Evaluating the Usability of Mobile Instant Messaging Apps on iOS Devices

Sergio Caro-Alvaro, Antonio Garcia-Cabot, Eva Garcia-Lopez, Luis de-Marcos, Jose-Javier Martinez-Herráiz
Computer Science Department
University of Alcala
Alcala de Henares, Spain
sergio.caro@edu.uah.es; a.garciac@uah.es; eva.garcial@uah.es; luis.demarcos@uah.es; josej.martinez@uah.es

***Abstract*— Instant messaging apps are experiencing a significant upturn in recent years in mobile devices. This paper shows the results of applying a systematic evaluation of these applications on iOS platform that was performed to identify their main usability issues. As a result of this evaluation some guidelines for improving the usability of these applications are proposed, such as a carefully designing the interface or not exceeding more than eight (and preferably not more than six) interactions to perform the main tasks. The results and the guidelines proposed will help in the future to create more effective mobile applications for instant messaging.**

Keywords- Instant messaging; mobile usability; keystroke level modeling; mobile heuristic evaluation

I. INTRODUCTION

The increased use of mobile devices [1] has led to the number of applications (apps) available in mobile markets has also increased significantly in recent years, such as instant messaging (IM) apps, which have become ubiquitous in contemporary society. To increase the chances of an app to be chosen by users among many others it is essential that it has a good usability. It is important to study usability in desktop applications, but it is even more important to study it in mobile apps because mobile devices have some limitations when compared to personal computers (PC) [2] [3], such as small-sized screens, limited input mechanisms, battery life, etc. These characteristics make necessary studying usability for mobile devices separately from usability for PCs.

A mechanism for systematic evaluation (also called protocol) was created by Martin et al. [4] for studying usability in existing mobile apps, and consists of five steps: (1) Identify all potentially relevant applications, (2) remove light or old versions of each application, (3) identify the primary operating functions and exclude all applications that do not offer this functionality, (4) identify all secondary functionality and (5) test the main functionalities using the following methods: Keystroke-Level Modeling (KLM) for estimating the time taken to complete each task to provide a measure of efficiency of the applications [5] [6] and Mobile Usability Heuristics (MUH) for identifying more usability problems using a usability heuristic evaluation.

Since mobile IM apps are becoming widely used in recent years, it is especially important to study the most common usability problems in this kind of apps, in order to get some guidelines or good practices for developers of mobile IM apps. This paper shows a systematic evaluation of instant messaging apps and the results obtained during this evaluation. Finally, some recommendations are proposed from a viewpoint of mobile usability.

The paper is organized as follows: Section 2 shows the evaluation carried out and the results obtained in the systematic evaluation. Section 3 presents a discussion of the results with previous work and, finally, Section 4 explains the recommendations and conclusions obtained from the results.

II. EVALUATION AND RESULTS

This section shows the systematic evaluation carried out and the different results obtained in the steps. iOS platform (an iPhone 4) was used along the evaluation steps.

A. Steps 1 to 4

In the first step, potential and relevant applications available in the iOS app store were identified. The "instant messaging" term was used to search in the app market. As a result, 243 applications were classified as potential applications.

In the second step, the applications that were not fully functional (i.e., demos, lite or trials) were removed from the list of potential applications. In all, 20 applications (8%) were removed from the initial list.

An application can be considered as instant messaging when it meets all the main functionalities, which were defined in Step 3 as follows:

- Task 1 (T1). Send an instant message to a specific contact ([7]).
- Task 2 (T2). Read and reply an incoming message ([7]).
- Task 3 (T3). Add a contact ([7]).
- Task 4 (T4). Delete/Block a contact (derived from [8] and [9]).
- Task 5 (T5). Delete chats ([10]).

Once the main functionalities were detected and defined, the applications that did not meet all these requirements were discarded. As a result, only 39 (18%) applications met the main functionalities to be considered as IM applications.

In the fourth step, it was necessary to discover the secondary functionalities on the applications selected in the previous step (11 apps were discarded because they ran

 ISBN: 978-1-61208-438-1

anomaly with unrecoverable errors). The most common secondary functionalities detected were including a user profile avatar (74.36%), sending pictures (66.67%) and sending videos (64.10%), among others.

B. Step 5-A: Keystroke-Level Modeling

In this step, the remaining applications (28 apps in total) were reviewed in order to count the number of interactions (KLM) required to perform each of the main functionalities established in Step 3.

TABLE I. TOP10 KLM RESULTS: NUMBER OF INTERACTIONS REQUIRED FOR COMPLETING THE TASKS

App	v.	T1	T2	T3	T4	T5	Total
Surespot encrypted messenger	6.00	5	5	3	4	4	21
Hike messenger	2.5.1	5	5	5	5	4	24
HushHush App	1.0.3	6	6	4	4	4	24
Hiapp Messenger	1.0.6	6	6	5	5	4	26
Kik Messenger	7.2.1	5	5	5	6	5	26
Touch	3.4.4	7	5	5	5	4	26
WhatsApp Messenger	2.11.8	5	6	5	5	6	27
BBM	2.1.1.64	6	5	7	6	4	28
iTorChat	1.0	7	6	5	5	5	28
XMS	2.31	6	6	6	6	4	28
Mean		**6.54**	**5.75**	**6.25**	**5.96**	**5.36**	**29.85**

Table 1 only shows the top 10 with the fewest interactions; the minimum number of interactions for completing all tasks was 21 (Surespot encrypted messenger) and the app with maximum (39 interactions) was Spotbros app, but obviously it is not shown in Table 1. An average of 6 interactions for each task can also be observed.

In order to send a new message (task 1), the apps with fewer interactions (5 interactions) obtained these results by showing the keyboard automatically, although only 6 apps have this feature (WhatsApp, Tuenti, IM+ Pro7, Hike, Surespot and Kik).

All analyzed apps required between 4 and 6 interactions to reply to a given message (task 2), except Spotbros, which required 8 interactions because chats were shown only when a button was pressed. The similarity in the number of interactions is because almost all applications had a section that contained the active chats grouped.

The most variations were observed in task 3 (adding a contact): from 3 interactions (Surespot encrypted messenger) to 10 interactions (Tuenti and Spotbros). This is because some applications (11 of all apps analyzed) used the agenda of the mobile device and others used their own contact list, causing alternative implementations of this process, thus requiring extra data in some cases.

Finally, for task 5 (deleting a chat) most of the examined apps required between 4 and 6 keystrokes, due to the similar implementation of the process.

For the next step (5.B. heuristic evaluation), not all apps were selected to continue in the process. As in previous studies [11]-[13], the four applications with fewer interactions were selected for the next step. In this case, applications with the same number of interactions were considered as one. Therefore, 7 applications in total were selected: Surespot (21 interactions), Hike Messenger and HushHushApp (both 24 interactions), Kik Messenger, Hiapp Messenger and Touch (26) and WhatsApp Messenger (27).

C. Step 5-B: Mobile Usability Heuristics

In this step, the mobile usability evaluation using heuristics was performed. Six (6) experts carried out the evaluation with the 7 applications selected in the previous step. As Bastien [14] and Hwang and Salvendy [15] indicated, from 5 to 10 users participating in the evaluation is enough to detect at least 80% of the usability issues in software. The Mobile Heuristic Evaluation (MHE) method [2] is based on a study in which each expert checks whether each application meets or not a set of directives for usability, which includes directives about usability features of the application that the expert has to answer, expressing their opinion with a numeric value from 0 to 4 (where 0 indicates that there is no problem, and 4 indicates a catastrophic problem), which is known as Nielsen's five-point Severity Ranking Scale [16], and they also had to justify their scores. The eight heuristics used were [2]: A (visibility of system status and losability / findability of the mobile device), B (match between system and the real world), C (consistency and mapping), D (good ergonomics and minimalist design), E (ease of input, screen readability and glancability), F (flexibility, efficiency of use and personalization), G (aesthetic, privacy and social conventions) and H (realistic error management).

The results of the evaluation are shown in Table 2. Applications with lower values (i.e., more usable applications) had mostly cosmetic problems (small obstacles) or no problems. On the other hand, applications with higher values had mainly minor and major problems, obstacles that affect the functionality of the application in a regular use.

TABLE II. RESULTS OF HEURISTIC EVALUATION ON APPLICATIONS

Mobile usability heuristics results									
App	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**Total**
#1[a]	0.08	0.28	0.92	0.67	0.13	0.42	1.58	0.00	**4.08**
#2[b]	0.92	0.61	0.67	1.00	0.23	1.17	0.50	0.39	**5.48**
#3[c]	0.00	1.17	1.67	1.42	0.70	0.50	1.25	0.89	**7.59**
#4[d]	2.54	2.17	1.75	0.58	1.03	1.50	0.50	0.61	**10.69**
#5[e]	2.63	1.61	2.08	1.17	1.33	1.25	1.67	1.00	**12.74**
#6[f]	0.00	2.00	1.67	1.92	1.07	2.08	3.17	0.89	**12.79**
#7[g]	1.00	2.39	1.75	2.08	1.57	0.67	2.33	1.44	**13.23**
Mean	**1.02**	**1.46**	**1.50**	**1.26**	**0.87**	**1.08**	**1.57**	**0.75**	**9.51**

a. WhatsApp Messenger
b. HushHushApp
c. Hiapp Messenger

 ISBN: 978-1-61208-438-1

d. Surespot encrypted messenger
e. Kik Messenger
f. Touch
g. Hike Messenger

The mean of heuristics show that G heuristic is the one that got the worst score (Table 2), due to a (generally) bad interface design and the lack of privacy and security information. The second worst heuristic was C heuristic, mainly because in some apps some objects were not expected on the interface. On the other hand, H heuristic got the best results, thanks to the ease on editing incorrect inputs and also the ease on recovering from errors. The E heuristic was the second best rated because of the ease on entering numbers, as well as it shows a back button on the screens and (mainly) the ease on navigation through the screens.

Finally, it is worth mentioning that a low number of interactions does not necessarily imply that the application has not usability problems. For instance, Hike messenger had 24 interactions and was the second best app on KLM results, whereas it was the worst app according to the MHE results. This implies that both techniques should be applied in order to evaluate the usability of an application.

III. DISCUSSION

In this section, the results and analysis carried out will be discussed. Firstly, we could compare the results obtained with those from two similar studies performed using a similar method: one for spreadsheet apps [13] and another one for diabetes management apps [12].

Step 1 (potentially relevant applications) produced 23 spreadsheet apps, 231 diabetes apps and we found 243 IM apps. The low number of apps in the spreadsheet study may be due to a more concrete term. Step 2 (delete light or old versions) discarded 9 (4.05%) diabetes apps and we discarded 20 (8.97%) IM apps. This variation may be due to IM apps are more popular. The analysis for spreadsheet apps does not indicate the number of discarded apps in step 2. Step 3 (identify main functionalities) got 12 (52.17% from step 1) spreadsheet apps, 8 (3.46%) diabetes apps and we obtained 39 (16.04%) IM apps. This difference can be due to the increasing number of applications or to the main functionalities chosen (different for each type of application).

Step 5A (KLM analysis) revealed that, in average, the tasks in spreadsheet apps took between 2.1 and 4 interactions, in diabetes apps they took between 3.16 and 6.33 interactions and in IM apps they took between 4.2 and 7.8 interactions.

Regarding the MHE (Step 5B), the study on spreadsheet applications was not performed. In diabetes apps, the main usability issues detected were related to heuristics G and H, and related to IM apps were about heuristics B, C and G. These results on the heuristics suggest that the design is generally not good and it can be improved.

The methodology used has a number of advantages [4, 11], but it also has some disadvantages and limitations, for instance: some steps take a long time and can be tedious to perform, large number of elements in the initial stages, results are time sensitive, etc. Furthermore, detecting all usability issues in this kind of experiment is not possible because the context of use is not taken into account [3] [14] but, on the contrary, as a laboratory experiment there is more control over the usability issues detected. An important limitation of this study is that it was conducted only on the iOS platform. On other platforms, different results could be obtained.

IV. CONCLUSIONS AND FUTURE WORK

According to the results of the KLM, sending a message, replying to a message and adding contacts are usually the fastest functionalities to be completed. Moreover, deleting a contact or a chat usually becomes a serious problem. The applications with lower levels of interactions were (from lowest to highest) Surespot encrypted messenger, Hike Messenger, HushHushApp, Kik Messenger, Touch, Hiapp Messenger and WhatsApp Messenger.

Regarding the Mobile Heuristic Evaluation with mobile experts, almost all applications had usability problems in performing the primary tasks. WhatsApp Messenger and HushHushApp obtained the best usability ratings. On the other hand, hike Messenger, Kik Messenger and Touch were negatively evaluated and presented critical usability, but did it well in KLM. This suggests that it is necessary to perform both the KLM and Heuristic Evaluation methods because if the results were based only in KLM the applications chosen would have many usability problems.

After finishing the study, we can propose some recommendations to improve the usability of instant messaging apps in mobile devices. Firstly, based on the KLM results, we can suggest the following recommendations:

- Each task should not exceed more than 5 or 6 interactions. It was observed that more than 8 interactions cause confusion in performing a task.
- Specifying the ID of a contact (username, phone number or email) should be enough for adding a contact. Other options (extra data such as name, last name, location, etc.) should be optional.

The heuristic evaluation results led us to propose the following guidelines:

- The interface should be carefully designed to ensure that all elements of the app are properly displayed in any position.
- Do not tolerate unrecoverable errors. It is always better displaying an error message than an unexpected shutdown of the app.

The usability recommendations proposed are a valuable resource for mobile app developers because they will improve the usability of their IM apps in mobile devices, thus achieving more downloads and users of their apps. As a future work, a new analysis will be carried out on other existing mobile platforms (e.g., Android) to compare results. Finally, after that we are planning to develop a mobile instant messaging application meeting the recommendations proposed, which will solve the main usability problems identified in existing applications.

 ISBN: 978-1-61208-438-1

ACKNOWLEDGMENT

Authors want to acknowledge support from the University of Alcala (TIFYC and PMI research groups).

REFERENCES

[1] M. L. Smith, R. Spence, and A. T. Rashid, Mobile phones and expanding human capabilities. Information Technologies & International Development, 7(3), 2011, pp. 77-88.

[2] E. Bertini, et al., "Appropriating heuristic evaluation for mobile computing," International Journal of Mobile Human Computer Interaction (IJMHCI), 1(1), 2009, pp. 20-41.

[3] D. Zhang, and B. Adipat, "Challenges, methodologies, and issues in the usability testing of mobile applications," International Journal of Human-Computer Interaction, 18(3), 2005, pp. 293-308.

[4] C. Martin, D. Flood, and R. Harrison, "A Protocol for Evaluating Mobile Applications," Proceedings of the IADIS, 2011.

[5] E. Abdulin, "Using the keystroke-level model for designing user interface on middle-sized touch screens," in CHI'11 Extended Abstracts on Human Factors in Computing Systems. ACM. 2011, pp. 673-686.

[6] D. Kieras, Using the keystroke-level model to estimate execution times. University of Michigan, 2001.

[7] B. A. Nardi, S. Whittaker, and E. Bradner. "Interaction and outeraction: instant messaging in action," in Proceedings of the 2000 ACM conference on Computer supported cooperative work.. ACM. 2000, pp. 79-88.

[8] R. E. Grinter, and L. Palen. "Instant messaging in teen life," in Proceedings of the 2002 ACM conference on Computer supported cooperative work. ACM. 2002, pp. 21-30.

[9] C. Lewis, and B. Fabos, "Instant messaging, literacies, and social identities," Reading research quarterly, 40(4), 2005, pp. 470-501.

[10] B. S. Gerber, M. R. Stolley, A. L. Thompson, L. K. Sharp, M. L. Fitzgibbon, "Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study," Health informatics journal, 15(1), 2009, pp. 17-25.

[11] C. Martin, et al., "A systematic evaluation of mobile applications for diabetes management," in Human-Computer Interaction–INTERACT 2011. Springer, 2011, pp. 466-469.

[12] E. Garcia,, C. Martin, A. Garcia, R.Harrison, D. Flood, "Systematic Analysis of Mobile Diabetes Management Applications on Different Platforms," Information Quality in e-Health, 2011, pp. 379-396.

[13] D. Flood,, R. Harrison, C. Martin, K. McDaid, "A systematic evaluation of mobile spreadsheet apps," in IADIS International Conference Interfaces and Human Computer Interaction. 2011.

[14] J. Bastien, "Usability testing: a review of some methodological and technical aspects of the method," International Journal of Medical Informatics, 79(4) , 2010, pp. e18-e23.

[15] W. Hwang, and G. Salvendy, "Number of people required for usability evaluation: the 10±2 rule," Communications of the ACM, 53(5), 2010, pp. 130-133.

[16] E. DIN, 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)–Part 11: Guidance on usability. International Organization for Standardization, 1998.

 ISBN: 978-1-61208-438-1