Vienna Surfing - Assessing Mobile Broadband Quality in the Field

Raimund Schatz
Telecommunications Research
Center Vienna - FTW
Donau City Strasse 1
1220 Vienna, Austria
schatz@ftw.at

Sebastian Egger
Telecommunications Research
Center Vienna - FTW
Donau-City-Strasse 1
1220 Vienna, Austria
egger@ftw.at

ABSTRACT

This paper presents our findings from a mobile broadband QoE field trial conducted in the city of Vienna, Austria. Using their own laptops in everyday contexts, participants regularly assessed the quality of their mobile broadband connection (tasks: web surfing, file downloads) which in the background was manipulated via traffic shaping throughout a period of three weeks. We discuss our study setup and observations of field trial participant behavior (rating patterns, times, contexts, etc.) as well as the results and lessons learned from correlating end user QoE ratings with measurements performed at client and network level. In addition, we compare our results with those from similar lab experiments, showing that the two evaluation contexts (lab, field) cannot be used interchangeably due to deviant QoE rating behavior of participants.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Evaluation, Methodology; C.4 [**Performance of Systems**]: Reliability, availability, and serviceability

General Terms

Human Factors, Measurement, Performance

Keywords

Quality of Experience, Quality of Service, Acceptability, Lab and Field Trials, Web QoE

1. INTRODUCTION

User satisfaction with application and service performance in mobile communication networks has attracted increased attention during the recent years. This development can be explained by increasing competition amongst service providers and network operators as well as the current proliferation of mobile broadband services. For example, Coda Research and Cisco expect global mobile broadband traffic volume to roughly double every year, with approx. 418 million users generating 1.8 exabytes by 2017 [4, 3]. These trends represent a huge challenge for operators and providers: on the one hand, they need to keep investing in faster networks with higher capacity while on the other, they have to operate on a profitable basis. Consequently, they face a tradeoff between economically investing in infrastructure and providing their end customers with maximum quality. This problem becomes eminent in the context of interactive web applications and file downloads, where high latency and long waiting times caused by low quality network access directly translate into user annoyance and churn [19, 24].

For these reasons, the concept of Quality-of-Experience (QoE) has gained strong interest, both from an academic research and an industry perspective. QoE refers to an understanding of the qualitative performance of communication systems and applications that transcends traditional technology-focused Quality-of-Service (QoS) parameters. Instead, the concept is tightly coupled to the subjective perception of the end user, which is also reflected in the most widespread definition originating from the ITU-T SG 12 [10] which describes QoE as *"overall acceptability of an application or service, as perceived subjectively by the end user."*, which *"may be influenced by user expectations and context."*

A cornerstone of QoE research is the development of perceptual models that adequately map underlying technical conditions (including QoS levels of the network) as well as "softer" factors such as user demographics and context to high-level end-user quality perception. In this context, user experiments where participants grade the different quality conditions presented to them are the norm. These results then need to be correlated with measurements of the underlying network conditions in order to identify generalizable relationships between QoS and QoE.

In this paper, we discuss the experimental setup and the results of a three week mobile broadband field trial conducted in the wild with the goal to assess the impact of different network settings on end-user QoE (tasks: web surfing, file downloads). We chose the field trial method due to the high degree of realism an results validity it is known to yield (cf. [14, 16]). However, since field network conditions cannot be fully controlled as in the lab, results data had to be supplemented with measurements on the network level. The remainder of this paper is structured as follows: after a brief overview of related work, Sections 3 and 4 outlines the

W-MUST'11, August 19, 2011, Toronto, Ontario, Canada.

project context, methodology and experimental setup used. Section 5 discusses the most relevant results (including a comparison with lab results) and Section 6 concludes the paper with lessons learned and an outlook on future work.

2. RELATED WORK

As mentioned in the introduction, mobile broadband traffic volume is growing at large. This growth together with the fact that HTTP traffic in residential broadband Internet accounts for a share of more than 50% as found by [15] has raised widespread interest in how to reliably assess perceived quality of interactive data services, also termed as 'Web QoE'.

In this domain, the most prominent application scenarios analyzed are web browsing and file downloads which also constitute the scope of this paper. Early studies such as [1, 2] have introduced measures of user satisfaction (3-point scale) and acceptability for web services. Later studies such as [17, 20] have mainly utilized this satisfaction scale and extended it to a 5-point scale. Recent work on Web QoE [9], [7] and [23] has converged so far, that utilization of the MOS methodology and ACR scales from video and audio quality assessment (cf. [8]) has emerged as a de-facto standard for Web QoE evaluation. On the other hand, acceptability measures have received far less attention in recent work on this topic.

In contrast to these user centric approaches which are based on direct user feedback, network centric approaches have also been proposed to estimate user perceived quality from pure network level information. The work in [13, 5] utilized the cancellation rate of HTTP requests as an estimation of user experienced quality. However, the authors did not report empirical validations of real user ratings with their computed QoE results. Thus, in order to overcome the shortcomings of aforementioned methods a joint approach which combines subjective user ratings with network level information as discussed in this paper is required.

As concerns subject quality assessment, lab experiments have become the prime method of choice within Web QoE. This can be explained by the fact that this method has been long established in the audio and video domain [8, 11]. In addition, early Web QoE assessment recommendations (cf. [9]) have inherited their methodological inventory from these domains. However, lab experiments do neglect the multitude of influence factors on QoE, for example usage context or device usability. The importance and the advantages of placing subjects as much as possible in his daily life for user testing has been shown by [14, 16]. The authors emphasize that natural, daily life situations are more representative of actual user experiences. Such field-based experiences are far more realistic compared to experiences the same users would go through in lab experiments. Furthermore, results presented by [12, 21] indicate that for acceptability-related questions, significant differences regarding interpretation and answering behavior do exist between users in lab and field contexts. Therefore it is necessary to conduct Web QoE studies also in the field and to contrast their results with comparable lab studies in order identify the merits and drawbacks of each approach.

3. METHODOLOGY

In this section, we introduce the research project ACE which provides the context for the QoE trials presented in this paper as well as the general research methodology behind. ACE (Advancing Customer Experience) has been conducted at Telecommunications Research Center Vienna (FTW) in collaboration with a leading Austrian mobile network operator as well as an systems provider and aims at better modeling, measuring and managing quality in 3G networks. The goal of the project is to realize the shift from QoS to QoE by investigating the links between technical network parameters and the end-user's QoE.

Fig. 1 depicts the project's guiding framework for developing efficient models predicting QoE of networked services. The framework suggests a layered modeling approach, distinguishing between the network, application and the user layers. As a first step, the most relevant performance indicators for each layer are derived from user studies and technical analyses.

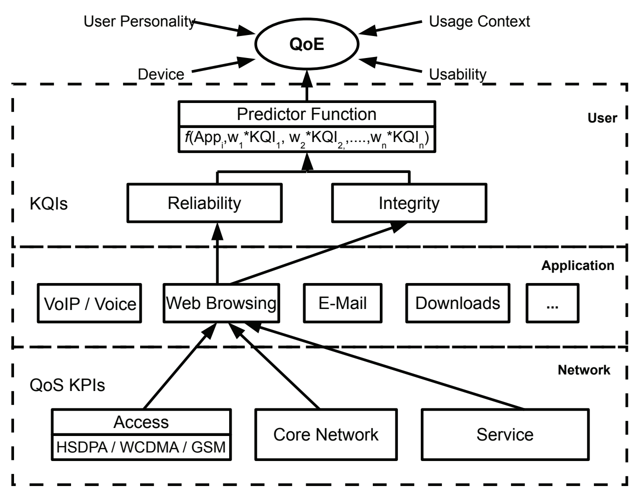

Figure 1: The ACE QoE Modeling Framework.

Since each application (VoIP, web surfing, e-mail, etc.) obviously has its own technical and user-experience characteristics, the QoE modeling process and its result are application/service specific. For example, while browsing the web, the perceived performance highly depends on the time from request/click until the start of rendering the new page as well as the time each page takes to load (therefore both application-level metrics need to be logged during user test sessions as part of the data acquisition). In turn, these performance-relevant metrics can be linked to the underlying network traffic by means of low-level performance indicators such as delay, effective throughput, and HTTP object load time.

This layered approach has immediate consequences on research methods used: user studies providing high-level subjective quality ratings as an output need to be combined with log data on application level and traffic measurements describing the QoS on the networking layer. These different layers of data are then combined using methods such as regression analysis, probabilistic modeling and machine learning, depending on the complexity of the application under investigation.

Guided by above framework, a number of comprehensive user studies have been performed in order to generate sufficient QoE modeling data, focusing on mobile broadband application scenarios like web surfing, file downloads and online video. The studies have been conducted as a series of controlled lab experiments and one field trial (focused in this paper), during which we exposed participants to a broad variety of test conditions (see [18] and [21] for further details).

4. EXPERIMENTAL SETUP

As previously mentioned, we conducted a mobile broadband field trial in order to complement our project's initial series of QoE lab studies with measurements performed in real-world settings. Participants were recruited via an announcement on the research center's homepage as well as FTW's database of test users. The test tasks addressed the two most popular mobile broadband application scenarios: web browsing and file downloads. 17 persons participated in the study and were instructed to regularly surf the web on their laptops (the most relevant type of mobile broadband user terminal in Austria at that time) and download mp3 files via a music service subscription we provided to them. This level of guidance was necessary in order to guarantee a sufficient number of comparable, consistently rated conditions at the end of the trial. Subjects were compensated for their efforts with vouchers and the downloaded music files they were allowed to keep. As we will show in Section 5, these measures were sufficient for motivating and sustaining high rating discipline.

	Field Trial	Lab 1	Lab 2
No. Subjects	17	26	32
Age (Mean)	30.51	35.42	28.39
Female/Male	8/9	10/16	15/17
Environment	Field	Lab	Lab
Services	Web+Download	Download	Web

Table 1: Demographics and Experimental Conditions of the Field Trial (Column 1).

As concerns network connectivity, participants used the 3G HSPA network of a leading Austrian mobile operator that provides excellent coverage within the city region of Vienna. In order to create different network conditions in terms of maximum bandwidth available, particular technical arrangements had to be made. Firstly, the IP traffic of trial participants (as identified by their usage of a dedicated APN) was routed from the provider's core network to our network emulator EMU (see Fig. 2). The emulator (based on Netem [6]) shaped each user's traffic according to different parameter sets. For example, downlink bandwidth limits were varied between 64 and 2048 kbit/s in log_2 steps as well as at times set to unlimited full HSPA speed. The parameter sets were automatically changed every 30 minutes via a script in a random fashion, thus participants experienced different quality levels even during a single session. In addition, we captured the network traffic of each user with a tracing probe (TRC) in order to track user behavior and to be able to measure low-level QoS parameters such as the actual throughput utilized.

In the context of the execution of both task scenarios

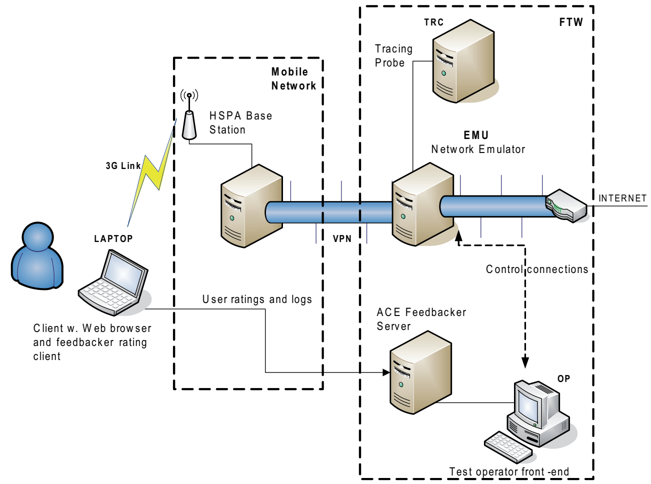

Figure 2: Technical setup of the field trial

(web surfing, file download), users were instructed to rate the quality of the connection using an electronic rating tool. Participants were asked to assess the quality experienced during the most recent mp3 download or throughout the last three minutes in the case of web surfing. Participants issued six ratings per day on average via a custom application running on their laptops which forwarded the QoE feedback to a central data collection server. This way, we collected approximately 142 ratings per subject over a period of three weeks.

As concerns the quality ratings themselves, we asked subjects for feedback regarding the following two aspects: *satisfaction* with the performance of the connection and *acceptability* of the quality level experienced. Subjects rated their degree of satisfaction on an ordinal ACR-9 Mean Opinion Score (MOS) scale [8] ranging from 'bad' to 'excellent'. In contrast, acceptability was treated as a binary yes/no measure for answering the question: "Would you continue to use the service under these conditions or not?".

5. STUDY RESULTS

This section discusses our trial results according to three aspects. Firstly, we discuss results preprocessing (which was necessitated by the field setting) that involves network-level measurements. Then we discuss some observations regarding user behavior in the field and finally, we report QoE rating results of the participants for web browsing and download services.

5.1 Data Processing Based on Network Level Measurements

As mentioned in the previous section, we used the 3G infrastructure of a local mobile network operator to provide our subjects with wireless HSPA access. Hence, the bandwidths we set on our network emulator represent the maximum throughput levels that participants could achieve when their radio reception conditions also allowed it. Due to the excellent coverage within Vienna, set throughput levels were also achieved most of the time. For all other cases, we had to reassess the achieved maximum throughput on behalf of network traces (cf. Fig. 2). We assessed maximum throughput via counting the TCP payload for time bins of certain length within the timeframe of the corresponding user rating (typically 3 minutes)[1]. Specifically, we chose the bin with the highest payload as representing the maximum bandwidth as perceived by the user within a respective condition. As a result of this re-assessment process approximately 19 % of the ratings had to be reallocated to bandwidths different from the set bandwidth on the EMU. Fig. 3 depicts the relation between achieved throughput and set bandwidth. The median of the achieved throughput reached the preset level for set downlink bandwidths up to 1024 kbit/s, whereas for higher downlink bandwidths the median did not reach the preset setting. This was an expected result as wireless access networks naturally underly throughput variations due to changing channel characteristics. In order to correctly relate the users QoE ratings to the actual bandwidth they experienced, we used these achieved throughput values for the abscissa in the following rating results analysis in the upcoming subsections.

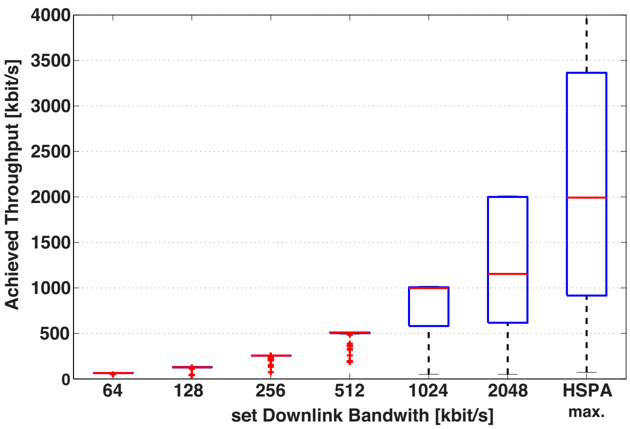

Figure 3: Set Bandwidth vs. estimated real Bandwidth assessed via Traces

5.2 User Rating Behavior

In total we obtained 2430 usable ratings from our participants, i.e. approximately 6 six ratings per participants per day. Most importantly, almost all participants fulfilled their daily rating quota voluntarily, thus automatic prompts or reminders were not necessary. Fig. 4 below illustrates when and where trial participants mostly issued their ratings. The most preferred location for using the mobile broadband connection and rating clearly was at home, followed by 'other' (mostly indoor locations like coffee house, friends' premises, etc.) and the workplace. This distribution confirms the fact that mobile broadband usage on laptops primarily takes place in static in-door settings an not while on the move. Furthermore, rating activity peaked around noon and in the evening (between 8pm and midnight), indicating that users mostly were testing in the context of leisure time usage.

Regarding content choice, we observed that the 10 websites that were most frequently visited[2] by trial participants constitute the context for almost 50% of all Web QoE ratings issued. In this respect, subjects' surfing behavior was fairly homogeneous although at the initial briefing, we encouraged them freely surf the web without any constraints during the trial period.

[1]Note, that if within the 3 minutes before a rating, the traffic shaper's bandwidth was changed, the rating was filtered out.

[2]A visit was counted when a site URL has been requested just before a Web QoE rating.

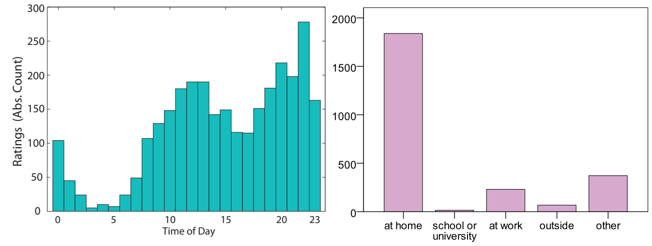

Figure 4: Distribution of participant rating times (left) and locations (right) for the field trial.

5.3 Rating Results

This section presents the results we obtained from the QoE user study in the field and contrast them with comparable ACE lab studies (For further details on the lab studies please refer to Table 1 and [18, 21]). We first discuss QoE assessment results for the case of web surfing, which we then extend towards file download scenarios. As a third example we compare subjects' binary acceptability opinion ratings with MOS satisfaction ratings.

The upper plot in Fig. 5 shows QoE ratings for web browsing at different instrumented downlink bandwidths. The figure shows a steady increase of the MOS scores for lower bandwidths which tends to saturates from 512 kbit/s upwards. It is noteworthy that only minor differences in the ratings exist between lab and field. These results indicate that for web browsing, the users' rating behavior is comparable in both contexts.

In the case of file downloads the relation between user perceived speed and downlink bandwidth is a bit more complex. We have shown in [18] that the size of the file accounts as another factor that influences user expectations and thus tolerance for waiting times. This results in file size dependent MOS rating behavior as can be seen from the set of curves in the right plot of Fig. 5. However, the MOS ratings for the 5.5 MB file size (as used in the field study) does not lie in between the 2.5 and 10 MB ratings from the lab study as they should. The most plausible explanation for this divergence is that file downloading in fact is a very simple, straightforward waiting task. In general, the perception of waiting times strongly depends on one's current attention span as influenced by task and situational context [22]. In the field study, participants were in their natural environment exposed to several sources of distraction, whereas in the lab they had to wait and stare at screen until the file was downloaded. Therefore, users subjectively experienced waiting times to be longer which resulted in worse MOS ratings in the lab study[3]. These results suggest that in case of file downloads results from lab tests can not be straightforwardly transferred to real life situations.

Finally, Fig. 6 maps QoE to acceptability ratings. For each value on the MOS scale, the relative share of cases in which participants rate a condition as acceptable is depicted. Accordingly, the plot shows which QoE levels the majority

[3]However, field trial participants were instructed to not multi-task during file downloads, similar to the lab.

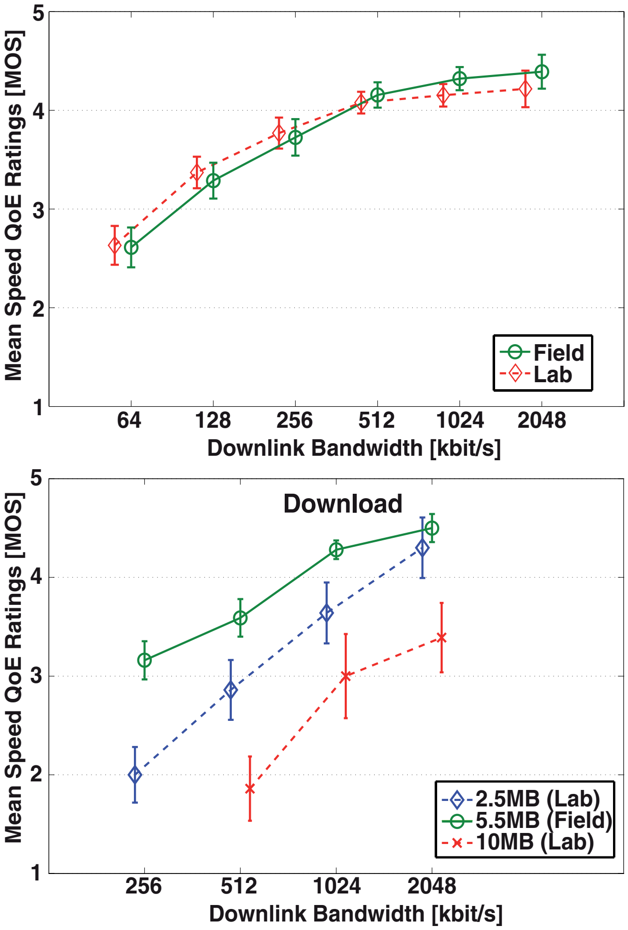

Figure 5: MOS ratings for web browsing (top) and file downloads (bottom)

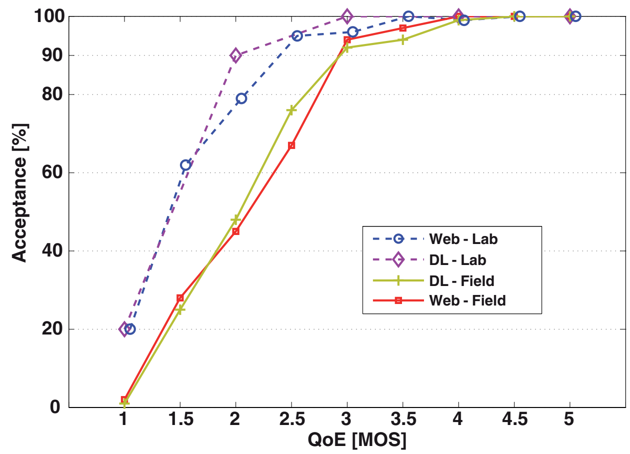

Figure 6: Mapping of MOS to Acceptability ratings for in the lab and the field

of subjects consider acceptable and which ones not. Ostensibly, the mappings between MOS and acceptability shares strongly deviates between lab and field. From these differences we conclude that subjects use a different concept of acceptability in everyday life contexts than in a lab setting. Thus, similar to the case of download satisfaction, acceptability rating results run the danger of lacking external validity when based on inquiry in lab environments.

6. CONCLUSION

In this paper, we presented the results of a field study on mobile broadband QoE that uses a mobile operator's 3G network as integral part of the test infrastructure. We demonstrated that our method and setup yields reliable results, if implemented with sufficient care. This does not only relate to briefing and compensation of participants, but also the underlying technical test infrastructure. In this respect, special attention needs to be paid to the fact that in the field, radio reception conditions are hard to control, which necessitates the inclusion of traffic monitoring and throughput estimation in the data gathering and processing process. Our results show that even if the same test tasks and QoS parameters are used, lab and field results can differ considerably, depending on the service type and quality dimension tested. This indicates, that context has a substantial influence on rating behavior and that consequently, certain concepts (such as acceptability) should be better tested in the field than in the lab in order to obtain results with high external validity.

As concerns future work, we will undertake deeper analysis of the traffic traces recorded in order to compute network- and application-level KPIs (like HTTP object load time or web page load time) and model their relationship with QoE. Furthermore, we are going to validate our methodology in another field trial that will feature additional service types like online video as well as smartphones and tablets.

7. ACKNOWLEDGMENTS

This research has been performed within the projects ACE and U-0 at the Telecommunications Research Center Vienna (FTW) and has been funded by the Austrian Government and the City of Vienna within the competence center program COMET. The authors would like to thank their FTW colleagues Kathrin Masuch, Matthias Fuchs and Alexander Platzer for their valuable input and discussions.

8. REFERENCES

[1] A. Bouch, A. Kuchinsky, and N. Bhatti. Quality is in the eye of the beholder: meeting users' requirements for internet quality of service. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 297–304, New York, NY, USA, 2000. ACM.

[2] A. Bouch, M. A. Sasse, and H. G. DeMeer. Of packets and people: a user-centered approach to quality of service. In *Proc. of IWQoS 2000*, 2000.

[3] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, Feb. 2010.

[4] Coda Research Consultancy. Mobile broadband and portable computers: Revenue, user and traffic forecasts 2009-2017, July 2009.

[5] D. Collange and J.-L. Costeux. Passive estimation of quality of experience. *Journal of Universal Computer Science*, 14(5):625–641, 2008.

[6] S. Hemminger et al. Network emulation with netem. In *Proceedings of Linux Conf Au*, 2005.

[7] E. Ibarrola, F. liberal, I. Taboada, and R. Ortega. Web qoe evaluation in multi-agent networks: Validation of itu-t g.1030. In *ICAS '09: Proceedings of the 2009 Fifth International Conference on Autonomic and Autonomous Systems*, pages 289–294, Washington, DC, USA, 2009. IEEE Computer Society.

[8] International Telecommunication Union. Methods for subjective determination of transmission quality. *ITU-T Recommendation P.800*, Aug. 1996.

[9] International Telecommunication Union. Estimating end-to-end performance in ip networks for data applications. *ITU-T Recommendation G.1030*, November 2005.

[10] International Telecommunication Union. Vocabulary and effects of transmission parameters on customer opinion of transmission quality, amendment 2. *ITU-T Recommendation P.10/G.100*, 2006.

[11] International Telecommunication Union. Subjective video quality assessment methods for multimedia applications. *ITU-T Recommendation P.910*, April 2008.

[12] S. Jumisko-Pyykkö and T. Utriainen. A hybrid method for quality evaluation in the context of use for mobile (3D) television. *Multimedia Tools and Applications*, 2010.

[13] S. Khirman and P. Henriksen. Relationship between quality-of-service and quality-of- experience for public internet service. In *Proceedings of the 3rd Workshop on Passive and Active Measurement*, Fort Collins, Colorado, USA, March 2002.

[14] M. Leitner, P. Wolkerstorfer, A. Geven, N. Höller, and M. Tscheligi. Evaluating a mobile multimedia application in field trials: the cost-benefit of self-report methods. *Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild*, page 27, 2009.

[15] G. Maier, A. Feldmann, V. Paxson, and M. Allman. On dominant characteristics of residential broadband internet traffic. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, IMC '09, pages 90–102, New York, NY, USA, 2009. ACM.

[16] K. D. Moor, I. Ketyko, W. Joseph, T. Deryckere, L. D. Marez, L. Martens, and G. Verleye. Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting. *Mobile Networks and Applications*, 15(3):378–391, 2010.

[17] K. Papamiltiadis, H. Zisimopoulos, M. Gasparroni, and A. Liotta. User quality of service perception in 3g mobile networks. *Telecommunications Quality of Services: The Business of Success, 2004. QoS 2004. IEE*, pages 64–69, March 2004.

[18] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo. The logarithmic nature of qoe and the role of the weber-fechner law in qoe assessment. In *Proc. of the 2010 IEEE International Conference on Communications*, pages 1–5, May 2010.

[19] G. M. Rose, R. Evaristo, and D. Straub. Culture and consumer responses to web download time: a four-continent study of mono and polychronism. *IEEE Transactions on Engineering Management*, 50:31–44, February 2003.

[20] J. Saliba, A. Beresford, M. Ivanovich, and P. Fitzpatrick. User-perceived quality of service in wireless data networks. *Personal Ubiquitous Comput.*, 9(6):413–422, 2005.

[21] R. Schatz, S. Egger, and A. Platzer. Poor, Good Enough or Even Better? Bridging the Gap between Acceptability and QoE of Mobile Broadband Data Services. In *Proceedings of the 2011 IEEE International Conference on Communications*, pages 1–6, June 2011.

[22] S. C. Seow. *Designing and Engineering Time: The Psychology of Time Perception in Software*. Addison-Wesley Professional, 2008.

[23] J. Shaikh, M. Fiedler, and D. Collange. Quality of experience from user and network perspectives. *Annals of Telecommunications*, 65:47–57, 2010. 10.1007/s12243-009-0142-x.

[24] Zona Research. The economic impacts of unacceptable web-site download speeds. Technical report, Zona Research, April 1999.