



Mobile User Experience Beyond the Laboratory: Towards a Methodology for QoE-QoS Evaluation in Natural User Environments (A Position Paper)

Katarzyna Wac and Anind K. Dey

Wi: Journal of Mobile Media, Spring 2011

The online version of this article can be found at:

<https://doi.org/10.65968/XCAS7893>

Wac, Katarzyna and Anind K. Dey. "Mobile User Experience Beyond the Laboratory: Towards a Methodology for QoE-QoS Evaluation in Natural User Environments (A Position Paper)". *Wi: Journal of Mobile Media, Spring* (2011). Web. <https://doi.org/10.65968/XCAS7893>

Mobile User Experience Beyond the Laboratory: Towards a Methodology for QoE-QoS Evaluation in Natural User Environments (A Position Paper)

Katarzyna Wac and Anind K. Dey

Introduction

The growing availability of diverse interactive mobile applications, envisaged to assist us in different domains of our daily life, make their perceived quality of experience (QoE) increasingly critical to their acceptance. Comments such as, “If it’s slow, I won’t give my credit card number” indicates the QoE expectations of a typical mobile commerce application user [1]. These expectations can be different given the user’s previous experiences with an application or an application’s criticality to the user’s task at hand. To date, the evaluation of QoE has been mainly conducted with qualitative methods that focus on an applications’ usability [2]. These studies are typically conducted within a limited time span in controlled laboratory environments, conditions that do not resemble users’ natural daily environments. The results of such evaluations may help to discover the mobile application’s serious and immediate usability issues in a design, but they may not help to uncover issues that are relevant to real life situations in the world outside the lab.

Issues relevant to the world outside the lab involve, amongst other things, a non-deterministic quality of service (QoS), and, in particular, the performance of the

underlying network infrastructures that support the execution of an application and mobile service delivery (Figure 1). This quality of service can be quantified by measuring delay, jitter and network capacity. This quality of service usually is provided at a 'best-effort' level; that is without any guarantees by a company or service provider that it will work at an optimum level. Yet, as we would argue, this quality of service is critical to the mobile user's quality of experience, especially for highly interactive mobile applications, delivery of which depends on frequent data transfers over the underlying network infrastructures.

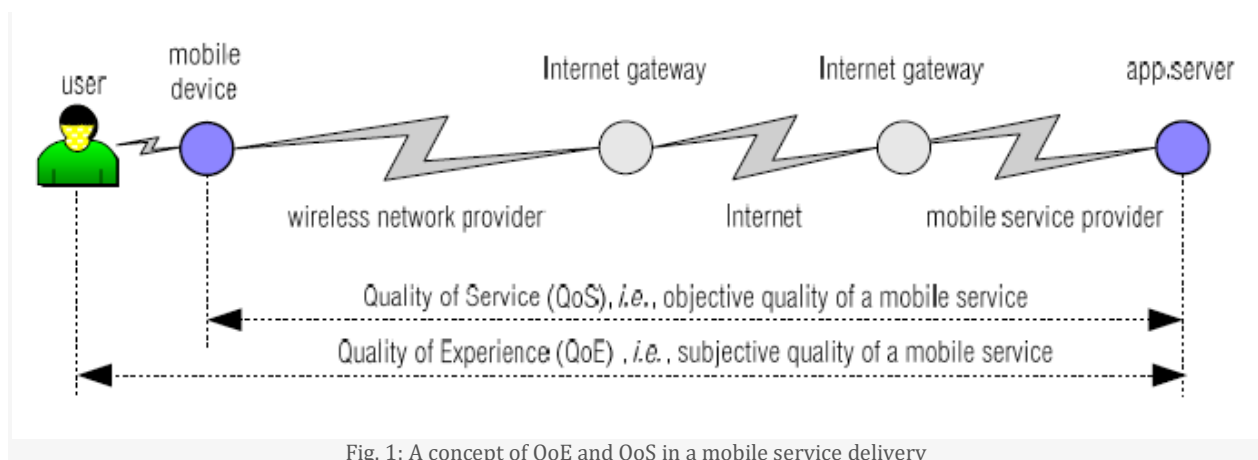


Fig. 1: A concept of QoE and QoS in a mobile service delivery

Fig. 1: A concept of QoE and QoS in a mobile service delivery

A common practice for quality of experience provision is that mobile application designers use their own judgment and perception of an application's ease of use as a bellwether to gauge an application's perceived quality of experience by an imagined mobile user [2]. But can the designer, who may have created the system and thus have an intimate knowledge of its capabilities and embedded logic, really stand in for the imagined user? The overall effect of this situation is that users whose QoE expectations

are not satisfied, may simply stop using the applications or switch to another provider. For example, it is estimated that there are, on average, 200 new applications available daily in the online store for Apple's iPhone platform. However, due to the low quality of experience, more than half of these applications do not achieve a critical mass of user acceptance and are withdrawn from the store's list of offerings within some months from the launch.

The challenge for designers and researchers studying these technologies and new applications is that no rigorous and robust scientific methods, tools, and systems exist for evaluating an application's perceived QoE in the user's natural environment(s) [3]. Rather, there are separate methods for usability evaluation in the HCI community [2, 4] and separate methods for the evaluation of the quality of service and performance of an application's underlying network infrastructures in the data networking community [5-7]. The former methods are largely qualitative, while the latter are largely quantitative. Both methods can acquire quality results in their dedicated areas of applicability. However, due to the dichotomy between these two scientific communities, there are no scientifically proven methodologies that combine both of these approaches.

Objective and Approach

Given this chasm between methodologies, approaches and the study objects, the focus of our research into these methods is to find a viable way to bridge this gap. To this end, we are in the process of developing a set of methodological procedures for the reliable real-time evaluation of an interactive mobile applications' perceived quality of experience in

user's natural environments. This investigation of the quality of experience will be conducted in conjunction with an assessment of the variable quality of service provisions. In our approach we will focus on already implemented and operational interactive mobile applications, which are now available to a typical mobile user. We assume that these applications have undergone a cycle(s) of (re)design and usability tests in a laboratory environment, although we do not necessarily have access to the results of these.

Our approach in this paper is as follows. We identify and analyze existing and emerging qualitative methods for the evaluation of usability, as well as quantitative methods for the evaluation of QoS and performance of mobile computing applications. Based on these methods, we propose a novel set of procedures for the real-time quantitative evaluation of users' perceived QoE of mobile applications in natural settings. This methodology is based on our long-standing successful history of research on measurements-based quality of service and performance evaluation methods for interactive mobile applications [8, 9]. We have successfully used this methodology in a healthcare domain, specifically in the creation of interactive applications for health telemonitoring and teletreatment depending on delay and capacity (i.e., quality of service metrics) of the underlying network infrastructures [5, 10, 11]. What is new is our desire to transfer, expand towards QoE and test this methodology in new areas and domains of mobile experience.

To quantify the mobile user's quality of experience, the methodology first requires defining a set of dependent (i.e., target) variables. We then define a set of mutually exclusive and collectively exhaustive variables influencing this quality of experience. These are the independent variables that can include, for example, user context like

location, time, social settings, etc. Both sets of variables should be based on the existing scientific literature and documented expert knowledge.

Furthermore, for a given interactive mobile application, the methodology requires a set of qualitative methods to derive new independent variables that are not indicated in the HCI or the networking communities so far, but are important to the experience a mobile user has in her natural daily environments. One qualitative method that can be used for this purpose is the Experience Sampling Method (ESM) [12]. The ESM is based on occasional user surveys, which can be administered over specific time intervals, after particular events, or at random. Since we aim to evaluate a user's perceived quality of experience while interacting with a mobile application, the ESM could be implemented in the form of a short, mobile-device based survey given to the users after each use of this application. The survey will pose some open-ended questions to get the user's in-situ real-time, spontaneous opinions on their mobile experience. These new independent variables will be 'grounded', as they are derived from the answers acquired from this user [13, 14]. The ESM method must be designed and deployed such that it does not influence the experience and behaviour of a mobile application user, but enables us to gather information that is relevant and predictive for this user's quality of experience evaluation.

As the evaluation will be conducted in the user's everyday environments, the methodology must provide requirements and guidelines for the effective and efficient implementation and application of (software) modules necessary for measurements of the mobile application usage, the QoS and performance of its underlying service and network infrastructures. In this way the state of these variables (including those generated by the

ESM), is continuously and accurately logged (i.e., measured) in real-time in an automatic manner; that is, in a way that is non-intrusive to the mobile user.

Moreover, having defined sets of dependent and independent variables, and having the modules implemented and operational for their measurement, the methodology would require reusing the existing analytical methods to discern relationships between variables and possibly establish causality. An example from our healthcare domain includes delineating boundaries for network delay and its jitter (i.e., independent variable), for which application data may have a clinical value for real-time diagnosis (i.e., as a part of dependent variable) provided to patients [5].

To analyze possible relations and causality between variables, the methodology requires the occasional involvement of a mobile user in the data analysis process. Namely, a mobile user needs to be interviewed about his application's usage patterns and experience. This data must then be matched to the data automatically logged in the application and service infrastructure. The interviews we propose to conduct will be based on the completion of a detailed diary of the previous 24-hour period, as suggested by the Day Reconstruction Method [15]. This breaks the day into episodes described by activities, locations and times, and the mobile application usage and experiences during these times. During the interview, users can explain, in more detail, their responses in the ESM, and these results will be compared to the state of other independent variables logged in the system. This way causalities and relations specific to this particular user can be identified, while any inconsistencies can be clarified.

The combination of these methodologies, which are both qualitative and quantitative, will then provide adequate guidelines on how to statistically analyze and interpret the acquired (qualitative) survey data and (quantitative) measurement data for analysis by a user for one or multiple interactive mobile applications. Focusing the analysis on a single user implies the idiographic approach putting the effort to understand the meaning of contingent, unique, and subjective phenomena of quality of experience state for this particular user. Given a population of users, an analysis could then be conducted within this for one mobile application, or between populations of users of different mobile applications. The subsequent data analysis might then be implemented using advanced statistical methods, such as a multivariate analysis, or machine learning techniques for pattern recognition in data. Example machine learning techniques include logic-based techniques (trees, rules), density-based techniques (Bayesian networks) techniques based on non-linear functions (neural networks and support vector machines), as well as so-called 'lazy' techniques ('k'-nearest neighbours). These techniques automatically learn to recognize and model complex patterns in the collected data, based on which they can predict the target variable, i.e., quality of experience state for a given application user in a given context.

Contributions and Evaluation

Our preliminary research on this topic, and approach, brings together and expands upon recent advances and methodologies in key scientific and technical areas, including the evaluation methods for human computer interaction, quality of service and performance

evaluation methods and tools for mobile computing, real-time machine learning and prediction. By laying out this set of principles and guidelines, in the future we will conduct research on critical issues such as:

1. The definition of quality of experience that is expected and required for interactive mobile applications. This definition must integrate multiple views including an examination of the application and its underlying infrastructure views, such as the interactions and provided QoS and performance, and the user view, such as past experiences and expectations, current application's perception and its criticality to the task at hand, as well as the user's context. The definition must also delineate a role of the user's affective response for an interactive mobile application use.
2. Implementing this method in future research involves a second challenge: that of gathering reliable real-time capturing of a user's perceived quality of experience and the parameters and conditions that influence this quality of experience as it is lived in their natural daily settings. This includes identifying the variable 'best-effort' state of quality of service for underlying service and network infrastructure.
3. The third aspect of our research in these areas will be to identify methods for documenting an automated and accurate inference of the user's quality of experience state based on the gathered data. An accurate inference of this state is challenged by the fact that a) quality of experience state may be temporally indirect, i.e., there may be a lag between a cause (e.g., user context change) and the change of the user's quality of experience state; b) sensors and modules used for real-time capturing of a user's quality of experience may be unreliable: their reading may be influenced by, e.g., a user's bodily position and movement artifacts;

and, c) there are individual differences in quality of experience perception and evaluation of state for each human.

4. The fourth challenge focuses on the accurate and real-time recognition of QoE patterns based on data mining and machine learning techniques. These challenges become even more complex if the system is required to be accurate and operational in real-time and to generalize to novel situations, e.g., in the case of novel mobile applications, or novel user's interaction patterns.
5. The final critical issue includes the ethical aspect of continuous real-time logging of possibly private information, implying that adequate security and privacy mechanisms must be in place for deployment of our methodology for a range of mobile application users. The users themselves must be informed about how and by whom the data is being collected, handled, stored, analyzed and used.

With use of the proposed complex multi-methodological approach, we hope to gain a deeper understanding of the use of interactive mobile applications, and to be able to quantify a user's quality of experience and her relationship with the underlying quality of service, and point out ways to improve the usability of these applications and generate higher user acceptance. On one hand, application providers could use such methods to improve the provided applications, and, on the other hand, consumer advocacy groups could use the methods to monitor the quality of provided applications. It is our intention to test our methodology on a set of widely available mobile applications for leisure, entertainment, communication or information, whose users expect to be able to easily stream multimedia content, such as YouTube, or use Internet-based radio; those engaging in highly interactive instant messaging or web browsing, such as e-banking or

e-commerce; those playing multiplayer online gaming. Finally, another area where quality of service should match expectations is in VoIP video-conferencing, such as Skype or Google Talk.

Concluding Remarks

In this paper we have presented our proposed research approach towards defining a methodology for quantifying a mobile user's experience (QoE) in their natural daily environments and relating this experience to the performance (QoS) of the underlying service and network infrastructures. This methodology is a blend of both quantitative and qualitative procedures. We propose a twofold methodology for evaluating user experience, where the user becomes an active participant in the research. First, it requires gathering in situ spontaneous information about the user's mobile experience by employing the Experience Sampling Method for interaction with the user directly after each mobile application usage. Second, it requires a retrospective analysis of the user's experience and of the state of factors influencing it, by employing the Day Reconstruction Method to assist with the recollection of the past 24-hours. While our current work focuses on defining the methodological steps, our future research includes an evaluation in a large-scale mobile user study for a set of widely used mobile applications.

Acknowledgments

Research conducted by K. Wac was sponsored by Swiss SSER (Co8.0025) and Swiss NSF (PBGEP2-125917). This work is also partially supported by and the US NSF FieldStream project (0910754).

References

- Bouch, A. et al. (2000). Quality is in the eye of the beholder: meeting users' requirements for Internet quality of service. in the *SIGCHI conference on Human factors in computing systems*. The Hague, The Netherlands.
- Bults, R. et al. (2005). Goodput Analysis of 3G wireless networks supporting m-health services. *18th IEEE International Conference on Telecommunications (ConTELO5)*. Zagreb, Croatia.
- Dix, A. (2003). *Human Computer Interaction* Prentice Hall.
- Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*.
- Hektner, J. M., et al. (2006). *Experience sampling method: Measuring the quality of everyday life*. Sage Publications, Inc.
- ITU-T (2008). Definitions of terms related to quality of service, *Recommendation E.800*.
- ITU-T (2001). Communications Quality of Service: A framework and definitions, *Recommendation G.1000*.
- Kahneman, D., Krueger, A., Schkade, D., Schwarz, N. and Stone, A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science* 306 (5702), 1776-1780.
- Kjeldskov, J. & Graham, C. (2003). A review of mobile HCI research methods. *Lecture Notes in Computer Science*(pp. 317-335).
- Martin, P. Y. & Turner, B. A. (1986). Grounded theory and organizational research. *The Journal of Applied Behavioral Science*. 22, 141.
- Michaut, F. and Lepage, F. (2005). Application-oriented network metrology: Metrics and active measurement tools. *IEEE Communications Surveys & Tutorials*, 2-24.
- Salamatian, K. and Fdida, S. (2001). Measurement Based Modelling of Quality of Service in the Internet: A Methodological Approach. *International Workshop on Digital Communications: Evolutionary Trends of the Internet*. (pp. 158-174).

Wac, K. et al. (2005). Measurements-based performance evaluation of 3G wireless networks supporting m-health services. *12th ACM Multimedia Computing and Networking Conference*. San Jose, CA, USA.

Wac, K. and Bults R. (2004). Performance evaluation of a Transport System supporting the MobiHealth BANip: Methodology and Assessment. *MSc Telematics*. University of Twente, Enschede, the Netherlands.

Biography

Anind K. Dey is an Associate Professor in the Human-Computer Interaction Institute at Carnegie Mellon University. His interests lie at the intersection of human-computer interaction, machine learning and ubiquitous computing. He has spent the last decade developing techniques for building context-aware applications, and for improving the usability of such applications. He is very interested in the development of truly smart systems, systems like sensor-enabled phones that can autonomously collect a vast amount of information about users and use that information to improve user experiences. Anind is the author of over 100 articles in the area of ubiquitous computing, has served as the Program Chair for several conferences on ubiquitous computing and serves on the editorial board for IEEE Pervasive Computing, Personal and Ubiquitous Computing Journal and the Journal of Ambient Intelligence and Smart Environments. Before joining Carnegie Mellon University, Anind was a Senior Researcher at Intel Research Berkeley and an Adjunct Assistant Professor at the University of California-Berkeley. He holds a PhD and a Masters degree in Computer Science, as well as a Masters degree in Aerospace Engineering, all from Georgia Tech, and a Bachelors of Computer Engineering from Simon Fraser University

Katarzyna Wac is a senior computer scientist currently associated with Institute of Services Science at University of Geneva (Switzerland). In 2009-2010 she has visited the Human-Computer Interaction Institute at Carnegie Mellon University. In 2003 she has received a BSc and MSc degree in Computer Science from Wroclaw University of Technology (Poland), and in 2004 MSc in Telematics from University of Twente (the Netherlands). In 2009 she has defended her PhD thesis in Information Systems at University of Geneva. Her research focuses on measurements-based methodologies for an evaluation of performance of interactive mobile applications and its relation with the end-user perceived experience. She builds tools that predict application's performance and hence facilitate development of mobile computing applications that improve end-user perceived experience.

