

Emotions Logging in Automated Usability Tests for Mobile Devices

Jackson Feijó Filho^(✉), Wilson Prata, and Thiago Valle

Nokia Technology Institute, Av. Torquato Tapajós, 7200 - Col. Terra Nova,
Manaus, AM 69093-415, Brazil
{jackson.feijo,wilson.prata,thiago.valle}@indt.org.br

Abstract. This work proposes the use of a system to perform emotions logging in automated usability tests for mobile devices. Our goal is to efficiently, easily and cost-effectively assess the users' affective state by evaluating their expressive reactions during a mobile software usability evaluation process. These reactions are collected using the front camera on mobile devices. The analysis of three different emotions - happiness, surprise and anger – and two “emotional events” – spontaneous smile and gazing away from screen – performed through server software. This automated test generates a graphical log report, timing (a) current application page (b) user events e.g. tap (c) emotions levels e.g. level of happiness and finally (d) emotional events e.g. smiling or looking away from screen.

1 Introduction

Due to the present saturation in the mobile software market, consumers frequently find themselves unable to decide which application to acquire, considering that many of them have the very same functional features. It is very likely they will prefer the application that presents their functionalities in the most usable and efficient manner [9].

In [8] Harty discuss how many organizations do not perform any usability or accessibility testing. It's seen as too expensive, too specialized, or something to address after testing all the “functionality” (which is usually prioritized because of time and other resource constraints). For these organizations, good test automation can be of great benefit.

Usability testing tends to be time consuming and hard to scale when it requires human observation of the people using the software being measured [8].

Most of the usability methodologies (e.g. usability inspection, heuristics, etc.) are both applicable to desktop as well as to mobile software, it is more difficult for mobile context to achieve relevant results with conventional assessment methods. The reason is that the emulation of real-world use during a laboratory based evaluation is only feasible for a precisely defined user context. Therefore, due to physical restrictions, it is difficult to generalize from quickly changing and possibly strongly varying user context [5].

The evaluation of usability in mobile software provides valuable measures about the quality of these applications, which assists designers and developers in identifying opportunities of improvement. But analyzing the usability of mobile user interfaces can be a tiresome assignment. It might be extensive and require expert evaluation

techniques such as cognitive walkthroughs or heuristic evaluations, not to mention often expensive usability lab equipment.

Relevant and recent work has been published concerning tools for low-cost, automated usability tests for mobile devices. In [1] such tools have been reported to aid small development teams to perform fairly accurate suggestions on user interface improvements. However, these tools do not consider emotional feedback of users towards mobile software.

1.1 Emotions and Usability

Emotional feedback is a significant aspect in user experience that chronically goes unmeasured in several user-centered design projects. [2] Human emotions are indispensable to understanding users, as these can enable the increase of persistence and strengthen interest in a subject or task. The examination of this affective aspect through well-known empirical user-centered design methods supports software creators in engaging and motivating users while using their systems [3]. The gathering of emotional cues will provide an additional layer of analysis for collecting user data, augmenting common evaluation methods and resulting in a more accurate understanding of the user's experience.

1.2 Automated Tests and Unsupervised Field Evaluations

Furthermore, it is essential to mention the importance of automated tests, while being executed through and for mobile devices. In contrast to desktop applications or web sites, mobile applications have to compete with external stimuli, as users might not be sitting in front of a screen for considerable amounts of time [4]. Due to the very nature of mobility in this scenario, in a real-world context, users might as well be walking on the street or sitting on a bus when interacting with mobile software. Therefore, it is imperative not to ignore the differences of such circumstances and desktop systems in isolated usability laboratories without distractions [5].

This contrasts with Kaikkonen et al. [11] and is supported by Lettner [1] which narrates that conducting unsupervised and automatic field studies will result in more in-depth results than supervised field studies. This statement is also supported by Hertzum [10], who confirms that the results of field studies differ between field tests and laboratory-based tests. Hertzum showed that conducting unsupervised field studies is inexpensive and does not require much preparation. Moreover, supervisors could influence the study by excluding interaction possibilities [1].

2 Related Work

Significant work has been published concerning automated software usability tests, specifically for mobile devices. Lettner et al. [1] approach the present matter implementing a framework to do user interaction logging as basis for usability evaluation on mobile devices. This paper compares commercial frameworks for logging user statistics

on mobile devices, such as Flurry¹, Google Analytics², Localytics³ or User-Metrix⁴. However, these frameworks focus on descriptive user statistics such as user growth, demographics and commercial metrics like in-app purchases. These solutions approach automation of usability tests, but ignore emotional feedback.

Some techniques and methodologies have been reported in significant publications about gathering affective data without asking the users what and how they feel. Physiological and behavioral signals such as body worn accelerometers, rubber and fabric electrodes can be measured in a controlled environment [6, 7]. It is also feasible to evaluate users' eye gaze and collect electrophysiological signals, galvanic skin response, electrocardiography, electroencephalography and electromyography data, blood volume pulse, heart rate, respiration and even, facial expressions detection software [2]. Most of these methods face the limitations of being intrusive, expensive, require specific expertise and additional evaluation time.

2.1 UX Mate

UX Mate [12] is a non-invasive system for the automatic assessment of User eXperience (UX). In addition, they contribute a database of annotated and synchronized videos of interactive behavior and facial expressions. UX Mate is a modular system which tracks facial expressions of users, interprets them based on pre-set rules, and generates predictions about the occurrence of a target emotional state, which can be linked to interaction events.

Although UX Mate provides an automatic non-invasive emotional assessment of interface usability evaluations, it does not consider mobile software contexts, which has been widely differentiated from desktop scenarios [5, 9, 11].

3 Contribution

Our proposal supplements the traditional methods of mobile software usability evaluation by monitoring users' spontaneous facial expressions automatically as a method to identify the moment of occurrence of adverse and positive emotional events. Identifying those events and systematically linking them to the context of interaction is clearly an advance towards overcoming design flaws and enhancing interfaces' strengths.

The automated test generates a graphical log report, timing (a) current application page (b) user events e.g. tap, (c) emotions levels e.g. level of happiness and finally (d) emotional events e.g. smiling or looking away from screen. The gazing away from the screen [2] may be perceived as a sign of deception. For example, looking down tends to convey a defeated attitude but can also reflect guilt, shame or submissiveness. Looking to the sides may denote that the user was easily distracted from the task.

This research has also produced a toolkit for automated emotions logging for mobile software that, in contrast to existing frameworks, is able to trace emotional reactions of users during usability tests and relating them to the specific interaction that

is being performed. This framework can be added to mobile applications with minor adjustments.

3.1 System Structure

The basic system structure is displayed in Fig. 1.

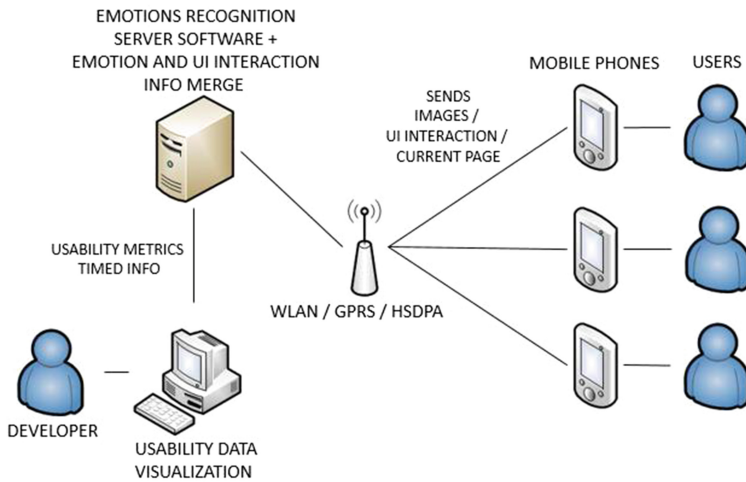


Fig. 1. System infrastructure

The running application uses the front camera to take photos of the user every second. This image is converted to base64 format and is sent via HTTP to the server. The server decodes the base64 information into image and runs the emotion recognition software, which returns the numerical levels of happiness, anger, surprise, smile (true/false) and gaze away (true/false). This information is sent back to the phone via HTTP and written to a text file, with a set of other interaction information. When the user exits the application, the log file is sent to the server, which stores and classifies the test results in a database, which can be browsed via a web front-end.

3.2 Interaction Information Logging

The applications to be tested are written using the library (.dll) we implemented. When the application is started by the user, a log file is created, marking the time, current page, level of happiness, level of anger, level of surprise, smile (0–1), gazing away (true or false) and tap/click (true or false). When tap is true, logs position of tap and name of the control object tapped e.g. button, item on a list, radio button, checkbox, etc.

The generated log file is comma separated value format, enabling visualization in tables, as displayed in Tables 1 and 2.

Table 1. Generated log (part 1)

Time	Page	Happiness	Anger	Surprise
0:00	mainPage	0.204	0.01	0.446
0:01	mainPage	0.252	0.017	0.421
0:02	mainPage	0.252	0.014	0.342
0:03	mainPage	0.276	0.013	0.389
0:04	loginPage	0.216	0.021	0.305
0:05	loginPage	0.204	0.034	0.281
0:06	loginPage	0.226	0.02	0.282
0:07	loginPage	0.223	0.017	0.243
0:08	loginPage	0.221	0.013	0.239
0:09	loginPage	0.204	0.009	0.221
0:10	loginPage	0.24	0.009	0.235
0:11	loginPage	0.214	0.011	0.273
0:12	loginPage	0	0	0
0:13	loginPage	0	0	0
0:14	loginPage	0.206	0.018	0.316
0:15	loginPage	0.276	0.015	0.339
0:16	loginPage	0	0	0
0:17	contactsPage	0	0	0
0:18	contactsPage	0	0	0
0:19	contactsPage	0.324	0.019	0.398
0:20	contactsPage	0.396	0.01	0.421

3.3 Emotion Recognition Software

The emotion recognition software was developed using the well documented Intel RealSense SDK [13]. Among many features, this software development kit allows face location and expression detection in images. This paper does not focus on analyzing any particular image processing algorithms to detect emotions.

3.4 Usability Information Visualization

The front-end web software display one test session as in Fig. 2. It is meant to supplement the traditional usability tools and methods.

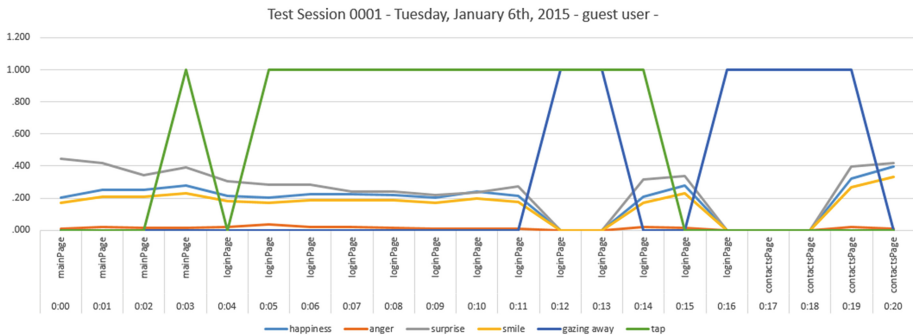
4 Experiments

In order to perform early system functioning check, we planned a test session that would induce negative and positive emotions, not necessarily related to the interface design.

To gather negative feedback, we asked one male adult (32 yo) to login to one of his social networks account and post one line of text to his timeline. During this task, we

Table 2. Generated log (part 2)

Smile	Gazing away	Tap	X	Y	UI object
0.17	FALSE	FALSE	0	0	none
0.21	FALSE	FALSE	0	0	none
0.21	FALSE	FALSE	0	0	none
0.23	FALSE	TRUE	245	467	btnConnect
0.18	FALSE	FALSE	0	0	none
0.17	FALSE	TRUE	201	359	txtUser
0.19	FALSE	TRUE	0	0	keyboard
0.19	FALSE	TRUE	0	0	keyboard
0.18	FALSE	TRUE	0	0	keyboard
0.17	FALSE	TRUE	211	420	txtPassword
0.2	FALSE	TRUE	0	0	keyboard
0.18	FALSE	TRUE	0	0	keyboard
0	TRUE	TRUE	0	0	keyboard
0	TRUE	TRUE	0	0	keyboard
0.17	FALSE	TRUE	237	476	btnLogin
0.23	FALSE	FALSE	0	0	none
0	TRUE	FALSE	0	0	none
0	TRUE	FALSE	0	0	none
0	TRUE	FALSE	0	0	none
0.27	TRUE	FALSE	0	0	none
0.33	FALSE	FALSE	0	0	none

**Fig. 2.** Emotions log automatically generated chart

turned the WLAN connection on and off, in intervals of 30 s. After 5 min of not being able to execute a considerably simple task, the test subject was clearly upset. The emotional feedback logged by our system was in successful accordance to the test session.

To gather positive feedback, we asked one male adult (27 yo) to complete an online quiz with charades and funny answers. The emotional feedback logged by our system

was in successful accordance to the test session, as the user smiled and even laugh about the funny text and imagery.

The test session displayed in Table 1 and Fig. 2 show an example of one test session we have run. The user was asked to login to one communications application in development stage in a research institute.

5 Future Work and Discussions

This work presents an early approach to emotional feedback logging for mobile software usability evaluation. The problem space was narrated through referencing other usability automation research. Some relevant related work was described and distinguished from the present proposal. A system was developed as a proof-of-concept tool to our hypothesis and experiments where performed to raise argumentation topics to provoke advances on the current matter.

Our system logs emotional feedback from users, using the front camera on mobile devices. It stands as a solution for automated mobile software usability evaluation.

The system functional features were tested for negative and positive emotional feedbacks by test sessions that where planned to fail and succeed/provoke smiles, respectively.

Future work will investigate a more in-depth applicability of the logged interaction information. For example, our system is not yet detecting and identifying usability problems. It is strictly logging emotional feedback and UI interactions, merging this information on a timeline, to aid usability evaluation.

References

1. Lettner, F., Clemens H.: Automated and unsupervised user interaction logging as basis for usability evaluation of mobile applications. In: Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia, pp. 118–127. ACM (2012)
2. de Lera, E., Garreta-Domingo, M.: Ten emotion heuristics: guidelines for assessing the user's affective dimension easily and cost-effectively. In: Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as We Know It, vol. 2, pp. 163–166. British Computer Society (2007)
3. Spillers, F.: Emotion as a Cognitive Artifact and the Design Implications for Products that are Perceived as Pleasurable. http://www.experiencedynamics.com/pdfs/published_works/Spillers-EmotionDesign-Proceedings.pdf. Accessed 18 Feb 2007
4. Madrigal, D., McClain, B.: Usability for mobile devices, September 2010. <http://www.uxmatters.com/mt/archives/2010/09/usabilityfor-mobile-devices.php>
5. Oztoprak, A., Erbug, C.: Field versus laboratory usability testing: a first comparison. Technical report, Department of Industrial Design - Middle East Technical University, Faculty of Architecture (2008)
6. Picard, R.W., Daily, S.B.: Evaluating affective interactions: alternatives to asking what users feel. Presented at CHI 2005 Workshop 'Evaluating Affective Interfaces' (Portland, OR), 2–7 April 2005

7. Anderson, R.E.: Social impacts of computing: codes of professional ethics. *Soc. Sci. Comput. Rev.* **10**(2), 453–469 (1992)
8. Harty, Julian: Finding usability bugs with automated tests. *Commun. ACM* **54**(2), 44–49 (2011)
9. Madrigal, D., McClain, B.: Usability for mobile devices, September 2010. <http://www.uxmatters.com/mt/archives/2010/09/usabilityfor-mobile-devices.php>
10. Hertzum, M.: User testing in industry: a case study of laboratory, workshop, and field tests. In: *Proceedings of the 5th ERCIM Workshop on 'User Interfaces for All'* (1999)
11. Kaikkonen, A., Kallio, T., Kekäläinen, A., Kankainen, A., Cankar, M.: Usability testing of mobile applications: a comparison between laboratory and field testing. *J. Usability Stud.* **1**, 4–16 (2005)
12. Staiano, J., Menéndez, M., Battocchi, A., De Angeli, A., Sebe, N.: UX Mate: from facial expressions to UX evaluation. In: *Proceedings of the Designing Interactive Systems Conference (DIS 2012)*, pp. 741–750. (2012)
13. Intel RealSense SDK. <https://software.intel.com/en-us/intel-realsense-sdk>