QoE Measurement of Mobile YouTube Video Streaming

István Ketykó[1], Katrien De Moor[2], Toon De Pessemier[1], Adrián Juan Verdejo[1], Kris Vanhecke[1], Wout Joseph[1], Luc Martens[1], Lieven De Marez[2]

[1]Department of Information Technology, Ghent University / IBBT, G. Crommenlaan 8, 9050 Ghent, Belgium
{iketyko, tdpessem, ajuan, kvanhecke, wjoseph, martens}@intec.ugent.be
[2]Department of Communication Sciences, Ghent University / IBBT, Korte Meer 7, 9000 Ghent, Belgium
{katrienr.demoor, lieven.demarez }@ugent.be

ABSTRACT
The scope of this paper is the interdisciplinary measurement and modeling of Quality of Experience (QoE) related to mobile YouTube video streaming in Living Lab environment. The paper introduces the implementation of a QoE measurement framework on the Android platform and discusses results from a first study using this framework. In this respect, a multi-dimensional QoE prediction model consisting of both objective and subjective parameters is presented. In this model, the test users' evaluations of the content, picture quality, sound quality, fluidness, and loading speed of streamed videos are taken into account and related to a set of objective parameters. To our knowledge, this model is the first to include unlimited, realistic video content. We found that the content has the largest influence on the QoE of online recommend video content in mobile context.

Categories and Subject Descriptors
C.2.1 [**Computer Communication Networks**]: Network Architecture and Design—*Wireless Communication*; H.5 [**Information Interfaces and Presentation**]: Multimedia Information System—*Evaluation/methodology, Audio, Video*; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors*

General Terms
Measurement, Performance, Experimentation, Human Factors

Keywords
QoE, QoS, recommended online video content, Google Android platform, user-centric measurement, mobile Living Lab, 3G

1. INTRODUCTION
Over the last years, the interest in the Quality of Experience (QoE) concept has grown steadily. Moreover, research focusing on the quality of the experiences that ICT users have with products or services seems to have evolved to a study field on its own.

Topics that are discussed in this relatively new research community include e.g., the definition and elements of QoE, the evaluation of both standardized and new measures and measurement approaches, etc. In the scope of this paper, we are particularly interested in those measurement approaches that try to bridge the gap between objective, technical aspects and subjective, user-related dimensions influencing QoE. The former are usually incorporated in definitions of QoE from a telecommunications perspective in which QoE is seen as an extension of the traditional Quality of Service (QoS) [1], thus ignoring the subjective character of a person's experiences. In the literature, the lack of attention to this subjective character is criticized. In [2] it is proposed to use the term QoE as an umbrella term, focusing on the broad range of aspects (e.g., usage context, expectations, previous experiences and network conditions, etc.) that might have an influence on QoE.

In this paper, we present results from an empirical study that combines objective and subjective aspects of QoE related to the use of the PersonalTV application (an application to watch video and receive personal suggestions for video content from YouTube [3]) in a mobile context. We investigate whether there is a relation between on the one hand the monitored, physical parameters and on the other the subjective parameters and overall QoE (explicit evaluation by the test users).

In Section 2, we discuss the relevance of Living Labs and the importance of real content for our study. Section 3 gives an overview of the measurement setup, describes the QoE measurement concept, introduces the PersonalTV Mobile application and discusses the parameters and experimental setup. Section 3 presents the most important results and introduces our QoE model, and Section 4 is dedicated to our conclusions.

2. RELATED WORK
In the literature, several objective (e.g., PSNR, PSQA, etc.) and subjective (e.g., SVQA) video quality evaluation metrics and methods can be found [4]. Although most of these have already been validated and standardized, they hold a number of limitations: usually, this type of research takes place in controlled research settings, focuses on very narrow technical parameters and often, users are not actively involved. As a result, few studies have already focused on the multi-dimensional evaluation of QoE in natural, real life environments. In some recent studies on mobile TV and User Experience related to mobile video, the extension towards such natural, daily life environments was made by organizing e.g. field trials or conducting "Living Laboratory" studies [5]. In [2], the relevance of such Living Labs – defined by [6] as "environments for innovation and development where users

MoViD'10, October 25, 2010, Firenze, Italy.

are exposed to new ICT solutions in (semi-)realistic contexts, as part of medium- or long-term studies targeting evaluation of new ICT solutions and discovery of innovation opportunities” – is discussed in the context of measuring QoE. As Living Labs “bring the lab to the people” and draw on real people’s experiences, QoE research in such settings will likely yield more accurate results than research in controlled environments.

Despite the high number of studies on mobile TV, we did not find any paper in the literature focusing on QoE of online recommended video content in a mobile, natural environment. Moreover, despite the importance and effects of realistic (or lack of realistic) video content [7-9], this paper is the first one – to our knowledge – to present a QoE model of watching mobile video streams: 1) in a natural context and; 2) taking into account the influence factor of real video content. The study described here draws on the framework for evaluating QoE in mobile, Living Lab settings [2].

3. MEASUREMENT SETUP

3.1 Concept

Our aim is to measure QoE of recommended online videos in a mobile, Living Lab context. According to the concept in [2], all of the three QoE dimensions (QoS, context and experience QoE dimensions) are taken into account. In terms of the usage context, the test users were able to take the Android smart phone home and to watch the videos when and where they wanted. Moreover, they were able to select and watch the recommended videos of their own choice. This approach of offering a broad video content to the test users expected to bring us a realistic QoE model [7-9].

3.2 Application

3.2.1 Hardware and software environment

Our requirements regarding the mobile platform are to have information related to the wireless infrastructure and network-level QoS data.

It is necessary to perform the QoS measurements at the client-side because the YouTube streaming servers do not provide public access to this information.

The Google Android platform (Google Android OS 1.5) exposes data related to the wireless infrastructure, such as received signal strength indication (RSSI), through its SDK API. Devices like the HTC Android Developer Phone 1 (ADP1) are capable of running native applications, allowing the integration of common Linux tools into the QoE measurement framework. Specifically, the *tcpdump* packet analyzer is used to collect network-level QoS data on the device.

3.2.2 QoE Measurement Architecture

Figure 1 explains the deployment of the QoE measurement architecture into six nodes (namely Android Mobile Device, QoS Data Back-end, YouTube Streaming Server, Facebook Database, PersonalTV Back-end and Google App Engine). The functional software entities in the nodes are called components.

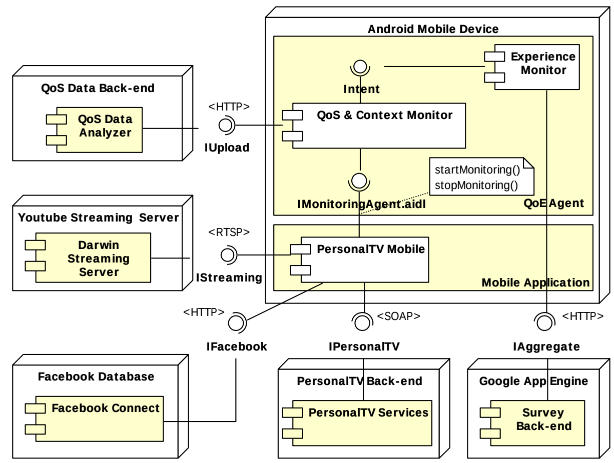

Figure 1. Deployment diagram of the QoE measurement architecture.

Three components (namely PersonalTV Mobile, QoS & Context Monitor and Experience Monitor) are deployed into the Android Mobile Device node and five components (namely QoS Data Analyzer, Darwin Streaming Server, Facebook Connect, PersonalTV Services and Survey Back-end) are deployed into five server nodes (Figure 1). The communication between the components deployed on the different nodes is performed through five interfaces (namely IUpload, IStreaming, IFacebook, IPersonalTV and IAggregate, indicated by circles in Figure 1).

Now, we explain the components deployed on the Android Mobile Device node:

- *PersonalTV Mobile component* is a video streaming player application for Android which has access to external services through three interfaces (namely IPersonalTV, IStreaming and IFacebook) in order to get sets of videos from PersonalTV Services, play streaming videos from the Darwin Streaming Server component and query user profile information from the Facebook Connect component.
- *The QoS & Context Monitor component* logs network QoS, physical and contextual parameters to files which are automatically transferred to the QoS Data Analyzer component via the IUpload interface.
- *The Experience Monitor component* displays surveys and automatically uploads filled-in surveys through the IAggregate interface to the Survey Back-end component deployed on the Google App Engine node.

3.2.3 PersonalTV Mobile application

Figure 2 shows a composition of three screenshots of the PersonalTV Mobile application. The first screenshot (number 1 in Figure 2) in Figure 2 shows four tabs, each of them displaying a list of videos. The first 2 tabs offer most viewed and top rated YouTube videos, respectively within a selected period of time (today, last week, last month, etc.). The third tab offers a set of recommended videos recommended by matching users’ profiles.

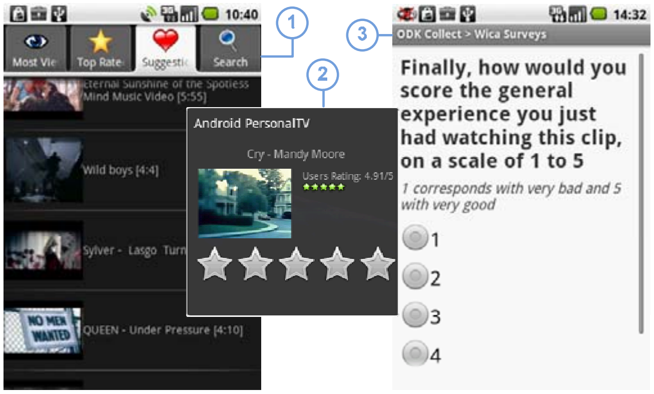

Figure 2. Screenshots from PersonalTV Mobile application.

Standard keyword searches among available YouTube videos can be performed using the fourth tab, listing the result either by relevance, view count or average rating. PersonalTV Mobile offers a rating possibility of the last watched video on a 5-point scale (screenshot number 2 in Figure 2). Screenshot number 3 in Figure 2 is an example of a question belonging to the survey by the Experience Monitor component (see Figure 1).

3.3 Observed Parameters

The QoE Agent provides data of both objective and subjective (user-related) parameters. Objective parameters are logged by the QoS & Context Monitor component, subjective parameters are logged by the Experience Monitor component.

3.3.1 Objective parameters

Table 1 lists the objective parameters (with their units and value intervals) used for the statistical analysis. These parameters were calculated for each individual video-watching session based on the data logged by QoE Agent. In the statistical analysis of Section 3.5 the peak jitter values will be used (peak video jitter denoted as *VJ* and peak audio jitter denoted as *AJ*). The packet loss rates (video packet loss rate denoted as *VL* and audio packet loss rate denoted as *AL*) are the percentage of lost Real-time Transport Protocol (RTP) packets. The Android platform does not allow simultaneous WiFi and cellular data (GPRS, EDGE, 3G & B3G) connections. Therefore, test users are connected either to a WiFi or a cellular data network during one video-watching session. However, in cellular data networks it is possible to have inter-system handovers between GPRS, EDGE and 3G & B3G (Beyond 3G) data connection types. The 3G tag corresponds with UMTS; the B3G corresponds with HSDPA and HSPA data connection. The percentage parameters (*WP*, *GP*, *EP*, *3P* in Table 1) express the duration connected to a specific data network type in ratio of the total duration of video-watching session.

The RSSI values are the arithmetic average of the observed RSSI values sampled in each second (*WR*, *GR*, *ER*, *3R* in Table 1). Handovers (*HO*) are all kind of radio cell reselection and Inter-System Handovers (*ISHO*) are different data connection-type cell reselections (e.g., between UMTS and GPRS). Mobility speed is measured in [m/s] based on GPS information. Four mobility (*M*) categories are determined: indoor (no GPS signal detected), no (0 m/s speed), slow (below 6 m/s) and faster mobility (above 6 m/s). The percentage shares of these mobility categories were 47.5, 25.4, 25.7, and 1.5% for indoor, no, slow, and faster mobility, respectively.

Table 1. Objective parameters and notations

Parameter	Unit	Value	Sampling rate
Video jitter (*VJ*)	[second]	[0, ∞[	Video RTP packets
Audio jitter (*AJ*)	[second]	[0, ∞[	Audio RTP packets
Video packet loss rate (*VL*)	[%]	[0, 100]	Video RTP packets
Audio packet loss rate (*AL*)	[%]	[0, 100]	Audio RTP packets
GPRS percentage (*GP*)	[%]	[0, 100]	In each second
GPRS average RSSI (*GR*)	[dBm]	[-113, -51]	In each second
EDGE percentage (*EP*)	[%]	[0, 100]	In each second
EDGE average RSSI (*ER*)	[dBm]	[-113, -51]	In each second
3G & B3G percentage (*3P*)	[%]	[0, 100]	In each second
3G & B3G average RSSI (*3R*)	[dBm]	[-113, -51]	In each second
WiFi percentage (*WP*)	[%]	[0, 100]	In each second
WiFi average RSSI (*WR*)	[dBm]	[-113, -51]	In each second
Num. of inter-system handovers (*ISHO*)	[integer]	[0, ∞[	In each second
Num. of handovers (*HO*)	[integer]	[0, ∞[	In each second
Mobility (*M*)	[state]	{indoor, no, slow, fast}	In each second

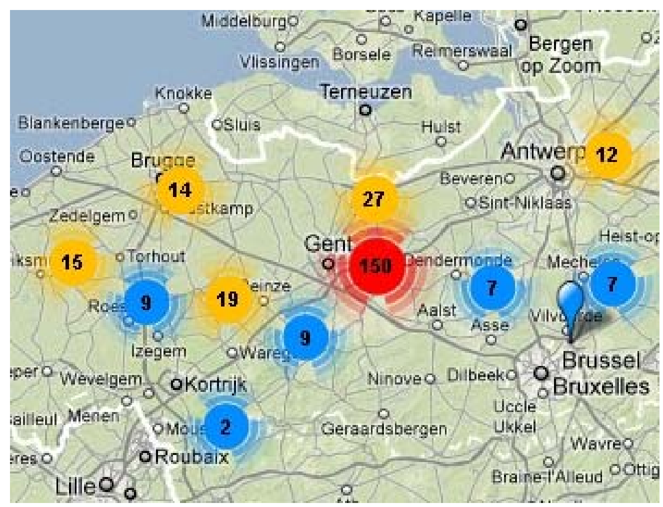

Figure 3. Location clusters based on GPS coordinates.

Figure 3 visualizes the observed location data (based on GPS coordinates) in clusters showing undoubtedly the true Living Lab environment. The numbers in circles correspond to the number of video-watching sessions at a location.

3.3.2 Subjective parameters

The test users' rating of the video (*R*), overall QoE score (QoE) and their evaluation of the content (*c*), accuracy of the recommendation (matching to interests or not)(*i*), loading speed (*l*), picture quality (*PQ*), sound quality (*SQ*), and fluidness (*f*) of the video were taken into account as subjective parameters. In addition, the influence of the content (*I_c*), picture quality (*I_PQ*), sound quality (*I_SQ*), matching to the interests (*I_i*), loading speed (*I_l*), fluidness (*I_f*), were also integrated. These subjective parameters were measured on 5-point scales and gathered via questionnaires displayed after each watching video event for the test users, cfr. Section 3.4.2).

3.4 Experiment and procedure

3.4.1 Sample

Using the convenience sampling method, we recruited 29 test users for the experiment. 79.3% of them are male and 20.7% is female. The overrepresentation of male test users in this case is largely due to the fact that we mainly recruited test users at the Faculty of Engineering from Ghent University. The age of the test users ranges between 23 and 36, with an average age of 28 years old (standard deviation is 3.8). The majority of the test users is working (10.3% is still a student).

3.4.2 Procedure

The procedure that was followed in the experiment consisted of 3 steps:

Firstly, the test users were asked to watch clips using the fixed PersonalTV application [4], so the recommendation algorithm could develop a profile for every test user. Secondly, the test users participated to a briefing (in small groups), which took place at the same location. After filling in a short general questionnaire (including e.g., socio-demographical questions, general questions on their current use of (mobile) online video sites, attitudes …), they were able to try the PersonalTV Mobile application for the first time using the ADP1. All test users received instructions for watching the same 3 pre-defined videos using PersonalTV Mobile. After the watching of every video, they were asked to rate the video using the application's star rating system and to complete a questionnaire. In the last part of the briefing, all test users were asked to select, watch and rate one video from their personal suggestions and to fill in the short questionnaire related to the subjective parameters described above, which was presented on the device immediately after usage. During this briefing, the objective parameters were also monitored. Thirdly, every test user took home an ADP1 for two days and was asked to select and watch (at least) 10 videos from the "suggestions" generated by the application. The test users rated every video and answered the same questionnaire as in the briefing, which popped up on the device after watching. The test users were able to watch the requested number of videos when they wanted and where they wanted, e.g. at home, at work, on the train. They also received a logging document, which could be used for sharing experiences or issues that came up during the test. During watching, the objective parameters were monitored as well.

3.5 Analysis and results

The total number of observations is 392 that is slightly less than the expected because of the missing measurements files for 8 video watching sessions. The dataset of the experiment is analyzed by using the statistical software package IBM SPSS Statistics 18 (see at http://www.spss.com/statistics/).

Table 2 lists the quality parameters of the RTSP video streaming provided by YouTube. The total bandwidth requirement is 64 Kbit/s which is rather low even in a mobile context.

Table 2.YouTube mobile video streaming quality

Audio		Video	
Codec	AMR-NB	Codec	H.263 2000
Bandwidth	12 Kbit/s	Bandwidth	52 Kbit/s
Channels	1	Resolution	QCIF (176*144)
Sampling frequency	8000 Hz	Framerate	15 fps

Figure 4 shows that the test users also found the picture quality (*PQ*) and the perceived overall quality (QoE) low. The box plots show that 50% of the observations were the score 2 and 3, furthermore score 5 values are treated as extreme cases. Figure 4 shows that the overall satisfaction with this video streaming service is at the score 2 and 3 levels. However, the distribution of sound quality (*SQ*) scores is symmetric with the median 3 showing that they found *SQ* quite normal.

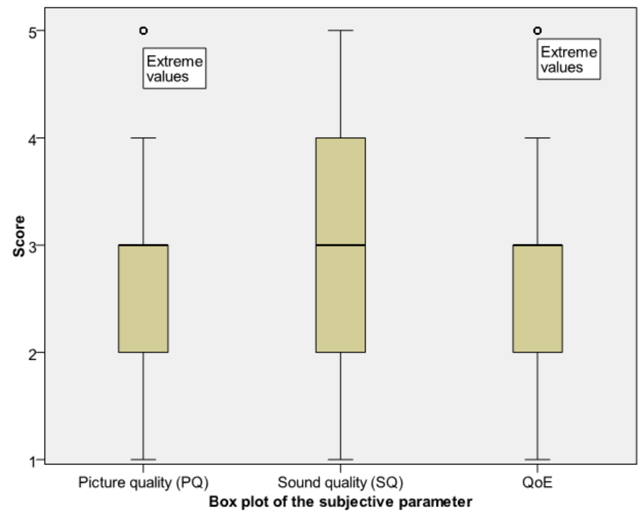

Figure 4. Box plots of the *PQ*, *SQ* and QoE rating scores.

Table 3. Wireless data connection percentage shares

Wireless data connection	Percentage
GPRS	8.70%
EDGE	0.90%
3G & B3G	85.90%
WiFi	4.50%

Table 3 lists the wireless data connection percentage shares during the tests. Test users could freely choose between WiFi and cellular data connection. In spite of this possibility only 4.50% of the watching video sessions were performed over WiFi access network. Mainly the 3G & B3G data connection was used with 85.90% share. The measurements were performed in Proximus UMTS/HSPA network in Belgium (see Figure 3).

3.5.1 QoE model

In this subsection we investigate how the QoE is influenced by the several QoE factors, and how In this subsection we investigate how the QoE is influenced by the several QoE factors, and how the QoE factors can be modeled by the objective parameters.

We propose a QoE model taking into consideration the content (*c*), picture quality (*PQ*), sound quality (*SQ*), matching to interests (*i*), fluidness (*f*) and loading speed (*l*) of the streamed video QoE metrics. All of these QoE factors influence the QoE by a given weight. We gathered this weight from the users themselves by asking their opinion about the influence level of these QoE factors in a 5-point scale.

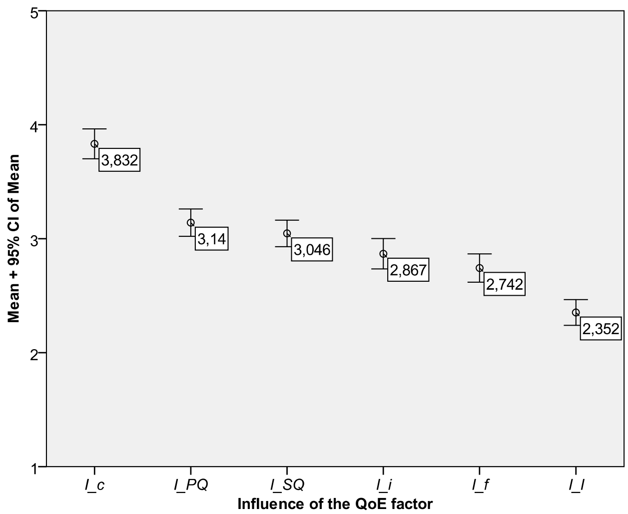

Figure 5. Mean influence scores of the QoE factors and 95% confidence interval.

Figure 5 presents the sample means (with 95% confidence interval (CI) of the mean) of the influence of content (*I_c*), picture quality (*I_PQ*), sound quality (I_*SQ*), matching to interests (*I_i*), fluidness (*I_f*) and loading speed (*I_l*) at rating (*R*). Figure 5 shows the different influence levels of QoE factors in a decreasing order by their influencing weight supporting that content (*c*) is the main determinant (21.3%) of the QoE. Moreover, *PQ* (17.5%) and *SQ* (17.0%) have closely equal influencing weights in QoE.

Table 4. Correlations between subjective parameters and QoE

Subjective Parameter	QoE
Rating (*R*)	*r*=0.833, *p*=0.000
Content (*c*)	*r*=0.787, *p*=0.000
Matching to interests (*i*)	*r*=0.702, *p*=0.000
Loading speed (*l*)	*r*=0.170, *p*=0.001
Picture quality (*PQ*)	*r*=0.277, *p*=0.000
Sound quality (*SQ*)	*r*=0.354, *p*=0.000
Fluidness (*f*)	*r*=0.215, *p*=0.000

Table 4 lists the correlations between subjective parameters and the QoE. Since the rating scores (*R*) and the QoE scores have a highly significant (*p*=0.000) and strong (*r*=0.833) positive correlation, the weights of QoE factors are the same in the QoE as well. Hence, in (1) a model (taking into account *c*, *PQ*, *SQ*, *i*, *f* and *l*) is proposed to predict the average QoE (denoted as $\overline{\text{QoE}}$). Subsequently, the averages are taken over all users and all videos.

$$\overline{\text{QoE}} = \frac{\frac{\overline{I\_c}}{5}\cdot\bar{c} + \frac{\overline{I\_PQ}}{5}\cdot\overline{PQ} + \frac{\overline{I\_SQ}}{5}\cdot\overline{SQ} + \frac{\overline{I\_\imath}}{5}\cdot\bar{\imath} + \frac{\overline{I\_f}}{5}\cdot\bar{f} + \frac{\overline{I\_l}}{5}\cdot\bar{l}}{\frac{\overline{I\_c} + \overline{I\_PQ} + \overline{I\_SQ} + \overline{I\_\imath} + \overline{I\_f} + \overline{I\_l}}{5}} \qquad (1.1)$$

$$\overline{\text{QoE}} = 0.213\cdot\bar{c} + 0.175\cdot\overline{PQ} + 0.170\cdot\overline{SQ} + 0.160\cdot\bar{\imath} + 0.153\cdot\bar{f} + 0.131\cdot\bar{l} \qquad (2.2)$$

In (1.1 and 1.2) the weighted-average of the average QoE factors predicts the $\overline{\text{QoE}}$. The $\bar{c}$ and $\bar{\imath}$ are fully subjective, thus have no correlation with objective parameters; while $\overline{PQ}$, $\overline{SQ}$, $\bar{f}$ and $\bar{l}$ correlate with objective parameters. Table V lists these correlations. The empty cells in Table V mean that there are no significant correlations between the compared objective and subjective parameters, while the filled cells contain significant correlation data.

Table 5 shows that the GPRS percentage (*GP*), Video jitter (*VJ*), Audio jitter (*AJ*), Video packet loss rate (*VL*) and Audio packet loss rate (*AL*) have a correlation with most of the subjective parameters. The low quality and low bandwidth requirement of the video streaming (see Table 2) cause for faster data connections (EDGE, 3G & B3G and WiFi) that the test users can not recognize as many impairments as in case of the slowest data connection (GPRS). That is why any network QoS degradation during the GPRS connection caused more perceptible and significant difference in QoE.

Table 5.
Correlations between objective parameters and QoE factors

Objective/ subjective parameters	Picture quality (*PQ*)	Sound quality (*SQ*)	Fluidness (*f*)	Loading speed (*l*)
GPRS percentage (*GP*)	*r*=-0.150, *p*=0.004	*r*=-0.137, *p*=0.009	*r*=-0.422, *p*=0.000	*r*=-0.180, *p*=0.001
Video jitter (*VJ*)	*r*=-0.147, *p*=0.006	*r*=-0.142, *p*=0.008	*r*=-0.265, *p*=0.000	*r*=-0.176, *p*=0.001
Audio jitter (*AJ*)	*r*=-0.147, *p*=0.006	*r*=-0.142, *p*=0.008	*r*=-0.265, *p*=0.000	*r*=-0.176, *p*=0.001
Video packet loss rate (*VL*)	*r*=-0.238, *p*=0.000	*r*=-0.163, *p*=0.003	*r*=-0.116, *p*=0.032	-
Audio packet loss rate (*AL*)	*r*=-0.112, *p*=0.039	-	*r*=-0.321, *p*=0.000	*r*=-0.238, *p*=0.000
WiFi percentage (*WP*)	-	-	*r*=0.147, *p*=0.005	*r*=0.170, *p*=0.001
Num. of inter-system handovers (*ISHO*)	-	-	*r*=-0.240, *p*=0.000	-
Num. of handovers (*HO*)	-	-	*r*=-0.255, *p*=0.000	-
EDGE percentage (*EP*)	-	*r*=0.114, *p*=0.030	-	-
3G & B3G percentage (*3P*)	-	-	*r*=0.134, *p*=0.011	-
3G & B3G average RSSI (*3R*)	-	-	-	*r*=-0.199, *p*=0.001
WiFi average RSSI (*WR*)	-	-	-	*r*=-0.584, *p*=0.022
Mobility (*M*)	-	-	-	*r*=-0.135, *p*=0.032

Based on the correlations in Table V, the average *PQ*, *SQ*, f, l scores can be modeled by multiple linear regression models (Equation (2)-(5)). The average picture quality score ($\overline{PQ}$) is modeled by the following multiple linear regression model (2):

$$\overline{PQ} = 2.876 - 0.005\cdot GP - 0.003\cdot VJ - 0.006\cdot VL \qquad (3)$$

Equation (2) is obtained by constructing a linear regression model for *PQ* with five dependent variables (*GP*, *VJ*, *AJ*, *VL*, and *AL*). This model is further refined by using a stepwise backward elimination procedure, where those dependent variables which do not significantly contribute to the model's accuracy (as decided by t-tests at the 5% significance level) are omitted. This procedure resulted in the elimination of *AL* and *AJ* from the model, while the other three parameters are kept. The R^2 value is 8.0%. Subsequently, the same type of regression analysis is applied on all models. The average sound quality score ($\overline{SQ}$) is modeled by the following multiple linear regression model (3):

$$\overline{SQ} = 2.974 - 0.005 \cdot GP - 0.003 \cdot VJ - 0.004 \cdot VL + 0.015 \cdot EP \quad (4)$$

Equation (3) is obtained by constructing a linear regression model for SQ with five dependent variables (*GP*, *VJ*, *AJ*, *VL*, and *EP*). This model is further refined by using a stepwise backward elimination procedure, where those dependent variables which do not significantly contribute to the model's accuracy (as decided by t-tests at the 5% significance level) are omitted. This procedure resulted in the elimination of *AJ* from the model, while the other four parameters are kept. The R^2 value is 6.3%. Average fluidness score ($\bar{f}$) is modeled by the following multiple linear regression model (4):

$$\bar{f} = 3.737 - 0.017 \cdot GP - 0.005 \cdot VJ - 0.014 \cdot AL - 0.17 \cdot ISHO \quad (5)$$

Equation (4) is the result of the stepwise regression analysis removing the factors which do not present a significant correlation with *f* (*AJ*, *HO*, *VL*, 3G & B3G, *3P*, and *WP*) and keeping GPRS percentage (*GP*), video jitter (*VJ*), audio packet loss rate (*AL*), number of inter-system handovers (*ISHO*). The R^2 value is 25.6%. The average loading speed ($\bar{l}$) is modeled by the following multiple linear regression model (5):

$$\bar{l} = 2.602 - 0.287 \cdot M - 0.004 \cdot VJ - 0.027 \cdot AL - 0.013 \cdot 3R \quad (6)$$

Equation (5) is the result of a stepwise regression analysis removing the factors which do not correlate significantly with *l* (*AJ*, *GP*, *WP*, and *WR*) and keeping *M*, *VJ*, *AL*, and *3R*. The R^2 value is 35.6%. Equation (2)-(5) can be used in the QoE model of (1). If the individual preferences of the users are known (by gathering their statistics and opinions during the usage of the product) related to the importance of the different QoE factors, it is possible to optimize the video streams for different network circumstances.

3.6 Evaluation of curve fitting

To evaluate the curve fitting of the proposed QoE model, a prediction of the QoE evaluation was calculated with (1) for each video watching session of the experiment. Subsequently, this prediction, which was based on the user's rating (of the content, picture quality, sound quality, matching to interests, fluidness and loading speed), was compared with the actual QoE evaluation values as expressed by the end-users through the questionnaire. Table 6 summarizes the results of this evaluation by mentioning the number of predictions based on the model together with the root mean square error (RMSE) of this prediction with respect to the actual values of the QoE evaluation.

Table 6. Evaluation of the QoE model

	Number values	Percentage values
Number Of Predictions	392	100%
RMSE	0.893	
Correct predictions	191	48.724%
1 star deviation	165	42.091%
2 stars deviation	35	8.928%
3 stars deviation	1	0.255%
4 stars deviation	0	0%

Table 6 lists the number of rounded predictions which equal the actual values (reported as correct predictions). The incorrect predictions are classified according to their deviation from the actual value (1, 2, 3 or 4 stars deviation). Table 6 indicates that almost none of the predictions suffer from a deviation of more than 2 stars, which proves the accuracy of the QoE model.

4. CONCLUSION

In this paper a fully interdisciplinary QoE model is created taking into account the content, picture quality, sound quality, matching to interests, fluidness and loading speed as QoE factors. The relation between QoE factors and technical parameters is also modeled involving network QoS indicators and context information. The concept and implementation of a QoE measurement framework on Android platform is discussed. QoE measurement of mobile YouTube video streams performed in a Living Lab test environment is presented. The result of the measurements is analyzed and the influence of the different QoE factors on the overall QoE related to watching mobile video streams are shown.

5. REFERENCES

[1] D. Lopez, F. Gonzalez, L. Bellido, A. Alonso, Adaptive multimedia streaming over IP based on customer oriented metrics, *Proceedings of the International Symposium on Computer Networks, ISCN 2006*, 2006.

[2] K. De Moor, I. Ketykó, W. Joseph, T. Deryckere, L. De Marez, L. Martens, Proposed Framework for Evaluating Quality of Experience in a Mobile, Testbed-oriented Living Lab Setting, *Mobile Networks and Applications*, 2010.

[3] T. De Pessemier, T. Deryckere, L. Martens, Context Aware Recommendations for User-generated Content on a Social Network Site, in *Proceedings of EuroITV2009*, 2009.

[4] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, PHASE I, http://www.vqeg.org/, 2000.

[5] D. Schuurman, T. Evens, A Living Lab Approach for Mobile TV, in *Proceedings of EuroITV2009*, 2009.

[6] A. Følstad, Living Labs for innovation and development of information and communication technology: a literature review, *The Electronic Journal for Virtual Organizations and Networks 'Special Issue on Living Labs*', 2008.

[7] M. A. Sullivan, J. Pratt, P. Kortum, Practical Issues in Subjective Video Quality Evaluation: Human Factors vs. Psychophysical image quality evaluation, *Proceedings of the First International Conference on Designing Interactive User uxTV2008*, San Francisco, CA, USA, 2008.

[8] S. Jumisko-Pyykkö, V. Ilvonen, K. Väänänen, The Effect of TV Content in Subjective Assessment of Video Quality on Mobile Devices, *Proceedings of IST/SPIE conference on Mobile Multimedia*, San Jose, CA, USA, 2005.

[9] P. Kortum, M. A. Sullivan, Content is King: The Effect of Content on the Perception of Video Quality, *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, pp 1910-1914, Santa Monica, CA, 2004.