 14, 93–118 (2002)
6. Wilson, G.M., Sasse, M.A.: From doing to being: getting closer to the user experience. Interacting with Computers 16, 697–705 (2004)
7. Hu, J., Nakanishi, M., Matsumoto, K.-I., Tagaito, H., Inoue, K., Shima, K., Torii, K.: A method of usability testing by measuring brain waves. IEEE Trans. Software Eng. 25, 474–491 (1999)
8. Li, X., Hu, B., Zhu, T., Yan, J., Zheng, F.: Towards affective learning with an EEG feedback approach. In: Proceedings of the First ACM International Workshop on Multimedia Technologies for Distance Learning, pp. 33–38. ACM (2009)
9. Masaki, H., Ohira, M., Uwano, H., Matsumoto, K.-I.: A Quantitative Evaluation on the Software Use Experience with Electroencephalogram. In: Marcus, A. (ed.) HCII 2011 and DUXU 2011, Part II. LNCS, vol. 6770, pp. 469–477. Springer, Heidelberg (2011)
10. Ohme, R., Matukin, M., Szczurko, T.: Neurophysiology uncovers secrets of TV commercials. Der Markt 49, 133–142 (2010)
11. Ohme, R., Reykowska, D., Wiener, D., Choromanska, A.: Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures. Journal of Neuroscience, Psychology, and Economics 2, 21 (2009)
12. Stickel, C., Fink, J., Holzinger, A.: Enhancing Universal Access – EEG Based Learnability Assessment. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 813–822. Springer, Heidelberg (2007)
13. Briesemeister, B.B., Tamm, S., Heine, A., Jacobs, A.M.: Approach the Good, Withdraw from the Bad—A Review on Frontal Alpha Asymmetry Measures in Applied Psychological Research. Psychology 4, 261–267 (2013)
14. Davidson, R., Schwartz, G., Saron, C., Bennett, J., Goleman, D.: Frontal versus parietal EEG asymmetry during positive and negative affect. Psychophysiology 16, 202–203 (1979)
15. Kline, J.P., Blackhart, G.C., Woodward, K.M., Williams, S.R., Schwartz, G.E.: Anterior electroencephalographic asymmetry changes in elderly women in response to a pleasant and an unpleasant odor. Biological Psychology 52, 241–250 (2000)
16. Park, M.-K., Watanuki, S.: Electroencephalographic responses and subjective evaluation on unpleasantness induced by sanitary napkins. Journal of Physiological Anthropology and Applied Human Science 24, 67–71 (2005)

17. Tomico, O., Mizutani, N., Levy, P., Takahiro, Y., Cho, Y., Yamanaka, T.: Kansei physiological measurements and constructivist psychological explorations for approaching user subjective experience during and after product usage. In: Proceedings of the DESIGN 2008, 10th International Design Conference, pp. 529–536 (2008)
18. Watson, D., Tellegen, A.: Toward a consensual structure of mood. Psychological Bulletin 98, 219 (1985)
19. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of Personality and Social Psychology 54, 1063 (1988)
20. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 63–76. Springer, Heidelberg (2008)
21. Brooke, J.: SUS-A quick and dirty usability scale. Usability Evaluation in Industry 189, 194 (1996)
22. Harmon-Jones, E., Allen, J.J.: Behavioral activation sensitivity and resting frontal EEG asymmetry: covariation of putative indicators related to risk for mood disorders. Journal of Abnormal Psychology 106, 159 (1997)
23. Wheeler, R.E., Davidson, R.J., Tomarken, A.J.: Frontal brain asymmetry and emotional reactivity: A biological substrate of affective style. Psychophysiology 30, 82–89 (1993)