ISSN:0975-9646

Affective Crowdsourcing Applied to Usability Testing

Victor H. M. Gomide [1] , Pedro A. Valle[1], José O. Ferreira[1], José R. G. Barbosa[1], Adson F. da Rocha[2] and Talles M. G. de A. Barbosa[1]

[1]*Department of Computing – PUC Goias, Goiania, Brazil*
[2]*UnB-Gama Faculty, University of Brasilia, Gama, Brazil.*

***Abstract*—Usability tests are very important to validate and improve software systems and software development processes. Nowadays, there are different methodologies to software usability evaluation. However, the most effective are based on real user's feedback although these methodologies are time-consuming and very expensive because depend on analysis of many human computer interactions. In this paper two hypotheses are launched. First, usability tests could be applied remotely based on crowdsourcing platforms. Second, outliers could be detected based on user's emotional behaviour. In order to investigate these hypotheses an affective usability evaluation process has been developed and supported by a low cost software interface called Hesitation Detector (HD). An HD is an affective component programmed to recognize events related to user's emotional states. In this paper a deterministic automata is presented with purpose of automatic detection of hesitation, which qualifies it to be applied in different computational platforms, including tablets and smartphones. Besides, it not requires obstructive sensors because user's movements are gathered and processed by means traditional interfaces. In order to validate and customize the proposed automata some usability tests were performed together with biomedical signals processing. Preliminary results reveal evidences that affective crowdsourcing can be a promising alternative in order to evaluated users satisfaction and their reactions remotely across the Internet.**

Keywords-Affective Systems; Affective Crowdsourcing; Usability Testing.

I. INTRODUCTION

Software verification and validation is one of most difficult activities related to software development. It can represent more than 60% of total costs particularly when agile methodologies are embraced. Usability tests are critical activities related to software testing because they are very influenced by participant's emotions. Moreover, financial costs regarding to usability tests are proportional to how many participants are involved. Thus, the software industry spends billions of dollars on software testing annually, but failures can still be easily found, maybe because the testing process is incorrect or participants' usability reports are not validated before statistical analysis. In other words, outliers were not detected neither removed.

On the other hand, Crowdsourcing is an emerging phenomenon used to get benefits through online community members who are unknown to companies in general [1]. According to Morris human computation work that involves large groups of people will be considered a form of crowdsourcing [15]. Using crowdsourcing to obtain relevance judgments presents challenges, however. Several studies found spam among the obtained crowdsourcing results, with some reporting as much as 50 percent spam [6]. Regarding to usability tests, crowdsourcing is a new approach supported by Internet to increase the number of participants without increasing costs because different users from different parts of the world can be invited on social networks to collaborate voluntarily on projects. However, a great challenge to include crowdsourcing into software usability testing is absence of efficient mechanisms to validate and verify each participant's responses presented in usability's reports.

This paper proposes a crowdsourcing software usability testing method supported by an events detection algorithm. It enables traditional Human-Computer Interfaces, such as an ordinary mouse or touch-based user interfaces to recognize users emotions by processing user's actions from mouse movements or touch events. Consequently, it can be useful to disqualify non-conforming usability reports. Non-conforming tests are very correlated with the increasing of mouse movements because it is highly correlated with hesitation. When hesitation frequency increases is possible to suggest that users do not comprehend usability questions or they are not be able to respond those questions. Thus, those usability questionnaires can be considered outliers. In order to validate and customize hesitation detection automata a testing process was proposed and applied. User's data sets were captured from mouse movements and correlated with users data sets obtained from Galvanic Skin Response (GSR) sensor. GSR is the most common used electrophysiological marker to analyse and quantify emotional users states. In this work non-obstructive GSR hardware was developed as a part of research methodology's tools used in the study. The success of this approach can lead to a paradigm shift, since software testing activity will be executed outside the software houses reducing costs while maintaining the same software quality.

II. MATERIALS AND METHODS

A. Usability Testing and Crowdsourcing

ISO (9241 Part 11) definition of usability is "the extent to which a product can be used by specified users to achieve specified goals with efficiency, effectiveness, and satisfaction in a specified context of use." Efficiency, effectiveness, and satisfaction are all measurable – some through observing performance of humans interacting with the technology, some through gathering data about the perception of the humans doing the interacting. Considering the context of use – that larger experience – is important to usability [9].

On the other hand, Crowdsourcing is one method for learning the relevant documents to each query in the test set [6], such as researchers can use data from social websites dedicated to patients to develop a deeper understanding of disease processes and devise more effective treatments [7]. However, the quality of relevance learned through crowdsourcing can be questionable, because it uses workers of unknown quality with possible spammers among them [6]. To crowdsource a task, its owner, also called the *requester*, submits the task to a crowdsourcing platform. People who can accomplish the task, called *workers*, can choose to work on it and devise solutions. Workers then submit these *contributions* to the requester via the crowdsourcing platform. The requester assesses the posted contributions' quality and might reward those workers whose contributions have been accepted. This reward can be monetary, material, psychological, and so on [4].

Traditionals usability testing are expensive and time-consuming processes. They require many activities such as *use case definitions* as a part of test planning; Environmental preparations for configuring physical space for applying all users tests; Selection and *convocation* of *participants* according to use case profile; *Executions* of *tests* should be prepared for recording and authenticating all participants in order to exclude outliers (or spammers), and finally analysis of reports. Fig. 1 shows a traditional usability testing process compared to a crowdsourced usability testing methodology. Crowdsourcing platforms are able to automatize three important activities, reducing time and costs.

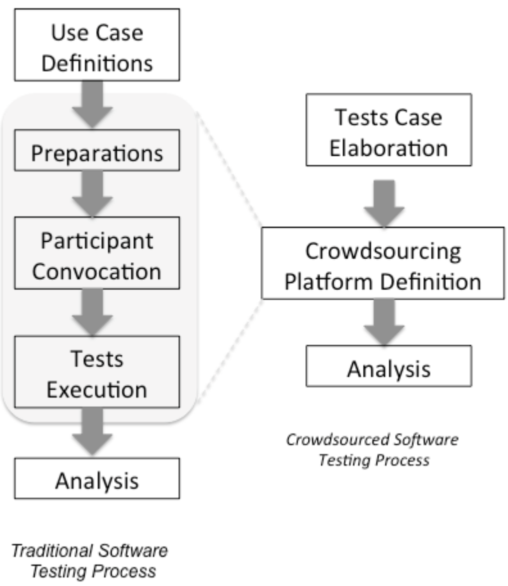

Fig. 1. Usability software testing approaches: traditional software testing process versus crowdsourced software testing process.

B. Affective Computing hypothesis

Emotions play an important role in human experience and it may influence decisions, perception, learning and communication. A great challenge in the human-computer interaction area refers to human adaptation instead of machine adaption to satisfy human necessities [2]. Affective computers equipped with cameras, microphones, physiological sensors, and sophisticated pattern recognition tools, can begin to recognize physiological components of emotions, and to infer the likely emotional state underlying these components [3].

Robert Morris at MIT Media Labs coined the expression Affective Crowdsourcing (in 2009) to suggest that "affective computing technologies can be used to measure affective states in real-time and provide feedback to coordinating the crowd" [15]. In 2013, a new study showed how affectivity could influence creative performance in mircrotask crowdsourcing environments, examining two affective techniques to boost creativity on crowdsourcing platforms: affective priming and affective pre-screening [16].

This paper argues that biofeedback related to emotions can be processed and analyzed to filter outliers (commonly called spammers) on usability tests. Outliers are most important artifacts to disqualify market research, which includes social and opinion research, and it is highly correlated with hesitation. Thus, this new functionality could be easily embedded in crowdsourcing platforms by traditional users interfaces.

III. AFFECTIVE CROWDSOURCING APPLIED TO USABILITY TESTING

A. Hesitation Detector Automata.

In order to implement and automatic hesitation detector three rules were defined: *i)* if *the duration time* related to each test participant was greater than *computed average time* for all participants the first evidence was encountered; *ii)* if *the amount of mouse clicking per unit time* was greater than that encountered in participants population avg. the second evidence was found, and *iii)* if the *click duration avg.* (computed as: release time – press time) was greater then average encountered in population avg. than the third evidence was faced. Fig. 2 presents the state diagram related to proposed automata.

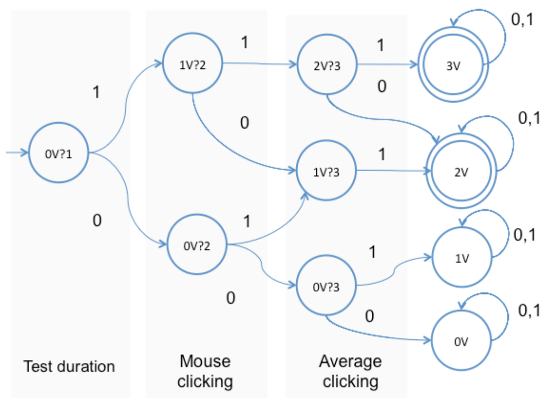

Fig. 2. Hesitation detector automata.

The vertices names were defined according to accounted evidences. For example, "1V?3" means one evidence in third stage. The edges are named "1" for true and "0" when the evidence was not perceived. Mouse pressure pattern detection has been used during episodes likely to be frustrating to the user [12] while the number of clicks, the duration of each click and distance traveled by the mouse has been investigated in [13], [14].

B. Hesitation Markers.

The galvanic skin response (GSR) is a simple, useful, and reproducible electrophysiological technique to investigate sympathetic nervous system function. In history GSR is also known as, or closely related to, the psychogalvanic reflex (PGR), electrodermal response (EDR), skin conductance response (SCR) and sympathetic skin response. Physically GSR is a change in the electrical properties of the skin in response to different kinds of stimuli [11]. In measurements changes in the voltage measured from the surface of the skin are recorded. The main parameters of GSR, such as basis threshold, peaks, or frequency variation, vary enormously among different individuals, and thus, no general features can be extracted from GSR signals. Therefore, the parameters extracted from GSR signals are strongly related to each individual. However, despite the differences among individuals, GSR signal is not distinctive enough to identify an individual in terms of biometrics [10].

In order to verify and validate the hesitation detector automata presented in Fig. 2, GSR acquisition hardware (see Fig 3) was developed. It was required to correlate obtained data filtered by hesitation detector automata with physiological signal.

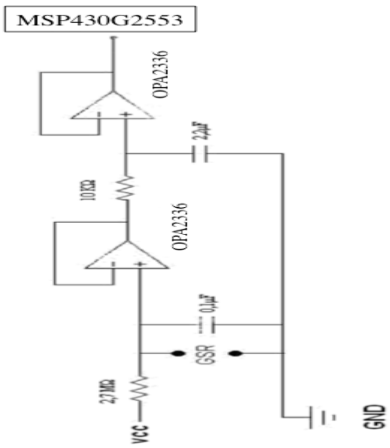

Fig. 3. GSR sensor.

This circuit works as voltage divisor comparing skin resistance with known resistance. Buffers are implemented by two ampops and two analog filters were used to remove undesired signals artifacts. On Fig. 4 an affective mouse prototype is showed. It was built to detect GSR events gathered from GSR Sensor presented in Fig. 3.

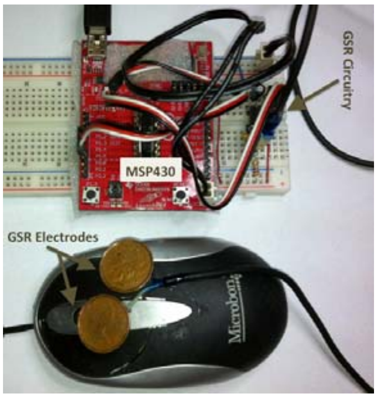

Fig. 4. Affective mouse detector.

IV. CASE STUDY: ECOMMERCE USER INTERFACE

A. Usability Evaluation Process.

In order to verify the proposed affective crowdsourcing usability an evaluation process was defined and applied. It consists of five activities that should be executed sequentially according to Fig. 5.

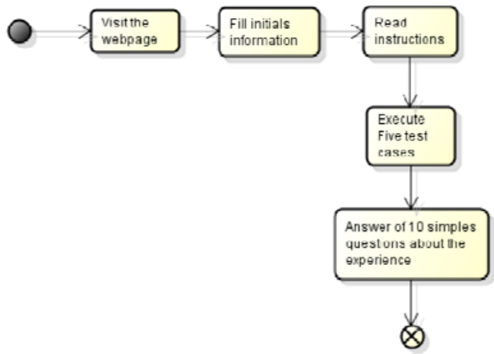

Fig. 5. Test case process.

The task "Visit the webpage" (Fig. 5) refers to the interface through which the participants should access and contribute to the usability tests. For the purpose of platform-independence a Web page related to ecommerce application was built. According to Kittur [5] a user-friendly interface can attract more participants (workers in crowdsourcing jargon) and increase the chance of a high-quality outcome. Fig. 6 shows an instance of usability testing questionnaire (desktop version) about users experiences after he manipulated the proposed web page during usability test. Fig. 7 shows the same web page presented in Fig. 6, but a Tablet version. Desktop and Tablet (Acer Iconia Tab A500) platforms were tested. HTML and JavaScript technologies were used for prototyping.

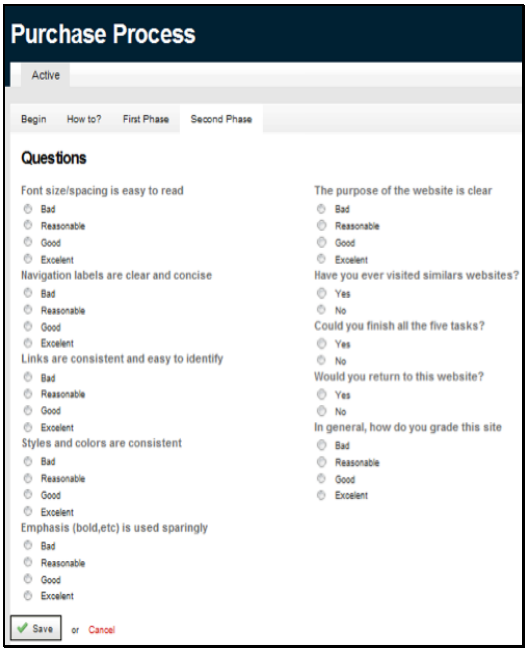

Fig. 6. Usability testing questionnaire. *(Desktop).*

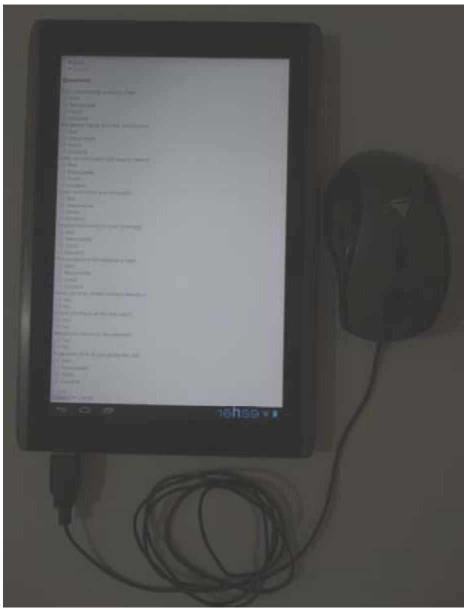

Fig. 7. Usability testing questionnaire. *(Tablet)*.

Preliminarily, the ecommerce system usability case has been evaluated with two groups: *i)* 50 computer science students (23 ± 2 years); *ii)* 15 computer-engineering students (22 ± 2 years) at PUC Goias. Both groups were selected for purpose of reducing misunderstandings about software manipulating and technologies. Users should be concentrated in software functionalities, labels and texts.

B. Preliminary Results

In Fig. 8 is presented a specific test case when recorded webcam video and GSR were correlated with mouse click and test duration obtained from mouse data input.

GSR and Webcam show evidences of hesitation when the participant 11 is faced with a specific situation, when he is using software interface. Based on events observed on hesitation markers (signals) gathered such as Fig. 8 the hesitation detector (Fig. 2) was defined and tested. It is able to automatically detect hesitation events comparing test duration, mouse clicking and the clicking average with data obtained from population average. In many situations this simple algorithm has been effectual. Table 1 and Table 2 summarize the encountered evidences.

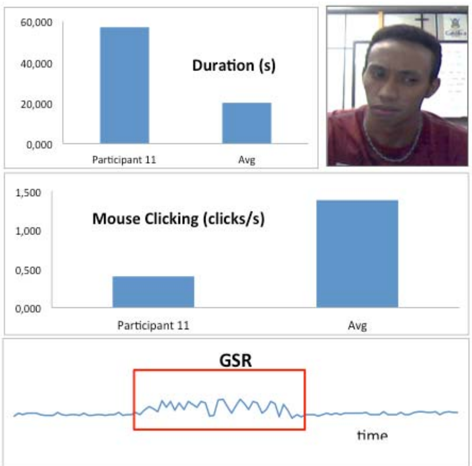

Fig. 8. The hesitation markers during a specific test case.

TABLE I. AUTOMATIC HESITATION DETECTION IN USABILITY TESTS (COMPUTER SCIENCE STUDENTS)

Webcam and Automata Verification		
Webcam AND Automata	10	20%
Webcam OR Automata	22	44%
NOT (Webcam AND Automata)	18	36%
Total	50	100%

TABLE II. AUTOMATIC HESITATION DETECTION IN USABILITY TESTS (COMPUTER-ENGINEERING STUDENTS)

Webcam, GSR and Automata Verification		
Webcam AND Automata AND GSR	3	20%
Webcam OR Automata OR GSR	7	47%
NOT (Webcam AND Automata AND GSR)	5	33%
Total	15	100%

V. CONCLUSION

As Internet and mobile technologies continue to advance, crowdsourcing can help organizations increase productivity, leverage an external (skilled) workforce in addition to a core workforce, reduce training costs, and improve core and support processes for both public and private sectors [4]. Besides, it provides unprecedented new opportunities for people to share their knowledge and observations with the rest of the world. And because of their relatively high cost and low scale, more conventional, centralized means of data collection are being displaced by crowdsourced alternatives [8].

In this paper a usability testing process was proposed. It has been developed for reducing time and costs related to traditional usability tests. Besides, it can support to a paradigm shifting because software industry has used agile methodologies in many software projects more and more.

Future works will be oriented to improve hesitation detection algorithm specially to treat uncertain situations. For these purposes some ensembles have been investigated and tested specially those combining Hidden Markov Models and metaheuristics. Besides, more usability tests should be executed in order to evaluate promising technologies such as smart watches and smart glasses.

ACKNOWLEDGEMENT

This study was financial supported by FAPEG *(in portuguese, Fundação de Amparo à Pesquisa de Goiás)* at Pontifical Catholic University of Goiás.

We would like to thank Jean Leles and Pedro Bacchini for collaborating on this work.

REFERENCES

[1] M. Houssain. "Crowdsourcing: Activities, incentives and users' motivations to participate". P*roc. of conference on Innovation Management and Technology Research (ICIMTR), 2012 International*, IEEE, Malacca, Malaysia, 2012.

[2] M. Schröder. Emotion Markup Language (EmotionML) 1.0, 2 W3C. Available: http://www.w3.org/.

[3] R. Picard. Affective Computing. MIT Press, 1997.

[4] M. Allahbakhsh et al. "Quality Control in Crouwdsourcing Systems: Issues and Directions". IEEE Internet Computing, March/April, 2013.

[5] A. Kittur, E. Chi, and B. Suh, "Crowd- sourcing User Studies with Mechani- cal Turk," *Proc. 26th Ann. SIGCHI Conf.* Human Factors in Computing Systems, ACM, 2008, pp. 453–456.

[6] J. Vuurens and A. Vries. "Obtaining High-Quality Relevance Judgments Using Crowdsourcing". IEEE Internet Computing, September/october, 2012.

[7] J. Riedl and E. Riedl. "Crowdsourcing Medical Research". IEEE Computer, January, 2013.

[8] P. L. Cox. "Truth in Crowdsourcing". IEEE Security and Privacy, v.9, issue 5, 2011, pp. 74-76.

[9] D. Chisnell. "Usability Testing: Taking The Experience into Account". IEEE Instrumentation & Measurement Magazine, v. 13, Issue 2, 2010, pp. 13-15.

[10] S. A. Sierra et al. "A Stress-detection System Based on Physiological signals and Fuzzy Logic". IEEE Transactions on Industrial Eletronics, v. 58, n. 10, october, 2011.

[11] P. M. Tarvainen et al. "Principal Component Analysys of Galvanic Skin Responses". *Proc. Of Proceedingsof the 22"dAnnual EMBS International Conference*, July 23-28,2000, Chicago IL.

[12] Q. Yuan Q. and Picard R. W. "Context-sensitive Bayesian classifiers and application to mouse pressure pattern classification". Proc of. IEEE Intenational Conference on Pattern Recognition, v. 3, 2002, pp. 448-451.

[13] J. Azcarraga et. "Use of Personality Profile in Predicting Academic Emotion based on Brainwaves Signals and Mouse Behavior". Proc. of Third International Conference on Knowledge and Systems Engineering, 2011.

[14] A. Kaklauskas et al. "Biometric Mouse Intelligent System for Student's Emotional and Examination Process Analysis". *Proc of Ninth IEEE International Conference on Advanced Learning Technologies*, 2009.

[15] R. Morris "Crowdsourcing Workshop: The Emergence of Affective Crowdsourcing". *Proc of Ninth IEEE International Conference on Advanced Learning Technologies*, 2009.

[16] R. Morris, M. Dontcheva, A. Finkelstein and E. Gerber. " Affect and Creative Performance on Crowdsourcing Platforms". *Proc of Humaine Association Conference on Affective Computing and Intelligent*, 2013, pp.67-72.