Quality of Experience in Mobile Applications: 
A Systematic Mapping of Metrics and Tools

https://doi.org/10.3991/ijim.v14i08.12819

Antonio Collazo Garcia (✉)
Centro de Investigación y Transferencia (CIT Santa Cruz) (CONICET-UNPA-UTN),
Rio Gallegos, Argentina
acollazoq11@gmail.com

Sandra Casas
Universidad Nacional de la Patagonia Austral (UNPA), Rio Gallegos, Argentina

Abstract—Context: Quality of Experience (QoE) enables the description of user perceptions about the performance of a particular application or service. In the mobile computing context, it is an important measure for service providers and users, since QoE makes it possible to improve it and make it more competitive to achieve user fidelity. In turn, the importance of QoE in mobile technologies increases due to the various factors that affect the applications that run on mobile devices.

Objective: The purpose of this study is to identify the metrics and tools relevant to the scientific community for the QoE analysis of mobile applications.

Method: A systematic mapping study was conducted.

Results: From a total of 751 studies, 33 were selected, and 13 metrics and 15 mobile QoE analysis tools were identified.

Conclusions: The existing mobile QoE analysis tools collect and calculate metrics automatically, combining objective and subjective metrics. However, they present limited approaches, making it difficult to carry out an integral analysis of the applications.

Keywords—Quality of experience, mobile applications, systematic mapping, metrics, tools.

1 Introduction

Mobile applications and services are increasingly present in most aspects of daily life, satisfying information, communication and entertainment needs. To achieve the acceptance of users in the use of mobile applications, it is necessary to consider the Quality of Experience (QoE), which allows to measure the quality of an application based on the user perception [1], as well as different factors that affect such quality.

Different concepts about QoE can be found in the bibliography, in [2] QoE is defined as the general acceptance of an application or service according to the subjective perception of the end user, being influenced by the context and user expectations. According to [3] QoE is the measure of the user performance based on objective and
subjective psychological measures about the use of an application or service. QoE represents the level of acceptance that users perceive when making use of an application or service, presenting as main characteristics subjectivity, user centeredness, integrality and multidimensionality [4].

QoE is an important measure in the context of mobile computing due to the characteristics of mobile devices: heterogeneity of devices and operating systems, types and functionality of applications, user profile, mobility, context, network connectivity and infrastructure. This leads to the appearance of multiple factors that directly intervene in the quality of mobile applications, making it necessary to take into account these factors for the analysis of the QoE in this type of devices.

Therefore, QoE in mobile applications constitutes a relevant issue for the scientific community, mobile application development communities, service providers and users, evidencing a growth in research related to this topic. This study focuses mainly on two aspects of QoE, the first is the theory to analyze it (QoE metrics) and the second is the software used to perform its analysis (evaluation, calculation, estimation, measurement, etc.) [1].

QoE metrics are classified as objective and subjective. Subjective metrics are related to the users opinion, evaluating the quality of the application or service based on their experience with the application [4]. This depends on subjective factors such as context and user expectations about the application. In general, subjective metrics are based on conducting surveys and questionnaires to the users. This process presents the difficulty of being costly in terms of time and money [5], [6].

Regarding objective metrics, its systematic nature stands out, they are exact and repeatable, and refer to different properties such as data presentation time, consumption of mobile data, energy consumption, user interface and content presentation [4].

Several investigations analyze the QoE in close relation with the Quality of Service (QoS), while QoS degradation causes an unacceptable QoE. The QoS parameters related to network performance such as delay, instability, loss rate, error rate, bandwidth and signal success rate are used to determine the QoE value, depending only on the calculation of the QoS [1].

The objective of this study is to carry out a systematic mapping study on QoE metrics and tools for mobile applications, identifying the most researched metrics and their application in the use of tools for QoE analysis. The paper is organized as follows: Section 2 describes the methodology used; Section 3 presents the results obtained; Section 4 analyzes and discusses the results of systematic mapping; and finally, in Section 5 the conclusions of the study are presented.

2 **Methodology: Systematic Mapping Study**

The systematic mapping technique allows us to review and categorize information related to a specific topic or area of interest. The objective of a systematic mapping is to determine the scope of the research conducted on a specific research topic and to classify knowledge [7]. Figure 1 presents the process that has been followed for the systematic mapping.
2.1 Definition of research questions

The research questions focused on identifying the most commonly used metrics and existing tools to analyze the quality of experience of mobile applications. Thus, the following questions for systematic mapping were raised:

1. What metrics have been proposed in the literature for the analysis of QoE in mobile applications?
2. How have these QoE metrics been applied?
3. What tools (framework, systems, algorithms, applications, etc.) have been proposed for the QoE analysis of mobile applications and what are their characteristics?

2.2 Conduct search

One search was conducted using ACM, IEEE, Springer and ScienceDirect as main data sources, and another search was made following the references in the found articles. The following keywords were used to perform the search: quality of experience, metrics and mobile applications. A small set of documents was obtained by forming a search string with these keywords. It was necessary to adjust the final search string to obtain more comprehensive results. Finally, the following search string was formed: “(QoE OR quality of experience) AND (metrics) AND (mobile)”. After executing the search on data sources and the reference review, a total of 751 articles were obtained.

2.3 Screening of papers

The selection process was done by analyzing the title, summary and keywords of the articles. The following inclusion criteria were considered:

- Studies carried out since 2008.
- Studies published in English.
- Studies on QoE in mobile applications.
- Studies on tools for the analysis of mobile QoE.

And the following exclusion criteria:

- Studies in the form of summaries or presentations.
- Studies whose main focus was not related to the mobile computing area.
2.4 Keywording of abstracts

The relevant information regarding the systematic mapping process was oriented mainly towards determining the elements related to the quality of experience in mobile applications:

- Quality of experience (QoE).
- Quality of service (QoS).
- Mobile application.
- Metrics.
- Tools.

These elements and the related aspects were considered for the grouping and classification of the studies using the classification scheme described below.

2.5 Classification scheme

Once the relevant articles were selected, a first classification was defined based on the contents of the study (Fig. 2):

- Dimension of Metric: empirical studies in which a set of metrics for the analysis of QoE in mobile applications are analyzed.
- Dimension of software: studies in which tools, frameworks, systems and/or applications are developed and/or applied to analyze QoE in mobile applications.

A second classification considered the types of metrics analyzed (Fig.2):

- Dimension of objective metrics: studies focusing on objective metrics.
- Dimension of subjective metrics: studies focusing on subjective metrics.

Fig. 2. Classification scheme
2.6 Data extraction and mapping process

After defining the classification scheme, the last step of the systematic mapping consists in the extraction of data and the process of mapping the different dimensions. To carry out this process, it was necessary to review the content of the selected articles. The results of this last stage are shown in the next section.

3 Results

This section shows the results of the mapping process and the data extraction from the selected studies in order to answer the research questions. Table 1 and Table 2 show the number of studies associated with the dimensions of the classification scheme, grouping the studies by the type of content and the type of metrics analyzed.

In the dimensions of the second classification an overlap of the studies in both dimensions can be observed. This is due to the fact that several articles address both the analysis of objective and subjective metrics.

Figure 3 illustrates a bubble chart with the dispersion per year of the studies selected according to the classification scheme. The size of the bubble is proportional to the number of items that are located in the intersection of each dimension with the corresponding year.

<table>
<thead>
<tr>
<th>Table 1. Number of articles by type of content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
</tr>
<tr>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Dimension of metrics</td>
</tr>
<tr>
<td>Dimension of software</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Number of articles by type of metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
</tr>
<tr>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Dimension of objective metrics</td>
</tr>
<tr>
<td>Dimension of subjective metrics</td>
</tr>
</tbody>
</table>

Fig. 3. Distribution of studies per year in each dimension
3.1 Data extraction and mapping process

In order to answer the first and second research question, the quality of experience metrics analyzed in the selected articles were identified, yielding a total of 13 metrics of QoE. Each of the identified metrics is presented below.

Latency [1], [4], [8]–[17]: It measures the length of time it takes for a packet to travel from the origin to the destination. It is measured by sending a packet that has a response. The latency is measured by the time it takes from the moment the packet is sent until the response is obtained.

Signal instability [4], [10], [13], [18], [19]: It defines the variation in the arrival time of the packets, caused by network congestion, loss of synchronization or by the different routes followed by the packets to reach the destination. One of the best-known ways to determine the instability of the signal consists of the average of the delay variation between the received packets.

Signal strength [12], [13], [15], [16], [20]–[22]: It is a reference scale for measuring the power level of the signals received by a device in wireless networks. The signal strength refers to the received intensity and not to the signal quality.

Bandwidth-Throughput [4], [8], [13], [17], [20]–[25]: Technically, throughput is the information capacity that a network element can transmit in a certain period of time. The bandwidth is the available theoretical capacity of a link and the throughput is the actual usage level of the link. To measure it, the Iperf tool can be used by sending data streams, indicating the amount of data transferred and the throughput measured.

Mobile data consumption [5]: It corresponds to the volume of data used by an application using mobile data, which can mean a cost to the user. One way to measure data consumption is by determining the volume of data used over a period of time by a particular application or service.

Energy consumption [4], [5], [12], [13], [22], [24], [26]–[29]: It is a factor that constantly influences the user experience on mobile devices, since it limits the use of the device, especially when the battery is completely discharged. In mobile devices it can be measured in percentage or in milliamperes-hour (mAh).

CPU consumption [12], [13], [22]: It describes the capacity of the processor used by an application. The CPU consumption of a mobile device may vary depending on the types of tasks that an application performs. This value is measured in percentage.

Memory consumption [12], [13]: It is the amount of memory (RAM) used by an application when it is running. It is used to store the internal data and instructions of the application. The memory limits the number of applications that the users can execute and the amount of data with which they can work.

Packet loss [4], [8], [10], [12], [13], [15], [17], [18], [20], [22]: It occurs when one or more data packets traveling through a computer network do not reach their destination. It is usually caused by network congestion and it is measured as a percentage of lost packets with respect to the total number of packets transmitted.

User perceived latency [5], [26], [30] [32]: It is a metric of the application level. It is defined as the time it takes the application from the moment that the user starts an action until the data is displayed in the interface.
Audio quality [8], [15], [18], [20]: It is the sound quality perceived by the user. This parameter can be analyzed subjectively. However, it can be determined by some properties of the audio file (Kb/s, signal/noise, frequency (Khz), etc.).

Video quality [4], [10], [14], [15], [18], [20], [24], [25], [33] [35]: Just like the audio quality, it is a subjective parameter that may depend on the user. However, there are several properties of a video file that allow for the automatic determination of the quality of the video (complexity of the scene, movement level, image quality, corrupted or missing data, etc.).

User opinion [1], [4], [8], [10], [12] [15], [18], [19], [21], [23], [33], [34], [36]: It is a subjective parameter for the analysis of the quality of experience. It indicates the level of general satisfaction of users regarding an application. The most common method is through the mean opinion score (MOS) representing the quality perceived by the user translated into values from 1 to 5 (a higher value means higher QoE).

3.2 QoE tools

Below is the list of QoE tools identified in the selected articles. The type of software, its objective (analysis, monitoring and prediction) and its description are defined. A total of 15 tools were identified, responding to the third research question.

QoE Doctor [5]: It is a software tool for the analysis and monitoring of QoE in mobile applications without the need to access the source code of applications. It consists of the implementation of a mobile application for the analysis of the main causes of QoE problems across the application, transport, network, and cellular radio link layers. It makes use of user interface automation techniques to reproduce user behavior related to QoE, and measures the latency perceived by the user directly from changes in the user interface.

Prometheus [20]: It is a monitoring and prediction system of QoE in real applications of video on demand and VoIP. It consists of implementations in both the mobile device and the service operator. It estimates QoE metrics that address the specific challenges of cellular operators of lack of control, limited points of view and complexity of the protocol. It does not require control over the services of the target application and makes QoE predictions using only passive network measures.

AppInsight [30]: It is an analysis and monitoring system that helps mobile application developers to diagnose the performance of their applications. It consists of the communication between an implementation on the mobile device and a web server. It collects tracking data to discover critical routes and exception routes in user transactions, indicating the optimizations needed to improve the user experience to developers. It is light and does not require any modification of the operating system, or any contribution from the developer.

Proteus [17]: It is a QoE monitoring and prediction framework that passively collects information about network performance observed by applications. It uses regression trees to decipher network performance patterns, and forecasts future network performance to benefit the application's performance. Proteus can predict the occurrence of packet loss, the occurrence of a long delay and the performance of the network with an error of 10kbps with an average performance range of 100 to 800 kbps.
Timecard [31]: It is a system of analysis, monitoring and prediction that helps manage end-to-end delays for interactive mobile server-based applications. It consists of the communication between an implementation on the mobile device and an API on a web server. It provides two abstractions: the first returns the time elapsed since the user started the request, while the second returns an estimate of the time it would take to transmit the response from the server to the client and process the response at the client. For any user transaction, Timecard tracks the elapsed time and predicts the remaining time, allowing the server to adjust its working time to control the end-to-end delay of the transaction. Timecard incorporates techniques for tracking delays in multiple asynchronous activities, managing the time bias between the client and the server, and estimating network transfer times.

Panappticon [32]: It is a system of services and applications for the analysis and monitoring of QoE through event tracking related to the application in the operating system and framework libraries. It correlates the related events and identifies the individual transactions perceived by the user. Panappticon determines the duration and critical path of each transaction, which helps to clarify the root causes of performance bottlenecks. It monitors application software, system and kernel layers. It can identify performance issues resulting from application design flaws, low power hardware and harmful interactions between seemingly unrelated applications.

PowerTutor [28]: It is a mobile monitoring application to show the power consumed by a set of system components such as CPU, network interface, screen, GPS and other applications. It receives the current values in mA from the controller and then multiplies the value by the voltage that is basically the battery of the smartphone. PowerTutor calculates the energy consumption of applications and services based on processing times and it is only available for specific types of phones.

YoMoApp [25]: It is a mobile analysis and monitoring application for the Android platform that accurately reproduces the behavior of YouTube service in order to monitor and store passively the key performance indicators of YouTube's adaptive video streaming on smartphones. It monitors user events, buffer level and video quality. The monitored data is used to analyze YouTube's mobile QoE.

IVQA [35]: It is an algorithm for the monitoring and prediction of instantaneous video quality. It presents a constant execution time appropriate for its implementation in real time applications on light devices. It is based only on the parameters that are calculated during the video encoding process.

ARO [22]: It is an analysis and monitoring tool that works through a mobile application and a desktop application. It exposes the interaction between the state of the radio resource channel and the transport, application and user interaction layers in order to reveal the inefficient use of resources by mobile applications. ARO helps developers identify resource usage inefficiencies and improve their applications.

Mobile Agent [12], [13]: It is a mobile application with three monitoring entities that check different aspects of a mobile application: quality of service, contextual and experience monitoring. This tool is the fundamental component of an architecture that runs in laboratories with controlled environments.

QX-probe [26]: It is a comparative and quantitative QoE analysis and monitoring tool to identify application adjustment points. It is based on the execution of a service
on the mobile device that communicates with a web server. QX-probe defines the latency perceived by the user and the energy consumption as critical factors for the analysis of QoE at user interface level. It provides a web tool for the analysis of QoE and the detection of adjustment points.

Agboma & Liotta [10]: It is a management framework for the analysis and prediction of QoE for different types of multimedia content on three types of mobile devices: mobile phones, personal digital assistants and laptops. It uses a statistical modeling technique that correlates the QoS parameters with the estimates of users’ QoE perceptions.

ExBox [9]: It is a middlebox based on a hardware device and a software. It can learn the experiential capacity of a network through the estimation of QoE and machine learning. ExBox performs the analysis and monitoring of QoE metrics in the application and uses light machine learning techniques that are designed for dynamic wireless environments.

Keytko et al [18]: It is a framework for the analysis and monitoring of mobile QoE in video transmission. It combines objective and subjective parameters to evaluate user experiences. It performs the measurements of network parameters on the server and user feedback on the device through a mobile application.

After having the list of metrics and software tools for the analysis of mobile QoE, Table 3 was designed, showing the identified tools in relation to the metrics considered for the analysis of the quality of experience.

Table 3. Relationship between tools and metrics of mobile QoE

<table>
<thead>
<tr>
<th>Tools / Metrics</th>
<th>Latency</th>
<th>Signal strength</th>
<th>Bandwidth</th>
<th>Mobile data consumption</th>
<th>Energy consumption</th>
<th>CPU consumption</th>
<th>Memory consumption</th>
<th>Packet loss</th>
<th>User perceived latency</th>
<th>Audio quality</th>
<th>Video quality</th>
<th>User opinion</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>QoE Doctor</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Prometheus</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>AppInsight</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Timecard</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Panappticon</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Mobile Agent</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>IVQA</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>QX-probe</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Agboma &amp; Liotta</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>YoMoApp</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>5</td>
<td>33</td>
</tr>
</tbody>
</table>

4 Discussion

According to the classification scheme defined in the methodology, the articles belonging to the dimensions of the first classification (Table 1) present a balanced distribution. Therefore, those dedicated to QoE metrics (17 articles) and those focusing on the development and/or application of tools, frameworks, systems and applications for QoE analysis (16 articles) have been addressed similarly by the scientific community. However, in the second classification (Table 2), it is observed that the objective metrics (30 articles) present greater interest of study and application by the scientific community.
community with respect to the subjective metrics (20 articles). Furthermore, it is clear that several of the studies address subjective metrics such as the opinion of users to analyze QoE, in order to make comparisons and validations of QoE analysis methods based on objective metrics.

The monitoring of values obtained from QoE metrics is a common feature in the tools identified. Another characteristic present in most of the tools was the subsequent analysis of the collected data associated with the calculation of the metrics. However, the prediction of future QoE values is only present in a minority of tools, which are mainly oriented to the video quality analysis of applications and multimedia services.

As seen in Table 3, the metrics most commonly used by QoE tools include: bandwidth, power consumption, packet loss, and latency perceived by the user. On the other hand, the least frequently used metrics correspond to mobile data and memory consumption, and audio quality.

The identified software tools present a wide dispersion of the metrics used to analyze QoE in mobile applications. On average, each of the metrics is analyzed in only three tools and at the same time each tool covers only three or four metrics of QoE. Therefore, if a user, developer, or network operator needs to analyze the QoE of an application or mobile service in order to identify failures and improve them, s/he cannot do it with a single tool. On the other hand, they offer little support when analyzing only certain quality attributes. Thus, multiple factors that affect the QoE related to the characteristics of the devices and the context are not taken into account. These tools present limited approaches, making it difficult to perform an integral QoE analysis in mobile applications.

5 Threats to Validity

A threat to validity of this study is whether a correct selection process was conducted. The research questions, inclusion and exclusion criteria were defined before the execution of systematic mapping in order to ensure an unbiased selection process. To improve validity and mitigate any partiality, the inclusion and exclusion of articles was decided jointly by the authors. However, the threats from a quality assessment perspective cannot be ruled out, since no scoring system was used for each article in the selection process.

Another threat to validity is the inclusion of all relevant articles in the area. Therefore, several main sources of data related to the area and the references of the studies found were taken into account. In addition, each of the studies returned by the data sources were analyzed, and it was necessary to adjust the first search string in order to obtain more comprehensive results. The classification scheme may also represent a threat to the validity of this mapping study. One of the problems with mapping studies is how to determine the correct way to categorize the resulting articles.
6 Conclusion

In the present research, a systematic mapping study was conducted on metrics and tools of mobile quality of experience, in which 33 articles relevant to the topic were selected from a total of 751 articles from different sources and references.

The preliminary results of the study allowed us to identify the main interests of researchers in the area of metrics and tools for QoE analysis in mobile applications. The principal proposed metrics in the selected relevant articles and their use in the context of mobile applications were identified. On the same lines, the tools proposed by the researchers for QoE analysis and their characteristics were identified. Differences and similarities among tools and their relationship with the metrics identified were determined.

Several tools were identified for the mobile QoE analysis that collect and calculate metrics automatically, combining both objective and subjective metrics. However, they present limited approaches which make it difficult to perform an integral analysis of the applications. This leads to the existence of a gap in the development of tools that integrate and combine mobile QoE metrics.

7 References


8 Authors

Antonio Collazo Garcia is a PhD student at “Universidad Nacional de la Patagonia Austral” (UNPA), Argentina. The research areas are Mobile Computing, User Experience, and Software Engineering. Email: acollazog11@gmail.com

Sandra Casas is the responsible researcher and associate professor of the Institute of Applied Technology at “Universidad Nacional de la Patagonia Austral” (UNPA), Argentina. The main research areas are Techniques for Improving Software Development and Mobile Computing.