A Framework for Usability Testing using EEG Signals with Emotion Recognition

Arwa Aledaily [1], Sofien Gannouni, Kais Belwafi, Hatim Aboalsamh [2]

[1] University of Hail , Hail, 2240, KSA

[2] King Saud University, Riyadh, 11362, KSA

ABSTRACT

The analysis of emotions has utility in several applications that cross multiple fields, including education, medicine, psychology, software engineering, accessibility in-habitation studies, healthcare, robotics, marketing, and business. Studying emotions can play an essential role in software engineering, particularly in the domain of usability testing. For example, emotions can be used to determine whether a specific software application achieves acceptable levels of user satisfaction. Furthermore, emotions can be used to test product usability and all its aspects. Emotion detection in usability testing is a first-of-its-kind tool that has the potential to improve software production (designing and interaction), thus enabling the ongoing revolution in software development to continue onwards. This work aims to build an original framework for emotion detection using electroencephalography (EEG) brain signals, which is then applied in usability testing as a case study. This will create opportunities to gain an in-depth understanding of user satisfaction in a precise and accurate way, especially when compared to traditional approaches.

Keywords: Brain-computer interface, Emotion Detection, EEG, Usability Testing

INTRODUCTION

A human being is an intelligent creation that is full of secrets. A person is a combination of emotions, feelings, behaviors and actions. Emotions are responses of human passions to objects or events in the environment (Barrett, 2016). Emotion analysis can be used in several applications crossing many fields such as education, learning, medical studies, psychological studies, software engineering, accessibility inhabitation studies, health care, robotics industrials, marketing and business (Barrett, 2016). Specifically, studying emotions can play an essential role in software engineering, precisely in usability testing. Emotions can be used to test a project in order to achieve a higher level of customer satisfaction (Thuering, 2007). Emotions can be used to test product usability and all its aspects. There- fore, this topic will be chosen as a case study in this work. Emotion detection in usability testing is a first-of-its-kind tool to improve software production (designing and interaction) to keep up with the revolution of software development. This work aims to build an original framework for emotion detection using EEG brain signals and to apply it in usability testing as a case study. This will lead to deep understanding of customer satisfaction precisely and accurately, more than the traditional way. The structure of this paper will be as follows: Section 2 provides a brief background of the major aspects in this work. Then, Section 3 is about the proposed framework of using emotion detection during usability testing. Followed by a discussion in Section 4. Finally, the conclusion is given in Section 5, followed by the list of references.

BACKGROUND

Emotion Detection using VAD model

One way to study emotions is to describe them as dimensions. The constructionist model adopts this idea and describes emotions under three dimensions: valence, arousal and dominance (called the VAD model). In this model, emotions are studied based on broader aspects instead of a specific brain structure and pattern for each emotion (Wyczesany, 2015). The three dimensions of the VAD model are defined as follows:

- Valence: This is an internal state involving pleasant and unpleasant feelings. This dimension ranges from a positive level to a negative level.

- Arousal: This is an internal state of excitement (active) or calmness (inactive). It ranges from low arousal (calmness) to high arousal (excitement).
- Dominance: This is the feeling of control during an emotional experience. It ranges from high dominance (high control, such as anger) to low dominance (weak control, such as fear).

EEG measures the electrical activity of the brain using electrodes on the scalp (Wyczesany, 2015). EEG is used in emotion classification. Many features are extracted from EEG signals. It provides a massive amount of information about activity in different parts of the brain. The feature extraction process is always a challenge in this field. All previous researchers tried to choose a fixed number of EEG electrodes to avoid the feature dimensionality problem. This limited number of fixed groups of electrodes differs from one study to another. Neuro- science experiments on the brain show that all regions (or almost all) exhibit a "firing" activity during emotional experiences (Murugappan, 2010). Thus, there is a need to study emotions subjectively as a prerequisite step to an objective conclusion. Based on the already mentioned gaps, in this work, we plan to apply adaptive emotion processing of the brain activity, then use the approach in usability testing.

Usability Testing and VAD Model

Usability testing is one of the most important issues in software testing. Usability testing is a method of evaluation that involves testing user interactions with a product or system to improve its design (Lee, 1999). The goal of usability testing is to test the ease-of-use of software, and to improve the process of design and development (Dumas,1999).

Several methods have been developed to study emotions in the context of usability testing, including traditional approaches (e.g., surveys, technical interviews, and questionnaires), and well as more novel approaches (e.g., physiological measures such as eye-tracking, heart rate monitoring, electromyography (EMG), and EEG) (Thuering, 2007) (Saariluoma, 2014). However, the application of brain-computer interface (BCI) systems in usability testing remains unexplored. This is particularly vital to recognize because the traditional methods that are applied in usability testing lack the ability to distinguish accurately between emotional states. It is important to understand the level of satisfaction in each step of testing. As a result, using brain signals can help in identifying the emotional state in every period of testing time (Puwakpitiyage , 2019) (Amaral, 2013).

Valence and arousal levels can be used as indicators to predict the satisfaction of a user during their experience with a system. These levels can indicate whether a system causes boredom, stress, anger, or relief, as well as many other emotional states (Ward, 2013). This is useful in achieving the functional specification of software, as well as user needs.

In addition, researchers, including (Jokela, 2013), have reported that various usability testing factors are mapped with different emotions depending on biological brain signal readings, including the following:

- The first impression factor is linked directly with excitement and the absence of emotions such as boredom or disgust. It is a crucial factor for measuring user satisfaction with webpages.
- The task-based testing method is directly linked to frustration and empowerment emotions. It detects the emotional status of the user when performing a specific task depending on a scenario.
- Overall emotions are used to test the system freely in the absence of any specific scenario. They are important for general measurements of usability and satisfaction. Engagement and discouragement emotions are important to differentiate between in order to achieve this goal.

PROPOSED FRAMEWORK

A generic framework is proposed in this section for application in the usability testing of any type of software. The proposed framework relies on the detection of a user's emotions while they perform tasks related to a given software system in serial sessions. Building this framework involves the following phases: preparation, training, testing, and reporting the usability aspects and linking them with the recognized emotions (see Figure 1).

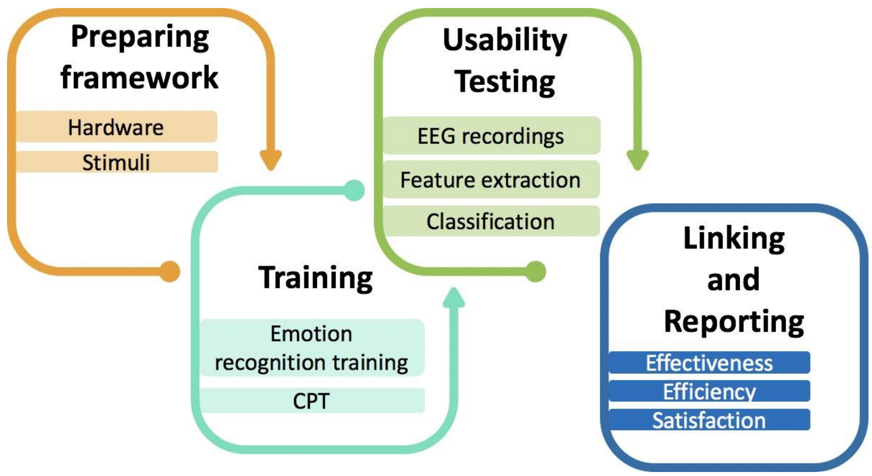

Figure. 1 Proposed framework for usability testing with emotion recognition

Preparation

The proposed framework is an interface of different choices of web-based systems. EEG acquisition hardware is required to record brain activity in the training phase, and then in usability testing. EEG signals are processed, extracted, and then classified. Several participants will be encouraged to help in this work. They all have a corresponding technical background, similar ages, and a similar cultural background. In addition, they are all right-handed. The two main factors involved in the experiment are discussed in the following sections.

Hardware

Hardware capable of recording and transmitting EEG signals is used in the proposed framework. Any EEG cap fits the framework's requirements (high- performance device and high-quality signals) can be used in this stage.

Emotion Stimulus

Two suggestions are given regarding the emotional stimulus in this work. It will be applied for participants with approximately the same age, experience, and cultural background, all of whom are right-handed.

- DEAP dataset: 40 videos will be used in the experiment, under the same setting as described in (Koelstra, 2011).
- Arabic culture dataset: Several studies (Alghowinem ,2014) (Al-Mutairi, 2015) have suggested that a new dataset should be constructed for emotion recognition that is suitable for Arabic culture and regulations. It consists of video stimulus from Arabic clips that link with different emotions.

Training Phase (Emotion Recognition Training)

This phase can be divided into two parts: emotion recognition training and continuous performance test (CPT).

- Emotion recognition training: This phase aims to train the proposed algorithm on the brain activities of the subjects who are enrolled in the experiment. The videos used in the chosen dataset will be applied for the subjects in this phase under the appropriate setting. In this stage, either the locationist or VAD models can be applied. In turn, a self-assessment questionnaire can be used to specify the emotion of the participant (either discrete status or valence-arousal-dominance status). In the last one, we can follow the same setting as described in (Koelstra, 2011). Furthermore, valence and arousal levels can be specified using the validated formulas given in (Ramirez, 2012). The later criteria should be studied more with the proposed method because they involve the selection of specific electrodes from specific frequency bands (Ramirez, 2015). However, it is applicable.

- Continuous performance test (CPT): This method assists in specifying the response time for each subject, which reflects the first impression period. It is an important factor for recognizing the emotion that arises in the first impression period, and so it plays an essential role in studying the satisfaction of each user during usability testing (Gunawan, 2017). CPT measures the response of clicking the 'space' key when any character appears on the screen, except the 'X' character. The response time is the first impression time for a given user (Gunawan, 2017).

Testing Phase (Usability Testing)

In this phase, the user will choose a website to test its usability. The procedure is divided into multiple 'to-do' tasks. Each task has a time that it begins and a time that it ends, and the user attempts to complete the task in the defined period. This phase can be divided into two implicit sessions:

- First-impression testing: As we previously measured CPT for each subject, this period is analyzed later to determine the user's first impression of the website at the beginning, as well as for each task.

- The rest of the time is the EEG recording to facilitate emotion recognition for a particular task.

EEG Feature Extraction and Emotion Classification

The proposed emotion recognition system, as outlined in our previous works (Gannouni, 2021) (Gannouni, 2020), will be applied in this stage. Feature extraction at this point will follow the same criteria that were proposed previously. The EEG signals are first pre-processed, and EEG electrodes are chosen and tuned. Finally, the feature vector is constructed from the values of the ERD/ERS of the chosen electrodes. The feature extraction method in (Gannouni, 2021) (Gannouni, 2020) is a novel approach. The EEG sources are chosen adaptively depends on the brain activity of the subject. As well as different classification methods discussed in (Gannouni, 2021) (Gannouni, 2020) will be applied separately at this stage. There are two types of final emotion labels: VAD model labels or distinct labels of emotional state. In turn, the results will be evaluated to determine which classifier method outperformed the others.

Adaptive emotion detection: A previous work

The work in (Gannouni, 2021) (Gannouni, 2020) is a novel feature selection approach for emotion recogni tion. We demonstrated how the activity of each brain region helps in choosing the optimal EEG channels that can be used for the accurate detection of emotions. These electrodes become the feature resources. The proposed method represents an adaptive method for selecting feature sources independently for each person. This is a new exploration in this area. Dissimilar to the proposed system, prior studies have used the same EEG sources for all subjects. We can summarize the results of the feature extraction and the classification in the mentioned works (Gannouni, 2021) (Gannouni,

2020) as follows

Table 1 Proposed Methods in (Gannouni, 2021) (Gannouni, 2020) that forms the base of this framework

Emotion model	Number of classes	Feature extraction algorithm	Classifier	AVG accuracy (%)
VAD model	3 dimensions	ERD/ERS	QDC	82.35%(V)
				79.95%(A)
				71.14%(D)
			RNN	85.88%(V)
				87.71%(A)
				86.63%(D)
Discrete emotion	9 emotions	ERD/ERS	QDC	80.23%
			RNN	84.23%
		ZTW and NGD	QDC	87.05%
		ZTW and NGD	RNN	89.33%
		asymmetrical ERD/ERS	QDC	71.98%

Linking and Reporting

This is the final phase in the proposed framework. The emotions will be linked to the usability testing aspects: of effectiveness, efficiency, and satisfaction, and a report will be generated. The details associated with linking emotions to aspects of usability testing are discussed in section 4.

USABILITY TESTING ASPECTS IN THE PROPOSED FRAMEWORK: A DISCUSSION

Related to usability testing factors, we linked the proposed framework processes with these factors in the following way:

- First impression factor: The classification result of the EEG period in CPT will allow the user's first impression to be determined. This helps to improve design aspects such as use of color, alignment, typography, and so on.

- Task-based testing factor: This is mapped with the classification results of each module. The feature vector will be constructed for each module record- ing. The emotion label of the classifier is the subject's emotional state during this task. This factor helps in improving the task design, and it has a direct impact on usability.

- Overall emotions factor: This factor is shown in the continuous method of the framework. It is an important factor to facilitate a measurement of usability when we divide the recording of each module and the satisfaction in general, if we take the results of the entire continuous period as a feature vector.

Furthermore, the approach to testing is linked to all aspects of usability testing, as shown in the following:

- Effectiveness: In the ‘task-based test’, the level of each dimension in VAD is an indicator of the quality of the design and the function of the interface (Ward, 2003). The correlation measurement is an accurate indicator of the emotion and the task/function of the system during the user experience, which reflects effectiveness (Stickel, 2009).

- Efficiency: Based on the definition of efficiency given in (Jokela, 2003), groups of emotions (e.g., calm or miserable) are linked to measurements of resource consumption. The type and level of correlation measurement leads to the level of system efficiency (Jokela, 2003).

- Satisfaction: The appearance of emotions that are either pleasant or unpleasant during the testing session can be used to draw an inference about user satisfaction. It is an indicator about each task, function, and the overall system. Scenarios such as the first impression test and free interaction test [20] will be chosen in this step. Certain measurements are suggested in this point to decide on the satisfaction level, including the percentage change relative to a baseline state (i.e., neutral emotion or calm), and different types of correlation (Jokela, 2003) (Kolakowska, 2014) (Marcus, 2011).

Regarding the VAD model, the dominance dimension has not been explored in any depth in the context of usability testing (Partala, 2009). Furthermore, there are no significant correlations in usability testing (Sonderegger, 2013) or in acceptance technology in general (Kulviwat, 2007). This aspect should be examined and, in particular, researchers should seek to uncover new facts about the relationship between dominance and usability testing. Another open question relates to the issue of how to map this dimension to usability testing factors. However, this does not omit the valuable role of dominance in determining an individual’s emotional state during usability testing.

Finally, for the purpose of evaluating the reliability of the proposed usability testing framework, its accuracy and performance will be compared with traditional approaches in usability testing.

CONCLUSIONS

This paper outlined an original usability testing framework that relies on the emotion recognition system developed in our previous work (Gannouni, 2021) (Gannouni, 2020). The framework seeks to derive objective, reliable, and valid measurements in usability testing that are linked to a user's emotional states. Two approaches – discrete and continuous mode – were applied for the framework, and while each is marked by consequential differences, each mode serves well-defined goals in usability testing. Crucially, by leveraging the power and, most notably, the accuracy of the proposed emotion classification system, we expect that the proposal given in this chapter will serve as a viable foundation for an efficient, reliable, and automatic usability testing framework.

REFERENCES

Barrett, L.F., Lewis, M., Haviland-Jones, J.M. (2016) "Handbook of Emotions. Guilford Publications"

Thuering, M., Mahlke, S. (2007). "Usability, aesthetics and emotions in human–technology interaction." International Journal of Psychology - INT J PSYCHOL 42, 253–264

Wyczesany, M., Ligeza, T.S. (2015) "Towards a constructionist approach to emotions: verification of the three-dimensional model of affect with eeg- independent component analysis." Experimental brain research 233(3), 723–733

Murugappan, M., Ramachandran, N., Sazali, Y., et al. (2010) "Classification of human emotion from eeg using discrete wavelet transform." Journal of biomedical science and engineering 3(04), 390

Lee, S.H. (1999) "Usability testing for developing effective interactive multimedia software: Concepts, dimensions, and procedures. " Journal of Educational Technology and Society 2(2)

Dumas, J.S., Dumas, J.S., Redish, J. (1999) " A Practical Guide to Usability Testing. Intellect books."

Saariluoma, P., Jokinen, J.P. (2014) "Emotional dimensions of user experience: A user psychological analysis." International Journal of Human-Computer Interaction 30(4), 303–320

Puwakpitiyage, C., Rao, V., Azizi, M., Tee, W., Murugesan, R., Hamzah, M. (2019) " A proposed web based real time brain computer interface (bci) system for usability testing." International Journal of Online Engineering 15(8)

do Amaral, V., Ferreira, L.A., Aquino, P.T., de Castro, M.C.F. (2013) "Eeg sig- nal classification in usability experiments." In: 2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), pp. 1–5. IEEE

Ward, R.D., Marsden, P.H. (2003) "Physiological responses to different web page designs." International Journal of Human-Computer Studies 59(1-2), 199– 212

Jokela, T., Iivari, N., Matero, J., Karukka, M. (2003) "The standard of user- centered design and the standard definition of usability: analyzing iso 13407 against iso 9241-11." In: Proceedings of the Latin American Confer- ence on Human-computer Interaction, pp. 53–60

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I. (2011)

"Deap: A database for emotion analy- sis; using physiological signals. " IEEE transactions on affective computing 3(1), 18–31

Alghowinem, S., Alghuwinem, S., Alshehri, M., Al-Wabil, A., Goecke, R., Wagner, M. (2014) " Design of an emotion elicitation framework for arabic speakers." In: International Conference on Human-Computer Interaction, pp. 717–728. Springer

Al-Mutairi, N., Alghowinem, S., Al-Wabil, A. (2015) “Comparison of user responses to english and arabic emotion elicitation video clips.” In: International Conference on Cross-Cultural Design, pp. 141–152 Springer

Ramirez, R., Vamvakousis, Z. (2012) “Detecting emotion from eeg signals using the emotive epoc device.” In: International Conference on Brain Informat- ics, pp. 175–184. Springer

Gunawan, F.E., Wanandi, K., Soewito, B., Candra, S., Sekishita, N. (2017) “Detecting the early drop of attention using eeg signal.” In: 2017 4th Inter- national Conference on Electrical Engineering, Computer Science and Informatics (EECSI), pp. 1–6 . IEEE

Gannouni, S., Aledaily, A., Belwafi, K., Aboalsamh, H. (2021) “Emotion detection using electroencephalography signals and a zero-time windowing-based epoch estimation and relevant electrode identification.” Scientific Reports 11(1), 1–17

Gannouni, S., Aledaily, A., Belwafi, K., Aboalsamh, H. (2020) “Adaptive emotion detection using the valence-arousal-dominance model and eeg brain rhythmic activity changes in relevant brain lobes.” IEEE Access 8, 67444–67455

Stickel, C., Ebner, M., Steinbach-Nordmann, S., Searle, G., Holzinger, A. (2009) “Emotion detection: application of the valence arousal space for rapid biological usability testing to enhance universal access.” In: International Conference on Universal Access in Human-Computer Interaction, pp. 615–624. Springer

Kolakowska, A., Landowska, A., Szwoch, M., Szwoch, W., Wrobel, M.R. (2014) “Emotion recognition and its applications. In: Human-Computer Systems Interaction: Backgrounds and Applications” 3, pp. 51–62. Springer

Marcus, A. (2011) “Design, User Experience, and Usability. Theory, Methods, Tools and Practice”. First International Conference, DUXU 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings vol. 6770. Springer

Partala, T., Kangaskorte, R. (2009) “The combined walkthrough: measuring behavioral, affective, and cognitive information in usability testing.” Journal of Usability Studies 5(1), 21–33

Sonderegger, A., Sauer, J. (2013) “The influence of socio-cultural background and product value in usability testing.” Applied ergonomics 44(3), 341–349

Kulviwat, S., Bruner II, G.C., Kumar, A., Nasco, S.A., Clark, T. (2007) “Toward a unified theory of consumer acceptance technology.” Psychology & Marketing 24(12), 1059–1084.