Acceptance Testing of Mobile Applications – Automated Emotion Tracking for Large User Groups

Simon André Scherr
Fraunhofer IESE
Kaiserslautern
Germany
simon.scherr@iese.fraunhofer.de

Frank Elberzhager
Fraunhofer IESE
Kaiserslautern
Germany
frank.elberzhager@iese.fraunhofer.de

Konstantin Holl
Fraunhofer IESE
Kaiserslautern
Germany
konstantin.holl@iese.fraunhofer.de

ABSTRACT

Mobile applications[1] are nowadays used by everyone. The success of a mobile app highly depends on its user acceptance, which must be checked as part of quality assurance. However, such tests are costly because they usually include testers using the app manually. An obvious solution for improving efficiency is to automate certain test steps. In this article, we present an approach that tracks user emotions automatically to support acceptance testing. Furthermore, we consider user motivation in a positive way with gamification solutions and focus on data privacy aspects in order to gain the trust of potential test users.

CCS CONCEPTS

Software and its engineering → Software creation and management → Software verification and validation → Process validation → Acceptance testing

KEYWORDS

Test, Emotions, Animojis, User Acceptance, Automation

1 Introduction

Since Apple's App Store started in 2008 with about 800 apps, it has grown to more than two million apps, and the Google Play Store even has more than three million apps. The development of many apps has been professionalized, as fast time to market at high quality is needed to stay competitive and to do business with apps. Typical quality assurance activities aimed at ensuring high quality are reviews of software artifacts, pair programming, unit and regression testing, continuous integration with automated tests, or crowd testing. However, app developers face growing pressure from users. The reasons are that users can, for example, provide feedback easily, and that feedback influences other users' decision of whether to install and use these apps. Furthermore, apps with bad quality are deleted quickly and alternative are used instead. Such often highly valuable user feedback is not yet widely considered by app developers, or only to a limited extent, as such analyses need effort and a clear process how to incorporate feedback into the development process is needed. For users, on the other hand, providing feedback is sometimes too complicated (e.g., many people no longer give meaningful written feedback on a smartphone). It would be ideal if an app developer could observe users while they use the app and see, for instance, at which steps the user easily navigates through the app or where the next step is unclear. Furthermore, monitoring users' emotions can also provide helpful insights into the quality of an app. existing solutions to capture emotions require buying dedicated hardware like [1], which the tester has to wear, often including a lot of cables. This prevents the user to act in his regular environment. A laboratory setting is usually not possible – especially if targeting large test user groups. However, today's smartphones all have a camera, which could be used instead. This, however, leads to the challenge of how to get users to participate. Concretely, we derive two research questions from this challenge:

1. Motivation: How can users be motivated to provide feedback based on being observed while using an app?
2. Trust: How can the trust of potential test users be gained in terms of privacy regarding the captured test data?

Our idea is to use emojis and animated emojis like the Apple Animojis [2], which can express what the user feels while using the app. An emoji is a pictograph that represents, for instance, faces, emotions, or feelings. Emojis are integrated into most on-screen keyboards of mobile devices. The main idea we propose in our contribution is to recognize the facial expression of a user using an app and then translate the emotions found into animated emojis, which are then provided to the app developer together with additional information tracked by a tracking library.

This library is bundled into the app in the form of a component. It consists of three main components: a logger, a pre-usage miner, and a sender. The logger tracks several interactions of the user and device data. The pre-usage miner takes the data from the logger and consolidates it. The goal is to identify the data that is worth further analysis. Finally, the sender component sends data packages to the backend, where the full analysis of the data, including merging of the data of all users, is performed.

MOBILESoft '18, May 27–28, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5712-8/18/05…$15.00
https://doi.org/10.1145/3197231.3197259

We motivate the user with a gamification mechanism (question 1) and focus on his/her emotions. This is highly valuable information for app developers as they can easily understand which steps in an app the users like and where problems exist. Furthermore, the barrier for users to provide feedback is reduced, as no real video (which would obviously be problematic due to the amount of data traffic) is sent to the app developers, but only emojis that translate the emotions of the users while using the app. This information privacy addresses the trust aspect (question 2). We believe that different kinds of feedback, such as textual feedback from stores combined with such emotion-based feedback, will be considered much more frequently in future quality assurance approaches and will add value to app development, as the required technical means are more and more common. Compared to a classical user acceptance test where a product owner or a sales person, for example, performs such tests based on user stories or requirements, and which is explicitly part of a dedicated development phase such as a sprint, we explicitly expand testing and consider real users of the app for providing feedback. Such an approach merges quality assurance und usage of an app, and will connect app developers and app users more strongly so that both sides can benefit: The user gets better quality faster, while the app developer understands the users better and gets fast feedback on what to improve in their app.

2 The Vision

Our vision is to enhance user acceptance testing for apps by allowing a large set of users taking part in testing, and to make it as easy as possible for users to be willing to provide feedback. This allows to detect how the software is improving over time. To achieve this, it is essential to motivate the users and to communicate trust by maintaining their information privacy. Therefore, we propose the use of animated emojis representing the users and their reactions while using the app.

A traditional form of user testing is user observation. Here an observer looks at the user's reactions and actions while performing a task. Obviously this method does not scale in terms of observation length and number of test persons, but we can use two characteristics of user observations to realize our vision: (1) The participant is performing real scenarios within the software and not fictional scenarios; (2) the observer immediately sees the reactions of the end user while using the application.

We incorporate these two advantages into an automated approach. What is essential for detecting the reactions are the emotions of the user while using the app. We record these emotions by using the camera of the mobile device. For this purpose, we intend to apply face detection techniques. To maintain the users' privacy, we are not recording and storing the actual face of the user, but symbolic representations in the form of animated emojis. This enables us to use not only non-fictional scenarios, but also real situations where the app is used.

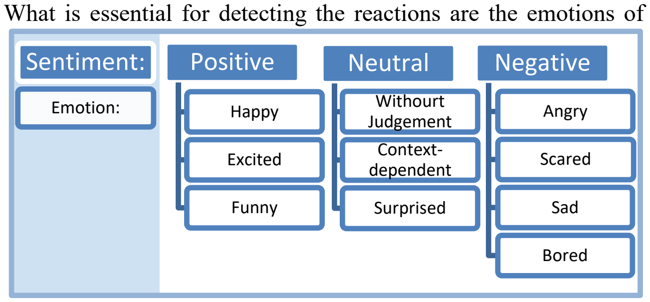

Figure 1: Classification of sentiments and emotions

Recently, we have created an emotional feedback model to analyze emotions by using emojis within texts [3] (initial presentation) [4] (results, submitted). Emojis are icons or pictographs used to represent emotions, activities, or objects. We conducted a survey with 107 participants. This gave us insights into how emojis are related to emotions. It turned out that people perceive emojis in a very homogenous way and we were able to classify about 600 emojis in terms of concrete emotions. Our sentiment and emotion model can be seen in Figure 1. We have not included emojis that simply replace words without having any kind of emotional connotation. For our model it is important that we first got a coarse-grained categorization into sentiments – positive, neutral, and negative – and then achieved a more precise classification in terms of emotions being part of each sentiment.

We built this model by considering the characteristics of emojis, emotion models such as those by Ekman [5] [6], Lazarus [7], and Plutchik [8], as well as sentiment analysis [9]. In the neutral sentiment, we have emojis that are neither positive nor negative as well as emojis that might be either positive or negative, depending on the context. We also added the emotion "surprised" to the neutral sentiment category, as a person can be surprised a positive or negative way. The negative sentiment category comprises the emotions angry, scared, sad, and bored. The positive category consists of the emotions happy, excited and funny.

Our goal is to enhance the tracking library described in the introduction with an emotional component. While the tracker records the interaction, application, and device data, the emotional component records the user in the form of an animated emoji representing the user's emotion at a certain point of time or during a particular interaction. These data are connected with each other. In order to be able to cover all the emotions as part of our emotion model, we have to use a set of emojis that is capable of showing all the emotions (e.g. the cat face as visualized in Figure 2).

Figure 2: Person showing different facial expressions represented as emojis [2]

[2] Image Source: https://blog.emojipedia.org/apples-new-animoji/

Representing the user's reaction in the form of an animated emoji has advantages compared to video recording. First, we are actively protecting the user's privacy as not the person is recorded, but only the animated emoji. Second, the emoji provides a more abstract and unified representation of a face.

An example of a person expressing various emotions and the corresponding representation in the form of a cat emoji can be seen in Figure 2. The facial expressions are communicated and visualized in a simple way by the emoji. In addition, emojis provide unification, as the characteristics of each individual human face are not visible anymore. This prevents being distracted by the face itself and allows us to focus more on the overall emotional expression. As the tracking library is able to log content changes it is possible to distinguish if users are reacting to content changes or to changes within the application itself.

Of course, it would also be possible to detect the user's emotions without using emojis, by just recording emotion values, but this is where the second aspect of our vision comes into play. We want to motivate users to actively take part in testing. Recording an emoji provides the advantage that we are able to show the user the exact data being recorded. The user can directly see and understand which data is being collected. This fact can be combined with the addition of playful gamification aspects to the acceptance test. One idea would be to enable users to send their emoji representation in the form of a picture or video with a comment to other people using the app or to social media sites. This might also be an opportunity to apply concepts for acquiring more users for the test. The introduction of the iPhone X and a feature called "Animoji" [10], which allows users to send messages in the form of animated emojis has shown that many people enjoy such features [11], [12].

Figure 3 shows the example process of our vision. Different users are getting a version of the app. On "Screen X", the emotion is captured and gets represented by an emoji. This data is sent to a backend in combination with the tracking data for interaction, app, and device data. Here the data of all the users and app executions is combined and processed. As part of the automated data processing, connections are made between the uploaded data and the use cases of the app. Part of this processing consists of creating an average flow of emotions per workflow element, screen and use case. In additions changes in the emotion of the users can be identified and related to the action they have performed. The test manager or quality assurance manager is able to see the captured data. He or she can analyze the results and especially check the elements containing negative emotions. The public opinion trend can be analyzed in various ways. One example would be to check the user's response to new or changed features. Furthermore, A/B tests can be performed to determine the differences in terms of emotions. We would also be able to get an understanding of how emotions change after certain features have been used for a longer period of time, leading to insights if there is any difference between new users of the app and long-term users. A critical aspect while using mobile apps is how the user copes with waiting times and failures within the app. Understanding the user's reactions to such things is also important. In our approach, we are able to find out when those issues are starting to annoy the user. This allows deriving implications for future product development in order to improve the app's quality. The tracking also allows us to understand how easy it is to perform the processes of the app and where the users are not sure what to do next by combining the emotions with the regular tracking data.

3 Why is this new?

Our approach takes into account state-of-the-art technology of mobile devices using capabilities that have existed only for a very short time. It is built on three foundations: the computing power of mobile devices, their camera capabilities, and the recent advances of augmented and virtual reality technologies.

In recent years, the computing power of mobile devices such as smartphones and tablets has increased to a huge extent. As those devices are becoming more powerful than high-performance desktop computers were just a few years ago, the marketing departments of those companies are starting to use terms like "Post PC Era" [13] or "mobile first strategy" [14]. These very new capabilities allow computing tasks to be performed that would not have been possible before. In addition, the camera systems of these devices are getting so advanced that even photographers [15] and movie makers are using them in professional productions such as creating TV ads or even cinema productions [16]. Besides creating better optical lenses for use in the camera systems, companies have also started to use multiple lenses within one camera to achieve better image capabilities, for instance for portrait or macro photography, or for better detection of objects.

This creates new possibilities for augmented reality becoming popular in the mobile world [17], [18]. Various apps for gaming, shopping, productivity, business, or navigation have become popular that enrich the camera image with augmented features or replace a part of the camera image with computer-generated objects [18]. Some of these apps allow creating photos of a person with AR changes by switching clothes or accessories or replacing

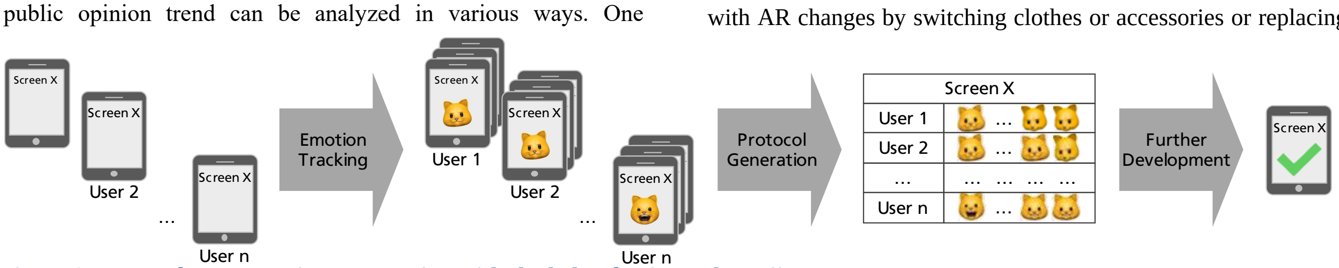

Figure 3 Process for automating user testing with the help of animated emojis

parts of the body [19], [20]. Nevertheless, traditional camera sensors do not have good ability to detect three-dimensional shapes. Early approaches for securing smartphones via the camera system by checking the user's face were therefore not robust [21], [22].

A very new change in the technology for mobile camera systems is the ability to detect the depth and shapes of three-dimensional objects. This is not achieved by using only optical lenses, which makes them more acting like a sensor and less like a camera. This has a positive impact on the energy consumption, as traditional computer vision is not used for data processing. These camera systems project a large number of invisible dots, which are recognized by the camera. This makes them able to detect surfaces in a suitable way. The detection works so well that for some devices, functionalities were built in to unlock the device by just looking at it, with an error rate of 1 in 1,000,000 [23] for a random person being able to unlock the phone. Fingerprint sensors in state-of-the-art mobile devices currently have an error rate of 1 in 50,000 [23]. These camera systems constitute a huge step also in terms of improving the capabilities of augmented reality, as they are capable to distinguish several dozen muscle movements in a face. In addition, these systems are able to better detect people, detach them from the surrounding area, and create a different background scene on the device [10].

These are enablers, in terms of recent technology, for our vison. Earlier systems on the mobile device were not able to do this in real time using photos, but current systems can even do it with real-time video processing and recording.

Besides the technological newness, our vision is also an innovation in terms of testing. Current testing approaches for getting feedback from real users do not scale in terms of number of testers or long-term observations. In addition, these approaches are not automated. On the other hand, automated testing approaches usually focus on smaller parts of apps, such as A/B tests, or do not use data acquired from real users, such as automated UI Testing, where the UI is controlled by a testing framework. We combine the advantages of the detailed results from acceptance tests being performed with real users with the automation support present in system and integration testing. This also means that our vison removes the weakness of non-scalable acceptance tests and fictional data from test automation available to date.

4 Risks

The idea of presenting the emojis to the users may lead to a bias of the evaluation results. This could be the case if a user is more interested in the representation of his emotions by an emoji than in the usage of the app itself. This is true especially if the representation by an emoji is shown during the usage of the mobile application. This may lead to situations where the user reacts to the emoji and not to application itself. Consequently, the recorded user emotion will not represent the response to the mobile application but to the emoji, which should not be part of the measurement. This means that the possible impact of the representations of emotions to the user needs to be investigated in a separate evaluation. In case of significant bias, the representing emojis will only be shown to the test user after the completion of the test run. Furthermore, additional feedback data might be considered together with emotional data in order to get a more widely validated basis.

Another risk is that potential test users might still have concerns regarding data privacy. The representation of emotions by emojis could make people feel scared because they reveal the opportunities enabled by modern technology. This may lead to the rejection of our test approach by potential test users. To mitigate this risk, we have to build up trust through clarification and clear participation policies. This could increase the test users' confidence and hence their acceptance of our approach.

From a technical point of view, our approach requires a viable tracking library. Restrictions of operating systems running the application and the library may hamper the realization of our approach. Furthermore, the tracking library must not influence the user experience in any way. If it requires too many resources, the execution of the mobile application under test may be influenced negatively. This would bias the test results. Consequently, a high-priority requirement for the tracking library is its resource efficiency during execution.

5 Next Steps

The definition of additional requirements of the tracking library is the main next step. As described above, the tracking library must be resource-efficient in order to prevent any influence on the mobile application. Further requirements will be derived on a more technical level and based on our overall testing process. Technically, we plan to track information about the user behavior such as times and tap patterns. On this level, we will track the user's emotions, too. All tracked information then needs to be presented to developers or managers responsible for the development of the system under test. This will be realized by a dashboard presenting the tracked data systematically. Based on this dashboard, the responsible persons will handle the analysis and further processing of the insights as part of our overall testing process.

The implementation of the tracking library and the corresponding dashboard will be evaluated in a further step. This will comprise several mobile applications that are publicly accessible. However, we will not perform the evaluation absolutely sequentially. Especially the impact of presenting emojis live based on the user's emotions will already be evaluated in an early phase of the development. In a later step, we plan to combine our emotion-based test approach with functional test plans. This will result in a process blending usability testing and classical testing.

ACKNOWLEDGMENTS

The research described in this paper was performed as part of the project Opti4Apps (grant no. 02K14A182) of the German Federal Ministry of Education and Research (BMBF). We also thank Sonnhild Namingha for proofreading.

REFERENCES

[1] Homepage - Emotiv. Emotiv. https://www.emotiv.com/.

[2] Todd Haselton. Watch how Apple's amazing new Animoji mimic your facial expressions on the iPhone X: 2017. *https://www.cnbc.com/2017/11/01/apple-iphone-x-animoji-mimic-your-face.html*. Accessed: 2018- 01- 08.

[3] Simon André Scherr, Frank Elberzhager and Konstantin Holl. 2017. An Automated Feedback-Based Approach to Support Mobile App Development. *2017 43rd Euromicro Conference on Software Engineering and* Advanced Applications (SEAA)

[4] S. Scherr, S. Polst, L. Müller, K. Holl, and F. Elberzhager „The Perception of Emojis for Analyzing App Feedback", International Journal of Interactive Mobile Technologies [submitted]

[5] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 2: 124–129. https://doi.org/10.1037/h0030377

[6] Paul Ekman. 1992. An Argument for Basic Emotions. *Cognition and Emotion* 6, 3–4: 169–200. https://doi.org/10.1080/02699939208411068

[7] Richard S. Lazarus. 1991. *Emotion & Adaptation*. https://doi.org/10.1007/s13398-014-0173-7.2

[8] Robert Plutchik. 1980. A General Psychoevolutionary Theory of Emotion. In *Theories of Emotion*. 3–33. https://doi.org/10.1016/B978-0-12-558701-3.50007-7

[9] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M. Mohammad. 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 50: 723–762. https://doi.org/10.1613/jair.4272

[10] Apple Inc. The future is here: iPhone X: 2017. *https://www.apple.com/newsroom/2017/09/the-future-is-here-iphone-x/*.

[11] Animoji: We Did Not Expect This. *Emojipedia*, 2017. https://blog.emojipedia.org/animoji-we-did-not-expect-this/.

[12] Welch, C. How to make Animoji karaoke. *The Verge*, 2017. https://www.theverge.com/2017/12/5/16735822/animoji-karaoke-tutorial-iphone-x.

[13] Tim Carmody. Fighting words: Apple's 'Post-PC' and Microsoft's 'PC Plus' were never that different: 2012. *https://www.theverge.com/apple/2012/7/12/3151491/fighting-words-apple-post-pc-microsoft-pc-plus*. Accessed: 2018- 01- 05.

[14] Satya Nadella: Mobile First, Cloud First Press Briefing | Stories: 2014. *https://news.microsoft.com/2014/03/27/satya-nadella-mobile-first-cloud-first-press-briefing/*. Accessed: 2018- 01- 05.

[15] Raymond Wong. A photographer used an iPhone 7 Plus to take this stunning 'Billboard' magazine cover: 2017. http://mashable.com/2017/02/16/billboard-magazine-cover-shot-with-iphone-7-plus-portrait-mode/#cUSZHQezMqq3. Accessed: 2018- 01- 08.

[16] Michel Gondry debuts film shot entirely on Apple's iPhone 7: 2017. http://mashable.com/2017/07/01/michel-gondry-iphone-7-film/#e5yFWPcv2kqh. Accessed: 2018- 01- 08.

[17] Clemens Arth and Dieter Schmalstieg. 2011. Challenges of Large-Scale Augmented Reality on Smartphones. *Graz University of Technology, Austria*: 1–4.

[18] Tobias H. Höllerer and Steven K. Feiner. 2004. Mobile Augmented Reality. *Understanding Augmented Reality*: 1–39. https://doi.org/10.1016/B978-0-240-82408-6.00007-2

[19] Lenskart: https://play.google.com/store/apps/details?id=com.lenskart.app&hl=en. Accessed: 2018- 01- 08.

[20] INKHUNTER Try Tattoo Designs on the App Store: https://itunes.apple.com/us/app/inkhunter-try-tattoo-designs/id991558368?mt=8. Accessed: 2018- 01- 08.

[21] Mike Orcutt. 2016. *MIT Technology Review*. Are Face Recognition Systems Accurate? Depends on Your Race.

[22] Keyurkumar Patel, Hu Han, and Anil K. Jain. 2016. Secure Face Unlock: Spoof Detection on Smartphones. *IEEE Transactions on Information Forensics and Security* 11, 10: 2268–2283. https://doi.org/10.1109/TIFS.2016.2578288

[23] Apple Inc. 2017. *Face ID Security.* Apple.