Where-How-What Am I Feeling: User Context Logging in Automated Usability Tests for Mobile Software

Jackson Feijó Filho[(⊠)], Wilson Prata, and Juan Oliveira

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

Abstract. This work proposes the use of a system to implement user context and emotional feedback and logging in automated usability tests for mobile devices. Our proposal augments the traditional methods of software usability evaluation by monitoring users' location, weather conditions, moving/stationary status, data connection availability and spontaneous facial expressions automatically. This aims to identify the moment of negative and positive events. Identifying those situations and systematically associating them to the context of interaction, assisted software creators to overcome design flaws and enhancing interfaces' strengths.

The validation of our approach include post-test questionnaires with test subjects. The results indicate that the automated user-context logging can be a substantial supplement to mobile software usability tests.

Keywords: Usability · Automation · UXDX · Mobile phones · User research

1 Introduction

In mobile software application stores, consumers frequently find themselves unable to decide which ones 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 manner [1].

In [2] Harty debates how several organizations do not execute any usability evaluation. It is considered too expensive, too specialized, or something to address after testing all the "functionality". This is habitually prioritized because of time and other resource constraints. For these groups, usability test automation can be beneficial.

The assessment of usability in mobile applications delivers valuable measures about the quality of these applications, which assists designers and developers in identifying opportunities for improvement. But examining the usability of mobile user interfaces can be an exasperating mission. It might be extensive and require expert evaluation techniques such as cognitive walkthroughs or heuristic evaluations, not to mention expensive usability lab equipment.

In addition to the resources constraints, there is the desktop versus mobile software matter. Most of the usability methods (e.g. usability inspection, heuristics, etc.) are both valid to desktop as well as to mobile phones software, although it is more difficult for a mobile usability testing context to accomplish 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 extract solid results from such varying user context [7].

Recent work has been published regarding tools for low-cost, automated usability tests for mobile devices. In [8], these tools have been reported to help small software development teams to perform fairly accurate recommendations on user interface enhancements. However, these tools do not consider neither emotional feedback nor contextual awareness of users towards mobile software.

1.1 Emotions and Usability

Emotional feedback is a significant aspect in user experience that chronically goes un-measured in several user-centered design projects [9], especially with small development groups. The examination of affective aspects through empirical user-centered design methods supports software creators in engaging and motivating users while using their systems [10]. Collecting emotional cues will provide another layer of analysis of user data, augmenting common evaluation methods. This results in a more accurate understanding of the user's experience.

1.2 Automated Tests and Unsupervised Field Evaluations

It is important to reference the importance of automated tests, while being performed for mobile devices. In contrast to desktop applications or web sites, mobile applications have to compete with stimuli from the environment, as users might not be sitting in front of a screen for substantial amounts of time [11]. Due to the natural 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. Hence, it is important not to ignore the differences of such circumstances and desktop systems in isolated usability laboratories without distractions [7].

2 Related Work

Previous work has been published about automated software usability tests, specifically for mobile devices. Here we divide related work in two groups: "UI Interactions, Automated and Unsupervised Logging and Analytics" and "Emotions Logging Systems".

2.1   UI Interactions, Automated and Unsupervised Logging and Analytics

Several commercial frameworks for logging user statistics on mobile devices, such as Flurry[1], Google Analytics[2], Localytics[3] or User-Metrix[4]. However, these frameworks focus on 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, user context and even UI interaction information.

Flurry Analytics. Commercial solutions such as Flurry, which is taken as an archetype for commercially available analytic frameworks (i.e. Localytics[5], Mobilytics[6], Appuware[7] and UserMetrix[8]), try to get an audience perception. They deliver usage statistics based on metrics like average users per day or new users per week. To use these frameworks, the development teams offer support and code snippets to integrate the framework with existing applications. However, developers are responsible for adding framework functionality at the right place in their applications. Besides the option of collecting demographic information (i.e. gender, age or location), the framework also offers a possibility to track custom events. However, metrics that provide assumptions about the quality of user interfaces are generally missing. In contrast to these frameworks, our approach directly focuses on associating UI interaction, emotions and user context logging.

EvaHelper Framework. In 2009, Balagtas-Fernandez et al. [16] presented an Android-based methodology and framework to simplify usability analysis on mobile devices. The EvaHelper Framework is a 4-ary logging system that records usability metrics based on a model presented by Zhang et al. [18]. Although Balagtas-Fernandez et al. focus on the part of automatically logging user interaction, they do not focus on the evaluation of the collected data. For the visualization of their results, they use third-party graph frameworks that are based on GraphML[9] to visualize a user's navigational graph. Compared to our approach, they solely provide navigational data for single application usages. They do not augment the graph with some kind of low-level metrics to enable the identification of emotional feedback or contextual adversities.

Automatic Testing with Usability Metrics. This work presents a methodology and toolkit for automatic and unsupervised evaluation of mobile applications. It traces user interactions during the entire lifecycle of an application [8]. The toolkit can be added to mobile applications with minor changes to source code, which makes it flexible many

---

[1] http://www.flurry.com.

[2] http://www.google.com/intl/de/analytics.

[3] http://ww.localytics.com/.

[4] http://usermetrix.com.

[5] http://www.localytics.com/.

[6] http://www.mobilytics.com/.

[7] http://www.appuware.com/.

[8] http://usermetrix.com/.

[9] http://graphml.graphdrawing.org.

types of applications. It is also able to identify and visualize design flaws such as navigational errors or efficiency for mobile applications.

2.2   Emotions Logging Systems

Some techniques and methodologies have been reported 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 [19]. 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 [9]. Most of these methods face the limitations of being intrusive, expensive, require specific expertise and additional evaluation time.

UX Mate. UX Mate [17] 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 [7, 11, 12]. Furthermore, it does not take into account the contextual awareness of the user.

Emotions Logging System. It [15] proposes the use of a system to perform emotions logging in automated usability tests for mobile devices. It 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. No aspects of user context are considered in this work.

3   Contribution

Our proposal supplements the traditional methods of mobile software usability evaluation by:

1. Monitoring users' spontaneous facial expressions automatically as a method to identify the moment of occurrence of adverse and positive emotional events.
2. Detecting relevant user context information (moving/stationary status; location, weather conditions; data connections availability;
3. Systematically linking them to the context of interaction, that is, UI Interaction (tap/drag) and current app view.

4. 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;
   (d) emotional events e.g. smiling or looking away from screen;
   (e) moving/stationary status;
   (f) location;
   (g) weather conditions;
   (h) data connection availability (WLAN/3G/4G);

3.1   Example Scenarios

According to [9], the gazing away from the screen 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.

The work in [15] clearly addresses this matter, by logging the event of "gazing away" from the screen. Although there are certain scenarios where this is perfectly acceptable even if the user is fully committed to the UI interaction:

- While waiting at a bus stop: user has to constantly gaze away from screen, to check for coming buses;
- In a conversation: user has to constantly gaze away from screen, to demonstrate e.g. any media on an application, to the conversation partner.

These are examples of the evident need of user context awareness during automated and unsupervised mobile software usability tests.

3.2   System Structure

The basic system structure is displayed in Fig. 1.

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.

For the user context gathering, an additional layer of software periodically logs user's location, appointments schedule (time to next appointment), moving status (walking running, etc.), data connection availability, weather information, UI interaction and current application page. 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.
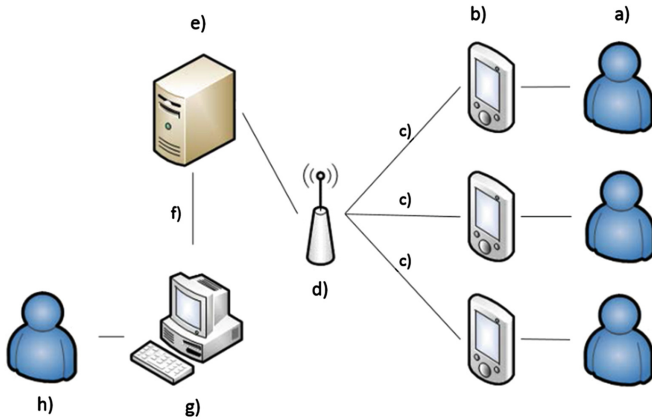
Fig. 1. (a) User; (b) mobile phone; (c) sends face images, GPS location, appointments schedule, moving status (walking, running, etc.), data connection availability, weather information, UI interaction (tap, drag, flick, etc.) and current application page; (d) WLAN, GPRS or HSDPA; (e) emotions recognition software, user context framework, UI interaction integration; (f) emotions and user context log; (g) log data visualization; (h) developer.

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, registering the time, current page, emotional feedback and user context. For simplicity analysis, we are logging only tap interaction - 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.

Emotion Recognition Software. The emotion recognition software was developed using the well documented Intel RealSense SDK [13]. Among many features, this
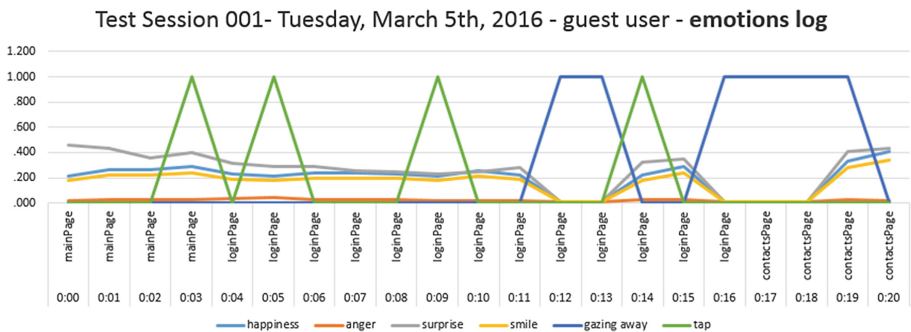
Fig. 2. Emotions log from a 20 s test session (Color figure online)

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.

Usability Information Visualization. The front-end web software display one quick (20 s) test session as in Fig. 2.

4   Experiments

4.1   Emotional Feedback

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 test negative feedback, we asked one male adult (32 years old) to login to one of his social networks accounts and post one line of text to his timeline. During this task, we 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 noticeably disappointed. The emotional feedback logged by our system was in successful accordance to the test session.
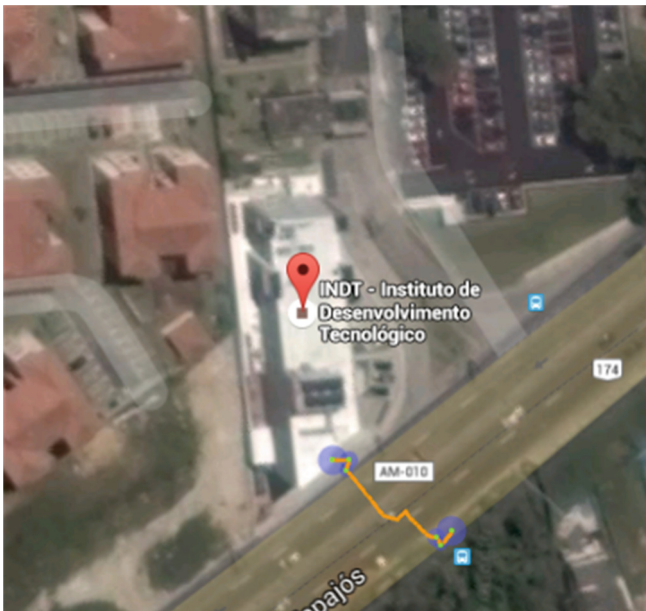
Fig. 3. Path of user during test. Circle size shows amount of stationary time. Clicking on each circle will show user context and emotional feedback (Fig. 4).

Fig. 4. Emotions and context information at given time

To test positive feedback, we asked one male adult (27 years old) to answer a quiz of charades and funny answers. The emotional feedback logged by our system was successful, as the user smiled and even laugh about the funny text and imagery.

The test session displayed in Fig. 2 show another example of one test session we have run. The user was asked to login to one instant messaging application in development stage in a research institute.

4.2   User Context

In order to perform early system functioning check, we planned a test session that would log an example user context information.

The user was asked to lo login to one instant messaging application in development stage in a research institute, while leaving an office premise, crossing the street and walking towards a bus stop. See Fig. 3.

5   Future Work and Discussions

This work presents an early approach to user context and 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 user context information and emotional feedback from users of mobile phone software. It stands as a solution for automated usability evaluation.

Future work will investigate a more in-depth applicability of the logged interaction information. More importantly, it will research the integration of the different possible variables to be added to this framework.

Additionally, this technique will be compared to other usability methodologies to validate the benefit of our approach.

References

1. Brajnik, G.: Automatic web usability evaluation: what needs to be done. In: Proceedings of Human Factors and the Web, 6th Conference, June
2. Harty, J.: Finding usability bugs with automated tests. Commun. ACM **54**(2), 44–49 (2011)
3. Witold, A.: Consolidating the ISO usability models. In: Presented at the 11th International Software Quality Management Conference and 8th Annual INSPIRE Conference (2003)
4. Rubin, J., Chisnell, D.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, 2nd edn. Wiley, New York, NY (2008)
5. Vukelja, L. et al., Are engineers condemned to design? A survey on software engineering and UI design in Switzerland. In: Presented at the 11th IFIP TC 13 International Conference on Human-Computer Interaction, Rio de Janeiro, Brazil (2007)
6. Dumas, J.S., Redish, J.C.: A Practical Guide to Usability Testing: Intellect Books (1999)
7. 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)
8. Lettner, F., Holzmann, C.: 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)
9. 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)
10. Spillers, F.: Emotion as a Cognitive Artifact and the Design Implications for Products that are Perceived as Pleasurable (2007). http://www.experiencedy-namics.com/pdfs/published_works/Spillers-EmotionDesign-Proceedingspdf. Accessed 18 Feb 2007
11. Madrigal, D., McClain, B.: Usability for mobile devices, September (2010). http://www.ux-matters.com/mt/archives/2010/09/
12. 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)
13. 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)
14. Kjeldskov, J., Skov, M.B.: Was it worth the hassle? Ten years of mobile HCI research discussions on lab and field evaluations. In: Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM (2014)
15. Feijó Filho, J., Prata, W., Valle, T.: Mobile software emotions logging: towards an automatic usability evaluation. In: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services. ACM (2015)
16. Balagtas-Fernandez, F., Hussmann, H.: A methodology and framework to simplify usability analysis of mobile applications. In: Proceedings of the 2009 IEEE/ACM International Conference on Automated Software Engineering (2009)
17. Staiano, J., Menendez, M., Battocchi, A., De Angeli, A., Sebe, N.: UX mate: from facial expressions to UX evaluation. In: Proceedings of the Designing Interactive Systems Conference, pp. 741–750. ACM DIS (2012)

18. Zhang, D., Adipat, B.: Challenges, methodologies, and issues in the usability testing of mobile applications. Int. J. Hum. Comput. Interact. **18**(3), 293–308 (2005)
19. Picard, R.W., Daily, S.B.: Evaluating affective interactions: alternatives to asking what users feel. In: Presented at CHI 2005 Workshop Evaluating Affective Interfaces (Portland, OR, 2–7 April 2005). Anderson, R.E.: Social impacts of computing: codes of professional ethics. Soc. Sci. Comput. Rev. 10(2), pp. 453–469 (1992)