

# Quantifying the influence of rebuffering interruptions on the user's Quality of Experience during mobile video watching

Toon De Pessemier, Katrien De Moor, Wout Joseph, *Member, IEEE*,  
Lieven De Marez, and Luc Martens, *Member, IEEE*

**Abstract**—Entertainment and multimedia are the key functionalities in emerging mobile markets. The ability to understand and quantify the Quality of Experience (QoE, i.e. the users' subjective perception of the 'overall acceptability of an application or service'), will play a major role in the success of these mobile services. This study explores the thresholds at which the technical quality of a mobile video service becomes unacceptable for users. A subjective experiment drawing on the logging of technical parameters combined with subjective evaluations by a user panel resulted in a model for quantifying the acceptability of video interruptions. The results of this analysis provide insights into the QoE and (un)acceptability regarding video interruptions for different network conditions and video parameters. The conclusions of this paper can be used as a guideline for service design and network dimensioning.

**Index Terms**— mobile, video, subjective evaluation, Quality of Experience

## I. INTRODUCTION

Mobile devices are becoming the primary tools that are used for internet access and communication. According to the latest version of the TNS Mobile Life survey, claimed to be the largest study into mobile consumers [1], this growth in the mobile communication domain is driven by an increased demand for mobile video services. Recent forecasts [2] state that mobile video transmission will generate 66% of the global mobile data traffic by 2015. However, offering a high Quality of Experience (QoE) to users remains challenging, and given the dependence on several influencing factors, this especially holds true in the context of mobile video applications [3]. This

Manuscript received xxx, 2012. This work was supported by the IBBT-GRASP project, co-funded by the IBBT (Interdisciplinary institute for BroadBand Technology), a research institute founded by the Flemish Government. W. Joseph is a Post-Doctoral Fellow of the FWO-V (Research Foundation - Flanders).

T. De Pessemier, W. Joseph, and L. Martens are with Ghent University/IBBT, Dept. of Information Technology, Gaston Crommenlaan 8, B-9050 Ghent, Belgium, Fax: +32 9 33 14 89 9 (e-mail: Toon.DePessemier@intec.UGent.be, Wout.Joseph@intec.ugent.be, Luc.Martens@intec.ugent.be).

K. De Moor and L. De Marez are with Ghent University/IBBT, Dept. of Communication Sciences, Korte Meer 7-9-11, B-9000 Ghent, Belgium (e-mail: KatrienR.DeMoor@UGent.be, Lieven.DeMarez@ugent.be).

emphasizes the necessity for service providers to investigate the quality of users' experiences in view of matching the produced video quality to users' subjective expectations. Therefore, visual quality assessments are usually conducted using subjective tests in which human subjects are asked to rate the perceived visual quality of the displayed media according to a provided quality scale [4]. These visual quality assessments play a crucial role in meeting the promised Quality of Service (QoS) and in improving the end-user's QoE [5].

In contrast to traditional QoS approaches, which are usually driven by multimedia signal degradation from the signal quality point of view, QoE considers how viewers perceive and experience multimedia content and/or multimedia communication services as a whole [6]. Since QoE relates to the user-perceived experience directly rather than to the implied impact of QoS, it is considered as a more important metric than QoS [7]. Different definitions of QoE exist, but all have similar notion, referring to user satisfaction [8]. QoE is defined by the International Telecommunication Union as 'the overall acceptability of an application or service, as perceived by the end-user', which might be influenced by 'user expectations' and 'context' [9].

Identifying, understanding, and quantifying the most determining aspects making or breaking the QoE of individual (or communities of) users and translating these rich insights into service and application optimization recommendations, is considered to be essential. QoE will continue to play a major role in the future development of broadcasting services and the design of multimedia applications, not the least in the dynamic mobile media domain. For video services, operators, and broadcasters, QoE has become a service differentiator next to the number of channels or the content they offer [10]. Moreover, QoE has become a key factor in routing mechanisms and resource management schemes for network operators and IPTV providers [11].

Various studies have been conducted to determine the QoE for UDP (User Datagram Protocol) based streaming as a function of the technical video parameters (resolution, frame rate and codec) and spatial and temporal video artifacts resulting from network imperfections e.g., packet loss, delay, and jitter [12]. However, nowadays a lot of video content is

available via Dynamic Adaptive Streaming over HTTP (DASH), a technique also known as progressive download which is based on HTTP (Hypertext Transfer Protocol) and TCP (Transmission Control Protocol) which assures a reliable, ordered delivery of video packets.

Using progressive download, buffering mechanisms and packet retransmissions can avoid the audiovisual distortions due to packet loss and jitter, but may incur rebuffering interruptions and additional start-up delays compared to UDP based streaming applications [13]. In other words, in case of a network bottleneck the displayed content does not suffer from video quality degradation, but playback suffers from rebuffering interruptions. In Section II, we discuss a number of studies that have looked into how such interruptions influence QoE. However, research investigating the acceptability with respect to loading time and rebuffering interruptions during mobile video watching is still rather limited. Moreover, results based on fixed video watching (using wired devices) cannot be applied to the mobile domain without adjustments because the user's expectations can differ depending on the platform, and the user's experience is influenced by the type of device and display.

The aim of this paper is therefore to investigate the influence of rebuffering interruptions on QoE during mobile video watching. It presents results from a subjective experiment, in which test subjects were asked to evaluate different measures of QoE and specific QoE features, while watching mobile video content in six different technical scenarios. These scenarios consisted of different combinations of connection types (low, medium, and high bandwidth) and video quality sources (low and high quality) and were characterized by a different number of rebuffering interruptions. For each of the scenarios, measurements drawing on a set of objective technical measures and subjective quantitative measures of QoE were collected. As a complement to the latter, qualitative data were gathered in order to gain a better understanding of relevant features and influencing factors according to the test subjects.

The remainder of this paper is organized as follows. In Section 2, an overview of related work is given. Section 3 describes the aims of the study and the setup of the subjective experiment. In this section, we provide details about the objectively-measured parameters and the subjective QoE measures. Section 4 elaborates on the results of this study and presents a model for quantifying the acceptability of video interruptions in a mobile context. Finally, the conclusions of the experiment are discussed in Section 5.

## II. RELATED WORK

In this section, we provide an overview of existing work starting with studies on the influence of technical parameters on the QoE during video watching. We discuss a number of studies regarding QoE modeling in the context of streaming media services. Subsequently, the focus is on studies

concerning the improvement of the QoE in the mobile application domain by approaches such as optimizing the handover process. Furthermore, this section discusses results regarding QoE research in WCDMA (Wideband Code Division Multiple Access) and UMTS (Universal Mobile Telecommunications System) networks. Finally, the pros and cons of controlled test beds as well as living lab experiments are reviewed.

Many of the studies on QoE have focused on how network-level parameters (such as delay, bandwidth, packet loss, and jitter) and video characteristics (such as codec, frame rate and resolution) affect the QoE of the multimedia content [14]. For instance, the effects of present-generation video compression and communication technologies on the perceptual quality of digital video were evaluated via a subjective study [15]. This user study consisted of a large-scale subjective evaluation of video quality on a collection of videos distorted by a variety of application-relevant processes. Furthermore, the performance of several full-reference video quality assessment algorithms was evaluated and compared with the users' mean opinion scores.

In the context of streaming media services, a statistical modeling technique was employed to correlate QoS parameters with estimates of QoE perceptions and to identify the degree of influence of each QoS parameter on the user perception [16]. The result was a classification model that predicts the user's perception of QoE based on the bitrate and frame rate of the video. The proposed prediction model allows network operators to anticipate the user's experience and then allocate network resources accordingly.

The impact of the underlying transport protocol on the QoE for streaming media services was studied by a comparison of UDP based and TCP based video streaming [17]. The results indicated that TCP based video-on-demand delivery, which is for instance used by YouTube, outperforms UDP based video streaming in terms of the user's perceived quality for network bottleneck scenarios. The user's personal satisfaction rating showed to be highly influenced by the number of video interruptions during video streaming based on TCP. All users rated their video experience with the maximum rating if no interruption occurred. In contrast, in case of two and more rebuffering interruptions, more than 30% of the users rated the video experience with the lowest rating score. From these results, the authors concluded that users tolerate one interruption of 3 seconds per clip but more interruptions, especially more than two, significantly reduce the user's perceived video quality. Since user expectations and experiences might be different for mobile applications due to other characteristics of the hardware (e.g., type of device or screen size) and the cellular data networks, we assume that these conclusions do not apply (without adjustments) to the acceptability regarding video interruptions on the mobile platform, which is investigated in our research.

The proliferation of multimedia applications over mobile, resource-constrained networks has raised the need for methods

that adapt these applications both to network resource constraints and to clients' QoE requirements. In this context, the upper-layer adaptation mechanisms have been investigated in order to achieve end-to-end delay control for multimedia applications [18].

Moreover, various studies have been devoted to the measurement of QoE by considering both measurable and non-measurable parameters in the domain of mobile multimedia services [19]. The measurable parameters are typically those related to the technological aspects of the service. Using these parameters, it is possible to produce quantifiable quality metrics for QoE evaluation of multimedia streaming by performing analysis at the application and network levels [19]. The so-called 'non-measurable' parameters [19] are related to the users' evaluation of the investigated QoE features, their behavior, expectations, emotions, etc. Increasingly, interdisciplinary research is set up in order to take these human aspects into account.

In the context of heterogeneous mobility, it is also investigated how network hand-overs can be optimized and made seamless, allowing the user to have the best possible experience. To improve this handover process across multiple link-layer access technologies, a modified Android user terminal using the IEEE 802.21 framework has been proposed [20]. The assessment of the handover process via an experimental test bed showed that under the proposed solution the handover delay and packet loss are significantly lower than the ones resulting from the normal operation.

Besides, various researchers have studied the QoE of mobile media sessions in cellular data networks. Based on experiments in WCDMA networks, a predictive QoE model for multimedia applications was proposed [21]. A stepwise regression analysis revealed the most relevant factors for the QoE: the number of transmission errors, buffering occurrences, and coding profile. Moreover the study pointed to the importance of the buffering duration and frequency. Experiments in UMTS networks found that the effect of the RTT (round-trip time) and bandwidth are very perceivable by the users while browsing web pages [22]. The same study showed that the initial startup time of streaming video is crucial, independent of the quality of the streaming. The test subjects were also very sensitive to any rebuffering that occurs after the streaming has started, and rated the overall quality regardless of the video quality after the rebuffering interruption.

Other studies considered the rebuffering length and rebuffering frequency as the properties that have the greatest impact on QoE. If interruption is unavoidable, a single rebuffering is a better solution than repeated rebuffering events [23]. Other subjective tests showed that also a single rebuffering interruption can reduce the users' QoE considerably [24]. A very recent study compared the impact of initial delays vs. interruptions and found that the latter should always be avoided, even at the cost of increasing the initial waiting time due to prebuffering [25]. Regrettably, these

studies do not evaluate how much time can be spent on the rebuffering of mobile video before this becomes unacceptable for the user, or in other words the acceptability regarding the initial loading time and the rebuffering interruptions during video playback. In this paper, we therefore investigate the acceptability of rebuffering interruptions during mobile video watching and provide a model that estimates this acceptability considering the initial loading time and rebuffering interruptions.

To evaluate QoE in the context of mobile applications or services, both traditional test beds with controlled parameters or living lab experiments in the field can be set up. Test beds allow for transparent, rigorous, and replicable testing of new technologies, scientific theories, and tools regarding the quantification and optimization of the QoE. Conversely, living lab experiments are less transparent and predefined but aim to provide more natural settings for studying QoE by involving the users in the innovation process [26]. Although these living lab experiments are an extension towards more natural and realistic research test environments [27], a strong tradition exists in experimental research taking place in controlled laboratory settings. Research using this kind of test beds makes it possible to investigate the relative influence of particular isolated parameters on users' quality perceptions. Therefore, the test presented in this paper has been carried out in such a controlled environment test room.

### III. TEST SETUP

#### A. Aims of the Study

The main aim of this study is to investigate the influence of rebuffering interruptions on Quality of Experience during mobile video watching. More concretely, we investigate if and how the test subjects' QoE is influenced by the number of rebuffering interruptions in six technical scenarios combining three simulated connection types and two video qualities. We investigate the influence of the objective measures mentioned in Table 3 on different measures of QoE, as dependent variables in our study. These include the overall experience rating, and the evaluation of both the overall technical quality, as well as specific QoE features, being interruptions, loading time, and fluidity (sometimes also referred to as fluentness or smoothness of the video playback). By the term QoE feature, we refer to 'a perceivable, recognized and namable characteristic of the individual's experience of a service which contributes to its quality' [28]. We included these specific features to investigate their relative importance and concrete evaluation by test subjects and because previous research pointed to their importance. Finally, we investigate possible differences in terms of the acceptability of video playback interruptions due to rebufferings. Following the definition given in [29], acceptability refers to a 'binary measure to locate the threshold of minimum acceptable quality that fulfills user quality expectations and needs for a certain application or system'. In addition, we also take into account the importance of specific QoE features related to mobile video watching

before and after the actual test. Furthermore, we complement the test subjects' ratings with qualitative feedback on expectations and assumed importance of features and influencing factors, indicated by test subjects themselves. We now describe the technical and experimental set-up in more detail.

## B. Procedure

The test consisted of three successive phases:

### 1) Phase 1: pre-questionnaire

Before the actual experiment started, participants were asked to fill in a traditional paper questionnaire consisting of closed and open questions. The questionnaire inquired after their socio-demographic characteristics, type and connection possibilities of their current mobile phone, and experiences and habits (in terms of viewing frequency, ranging from never to several times a day) regarding the watching of video content on a mobile phone. Next, by means of the first open question, the test subjects with prior experience were asked to specify which characteristics (QoE features) related to mobile video watching they personally consider to be essential for having a good experience. 'Open' in this context means that no pre-defined answer categories were given and that test subjects were able to express themselves in their own words. Thereupon, the test subjects who had no prior experience with mobile video watching were able to indicate what according to them might influence their experience. By means of this second open question, we wanted to gain more insight in the test subjects' expectations with regard to possible influencing factors. Finally, the test subjects were asked to indicate how (un)important they considered a number of listed aspects in order to have a good experience during the watching of video content on a mobile phone. These aspects and their importance for mobile video sessions, assessed on a 5-point rating scale going from 1 (not important at all) to 5 (very important) are listed in Table 5.

### 2) Phase 2: mobile video watching during actual test

After this preliminary questionnaire, test subjects received instructions on how to switch on/off the device (Google Nexus One running on Android 2.1), how to use the touch screen, how to access the test application, and how to select and watch the videos. Since each video watching is followed by a small questionnaire on the device, test subjects were also shown how to fill in this electronic questionnaire using the touch screen and given instructions concerning the interpretation of the questions and operational definitions of the QoE measures. After this briefing session, every test subject was given a device and asked to watch 14 videos, each with a length of approximately two minutes, in a controlled environment (i.e., the research lab of our university).

During the setup of the experiment, these 14 videos were preselected from a large content pool and hence cover a large variety of genres including entertainment, technology, music, film, animation, science, cartoons, and news. Since progressive download is used as transport protocol, video playback can start before the video file is completely downloaded to the

device. For this video playback, the standard video player of the Android operating system is used. The videos are transmitted to the device over a WiFi connection (802.11g) of which the maximum available bandwidth per device can be configured. By limiting the bandwidth of the WiFi connection, (the bandwidth of) different cellular data networks can be simulated.

Table 1 lists the three different connection types used in this experiment, together with their theoretical and measured throughput (i.e., the average rate of successful data delivery over the communication channel) and the standard deviation of the measured throughput. The theoretical throughput matches the maximum available bandwidth per device as configured in the wireless access point during the experiment. Because of protocol overhead, the actually measured throughput is slightly lower than the theoretical throughput. This actual throughput is calculated by averaging 10 measurements of the download speed, which are performed by an application called SpeedTest [30]. The standard deviation is calculated based on these 10 measurements of the actual throughput. Since the download speed is far more important than the upload speed for video reception on the device, only the download speed was considered. The low standard deviation of the measurements indicates that the throughput was only slightly varying during the experiment.

The high-bandwidth connection used in this experiment represents a WiFi communication channel and has no imposed restrictions. Although the device does not use the full theoretical bandwidth of the wireless connection, the measured throughput is representative for a WiFi connection and sufficient for all mobile services. The low-bandwidth connection of this experiment has a throughput that is typical for a UMTS connection since the planned transceiver capacity of a UMTS network is typically ranging from 400 kbits/s to 700 kbits/s [31]. Belgian network operators plan HSPA (High Speed Packet Access) networks to provide each mobile user with a bandwidth capacity of 1.5 Mbit/s in the downlink channel (this information is based on confidential interviews with an operator). So, the medium-bandwidth connection of the experiment has a throughput that is typical for a HSPA connection that is available for end-users in Belgium. Although real mobile data networks (such as UMTS or HSPA networks) and the proposed shaped WiFi network have a different behavior regarding packet loss and jitter, these differences are hidden by packet retransmissions and data buffering of the reliable transport protocol (TCP/HTTP).

These three connection types enable us to investigate the influence of network throughput on the subjectively evaluated experience, technical quality, and acceptability of a video session. To demonstrate the reference quality of the mobile videos, the first two videos were transmitted to the mobile device using a high-bandwidth connection. Since this high-bandwidth connection has no network limitations influencing the audio-visual quality during video playback, the quality of the video source is the only variable that affects the quality of

the video playback. Therefore, this case is less interesting to study and so the number of videos using this high-bandwidth connection is limited to two in the experiment. The following 12 videos that users had to watch were transmitted over a low- or medium-bandwidth connection, which may introduce video interruptions during playback (6 videos for each connection type).

TABLE 1 THEORETICAL AND MEASURED THROUGHPUT OF THE DIFFERENT CONNECTION TYPES

Connection type	Theoretical maximum throughput (kbit/s)	Average of the measured throughput (kbit/s)	Standard deviation of the measured throughput (kbit/s)
High-bandwidth	54000	14853	342
Low-bandwidth	750	564	9
Medium-bandwidth	1800	1516	44

To investigate the influence of the quality of the video source on the user's QoE during video watching, videos are transcoded into two different quality versions. Table 2 lists the characteristics of these two video versions and shows that the high-quality version has a higher resolution, bitrate, and frame rate compared to the low-quality version. During playback, both quality versions were upsampled by the device and displayed in full-screen. For both versions, the ITU-standard (International Telecommunication Union) H.264 AVC (advanced video coding) is used since it is currently one of the most commonly used formats for the recording, compression, and distribution of (high definition) video. The AAC LC 3 (Advanced Audio Coding, Low Complexity profile 3) compression scheme is used for the audio track. The average audio bitrate of 62kbit/s is rather low, but satisfactory for streaming video on the mobile devices given the moderate quality of the speakers of the smartphone. No noticeable disturbances were audible in the sound. Since test subjects did not have to evaluate the audio quality separately in the experiment, all videos are coded with the same audio bitrate. For each connection type, as many low-quality videos as high-quality videos are used in the experiment. To avoid boredom, the test subjects had to watch all videos only once during the experiment. So summarized, test subjects had to watch 2 videos without network limitations (1 in low quality, 1 in high quality), 6 videos transmitted using a medium-bandwidth connection (3 in low quality, 3 in high quality) and 6 videos transmitted using a high-bandwidth connection (3 in low quality, 3 in high quality).

TABLE 2 TECHNICAL PARAMETERS OF THE MOBILE VIDEO

Low Quality Video Source			
Audio		Video	
Codec	AAC LC 3	Codec	H.264/AVC
Average bit rate	62 kbit/s	Average bit rate	109 kbit/s
Maximum bit rate	81 kbit/s	Maximum bit rate	507 kbit/s
Channels	2	Resolution	144*256
Sampling frequency	44100 Hz	Frame rate	13 fps
High Quality Video Source			
Audio		Video	
Codec	AAC LC 3	Codec	H.264/AVC
Average bit rate	62 kbit/s	Average bit rate	765 kbit/s
Maximum bit rate	81 kbit/s	Maximum bit rate	1815 kbit/s
Channels	2	Resolution	288*512
Sampling frequency	44100 Hz	Frame rate	25 fps

The test subjects were not informed about these changing network characteristics and the variable quality of the video source but received a list of videos with just a thumbnail and the title as additional information (Figure 1 (a)). Selecting a video from this list starts the transmission to the mobile device and the playback of that video. Each user received the same list of videos and each of these videos had a predefined quality and transmission condition which remained the same in every test. The videos of each connection type / quality combination are covering a variety of content genres to ensure that there is no link between on the one hand the content and on the other hand the quality of the video source or the bandwidth of the communication channel.

During each video playback, various technical parameters regarding the network and video are logged. Table 3 shows these objectively-measured parameters with their unit, value, and sampling rate. For each video, the bandwidth of the communication channel and the quality of the video source were determined during the setup of the experiment. The loading time, which is also measured for each video playback, is defined as the time between selecting a video and the moment when the video starts playing. During the playback of the video, multiple rebufferings may be required. The rebuffer time is defined as the time period that video playback is interrupted because the video buffer is (almost) empty and waiting for new data from the network connection. The loading time and rebuffering time are used to investigate the subjective acceptability of video playback interruptions (Section IV E). Through an application called Wireshark [32], the average RTT (round-trip delay time) is measured during each video playback. Wireshark defines the RTT as the difference in capture time of TCP packets with a certain sequence number and the corresponding follow-up acknowledgement packets from the receiver. The objectively-

measured RTT, enables us to investigate the influence of the network delay on the subjective user evaluations (Section IV C).

TABLE 3 OBJECTIVE PARAMETERS

Parameter	Unit	Value	Sampling rate
Video quality	State	{high, low}	Each video
Bandwidth	State	{low, medium, high}	Each video
Loading time	Seconds	$[0, \infty[$	Each video
Rebuffer times	Seconds	$[0, \infty[$	Each rebuffering
Average RTT	Seconds	$[0, \infty[$	Each video

After each video playback, a short questionnaire pops up on the screen of the mobile device. Figure 1 (b) shows a screenshot of this digital questionnaire that test subjects were asked to fill in during the experiment. Through this feedback form, test subjects can evaluate the content (semantic content of the videos), their general experience, the general technical quality and specific features, and finally, the acceptability of the video quality. After the evaluation of the (semantic) content itself, users were firstly asked to rate the overall technical quality of the video. In the briefing for the test subjects, the following operational definition was given: ‘By technical quality, we mean the overall quality of the different technical features that you – as a viewer - can perceive (these include e.g., the sharpness of the image, the synchronization between the sound and image, the fluidity of the video, loading speed, visual artifacts or errors in the video, ...). Other aspects, such as the appreciation of the content of the clip, are not part of this technical quality’. Then, separate questions were provided to assess a number of specific QoE features, being interruptions, fluidity, and loading speed.

The choice of the rating scale might be seen as an important element in the subjective testing methodology. Nevertheless, a direct comparison between four different rating scales based on experimental data showed no overall statistical differences between the different scales [33]. Table 4 lists the questions and the used measurement scales as recommended by ITU-T [34].

A limitation of the followed approach is linked to the instructions given to subjects to focus on specific features. This may have biased the obtained results to some degree as test subjects may have become more aware of and sensitive to these aspects. However, this possible bias is inherent to this type of set-up and hard to avoid. Moreover, without clear instructions and tasks, the results may also be biased as test subjects might be basing their evaluation and ratings on completely different aspects, or have a different understanding of specific features.

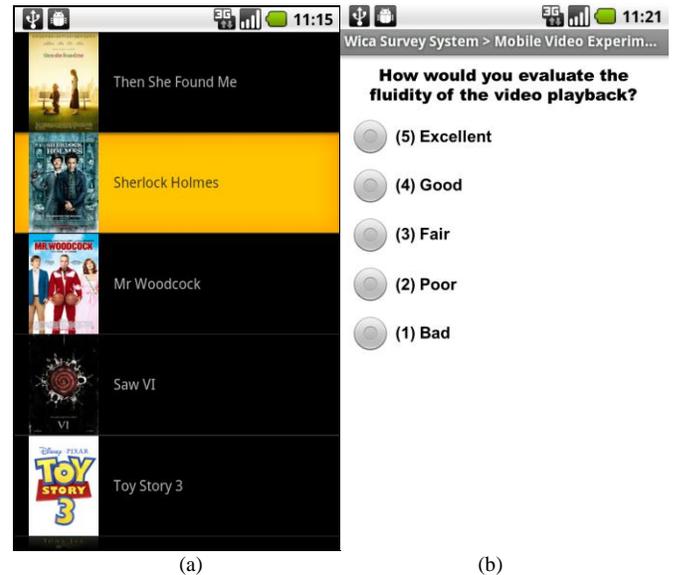


Fig. 1. Screenshots of the video application on the mobile device. The left part (a) shows the video selection mechanism. The right part (b) illustrates the questionnaire which is shown after each video.

TABLE 4 THE QUESTIONS THAT WERE USED TO EVALUATE THE VIDEO IMMEDIATELY AFTER THE PLAYBACK, TOGETHER WITH A REFERENCE TO THESE QUESTIONS AND THE POSSIBLE ANSWERS

Reference	Question	Possible answers
Content	How would you evaluate the content of the video?	5-point rating scale: 1 = bad; 5 = Excellent
Technical Quality	How would you evaluate the technical quality of the video in general?	5-point rating scale: 1 = bad; 5 = Excellent
Interruptions	Did you experience distortions or interruptions as annoying during video playback?	5-point rating scale: 5 (not perceptible); 1 (perceptible and very annoying).
Fluidity	How would you evaluate the fluidity of the video playback?	5-point rating scale: 1 = bad; 5 = Excellent
Loading	How would you evaluate the loading time of the video?	5-point rating scale: 1 = bad; 5 = Excellent
Experience	How would you evaluate your general experience during video playback?	5-point rating scale: 1 = bad; 5 = Excellent
Acceptability	Would you evaluate the technical quality of this video as acceptable?	Binary answer: a) acceptable, b) not acceptable

In total, 12 sessions were organized (in groups of maximum five test subjects), since five Nexus One devices rotated among the test subjects. During a period of two weeks, 57 users (38 men and 19 women) participated in the experiment. The average age of the participants is 29.5 with a standard deviation of 5.2. As most of them work or study at the university, the sample is composed of researchers, project managers, students, secretaries, and maintenance personnel. After checking the data in terms of their completeness, the technical data and subjective evaluations from the questionnaire were coupled and integrated into one data file,

containing 785 samples, which could be used for further analysis and which is enough for drawing statistically-founded conclusions [35].

### 3) Phase 3: post-questionnaire

After the subjective experiment, test subjects were asked again to evaluate the importance of the aspects of Table 5 in order to have a good experience during mobile video watching. Given the variable quality of the video source and the variable bandwidth of the network connection, test subjects might have changed their opinion about the importance of the various technical aspects of mobile video. Additionally, there was an open question where the test subjects could indicate the three dimensions that according to them are most essential in view of having a good QoE in the context of mobile video watching.

## IV. RESULTS

We first discuss the results of the pre-and post-questionnaire in section A. Thereupon, Section B investigates the differences in terms of the objective measures for each combination of connection type and source quality of the video. Section C discusses the differences in terms of the subjective measures and the correlation between the objective and subjective measures. Section D elaborates further on these subjective measures and investigates which combinations of connection type and source quality receive a significantly different evaluation regarding technical quality and QoE. Finally, Section E discusses the acceptability of the technical quality and the influence of rebufferings on this acceptability.

### A. Pre- and Post-Questionnaire

Figure 2 shows a pie chart visualizing the types of mobile phone, characterized by their technical capabilities, and the number of test subjects owning such a device: the majority of the users (31 of the 57) owns a smartphone (with or without touch screen) enabling them to watch mobile video. However, a question regarding mobile video consumption learned that many of these smartphone users never use their phone for watching mobile video.

Figure 3 shows a pie chart illustrating the test subjects' habits regarding mobile video watching. Although the widespread use of smartphones capable of playing video, the vast majority of respondents (41 of the 57) never watched a video via their mobile phone and only a minority of them (7 of the 57) watches mobile videos on a daily to weekly basis. Reasons for this limited usage of mobile video might be the high expenses of the cellular data transfer (in Belgium), and the battery consumption associated with the video playback.

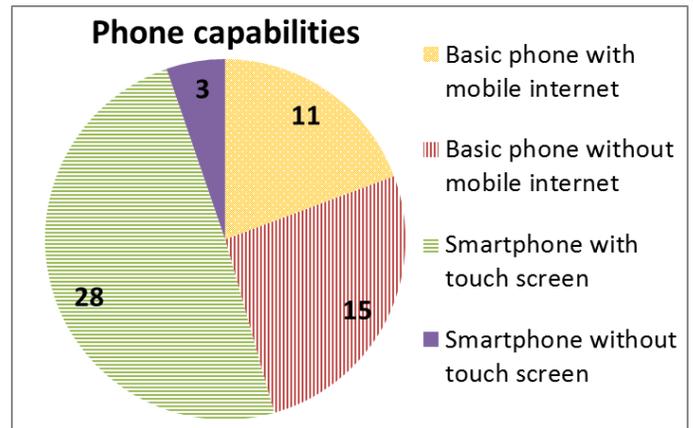


Fig. 2. Pie chart showing the capabilities of the mobile phones that the test subjects own.

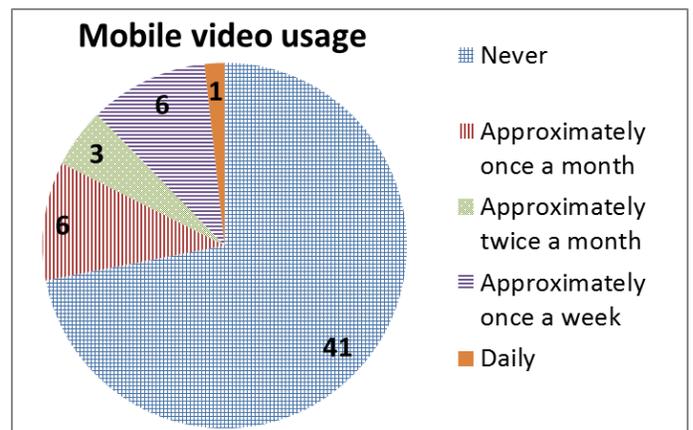


Fig. 3. Pie chart showing the test subjects' habits regarding mobile video watching.

The first open question was answered by almost one third of the participants. These test subjects who had prior experience with mobile video watching indicated which aspects and features that they think are important in view of having a good experience while watching a video on a mobile phone. The answers were coded in broader categories and counted. The figures mentioned here represent a percentage of the number of entries to this question. In general, the most important aspects that were mentioned are the general video quality (22.7%), fluidity during the playback of the video (22.7%) and the audio quality (11.3%). Other aspects mentioned include the screen size and resolution, the absence of distortions, and the loading time of the video. Looking at the aspects that were mentioned first, the most important aspects are fluidity, video quality, and fast loading of the video. As these aspects were mentioned first, we can assume that they are so-called 'top of mind' for several of the test subjects and thus more important.

The answers to the second open question, which was inquiring after the expectations of the test subjects who had no prior experience with mobile video watching, were more diverse. More specifically, they were asked to indicate possible influencing factors, aspects of which they expected that they would influence their personal experience. Most

mentioned in this respect are the loading speed (17.0%): fast loading of the video is expected to contribute positively to the experience, the screen properties (16.0%): the screen should be bright, big enough and have the right resolution for the content, the quality of the audio and video (17.0%), and the fluidity of the video while playing (9.6%): there should be as little buffering as possible. Also mentioned several times are: synchronization of audio and video (3.2%), absence of distortions (4.3%), video player characteristics (4.3%), viewing context (7.4%), content (8.5%): the likeability of the content but also the availability of content, reliability of the internet connection (5.3%), and associated costs (6.4%). One participant also mentioned the battery of the device.

Finally, all the aspects of Table 5 are evaluated by the test subjects in terms of importance in order to have a good experience during mobile video watching. The second column of Table 5 lists the averages of the users' ratings obtained before the actual video experiment whereas the third column shows the averages of the ratings gathered after the experiment. All aspects are evaluated as important to have a good experience during video watching. "The fluidity of the image during video playback" and "the synchronization of image and sound during video playback" received the highest ratings.

The third column of Table 5 shows how test subjects evaluated the listed aspects immediately after the video experiment. Possibly influenced by the variable quality of the video sessions during the experiment, users slightly changed their assessment compared to their initial ratings. The bold values in Table 5 indicate a significant difference between the ratings that users gave before the experiment and the ratings specified by these users after the experiment for the aspects mentioned in the first column. The significance of these differences is determined by the Wilcoxon signed rank test at a significance level of 0.05. The Wilcoxon signed rank test is a non-parametric statistical hypothesis test used to compare two related samples or repeated measurements on a single sample to assess whether their population means differ [36]. In this case, the two subjective ratings (before and after the experiment) originating from the same test subject are compared. Given the discrete values of the users' ratings, we opted for the non-parametric Wilcoxon signed rank test. However, the parametric counterpart of Wilcoxon's test, the statistical T-test identifies the same statistical significant differences.

Especially for those features that were impacted by the different connection types and video qualities, the differences could point to the adjustment of the test subjects' expectations and evaluation, based on their previous experience (i.e., during the test). For future research, it would be interesting to investigate how different levels of expectations (e.g., not met, met, exceeded) relate to specific quality levels and acceptability thresholds and how current experience help to form or adjust those expectations.

As mentioned in section III B 3, there was an additional

open question asking the participants to prioritize the three dimensions that according to them are most important in view of having a good QoE in the context of mobile video watching. When looking at all dimensions that were mentioned, most important are the fluidity of audio, the fluidity of audio and video in general, and the synchronization of audio and video. Additionally, the absence of distortions, the content, sharpness of the video image, and the loading time are considered to be of high importance. When we zoom in on the aspects that were mentioned first (highest priority), the most important aspects are the fluidity (both in general and of audio in particular), the absence of distortions, and the content.

To summarize, after the experiment test subjects attach significantly more importance to the technical quality of the video, the lack of distortions in the image, and the fluidity of the image during video playback. This increased importance is due to the fact that some test subjects assess these technical aspects as unacceptable for some video sessions of the experiment. In contrast, the loading speed of the video is evaluated as less important after the experiment. This indicates that test subjects assess all loading times of the experiment as acceptable and attach less importance to a short loading time than to a fluent playback of the video.

TABLE 5 BEFORE AND AFTER THE EXPERIMENT, USERS WERE ASKED TO EVALUATE THE IMPORTANCE OF THESE ASPECTS IN ORDER TO HAVE A GOOD EXPERIENCE DURING MOBILE VIDEO WATCHING

	Average rating before the experiment	Average rating after the experiment
The content of the video	4.2	4.0
The technical quality of the video	<b>4.0</b>	<b>4.3</b>
The lack of distortions in the image during video playback	<b>4.3</b>	<b>4.5</b>
The fluidity of the image during video playback	<b>4.6</b>	<b>4.8</b>
The lack of distortions in the sound during video playback	4.3	4.5
The fluidity of the sound during video playback	4.4	4.6
Synchronization of image and sound during video playback	4.5	4.4
The loading speed of the video	<b>3.9</b>	<b>3.6</b>
The readability of text on the screen during video playback	3.8	3.6
The sharpness of the image during video playback	3.9	4.1
Other aspects: ...(to be completed by the user)	/	/

### B. Objective Measures

Table 6 shows the technical details regarding the video rebufferings and loading time, which are logged during each video playback. Although the loading time is acceptable for all connection types and quality versions of the video source, the median shows some characteristic differences for the six cases. Low-bandwidth connections induce longer loading times than medium- or high-bandwidth connections. As expected, the

fastest loading times are measured for videos transmitted over a high-bandwidth connection and the high-quality videos require higher bitrates thereby causing longer loading times compared to low-quality video sources.

As mentioned earlier, bandwidth limitations can introduce interruptions during video playback due to rebufferings. However, the number of rebufferings and the point in time when such a rebuffering occurs, i.e. the rebuffering pattern, is non-trivial due to a number of interactions and correlations on several layers of the ISO/OSI stack [17]. The streaming server might implement flow control on the application layer; TCP implements flow control on the transport layer; the video player implementation (the build-in Android player in this experiment) tries to overcome interruptions by means of a video buffer; and the videos are encoded with variable bit rates. Still, differences in the rebuffer times are noticeable between low- and high-bandwidth connections as well as between low- and high-quality video sources. The median as well as the maximum of the measured rebuffer times are slightly higher for low-bandwidth connections and high-quality video sources compared to high-bandwidth connections and low-quality video sources.

Table 6 illustrates that only a small number of rebufferings is required (median = 1) if a high-bandwidth connection is used or if a low-quality video source is transmitted over a medium-bandwidth connection. In these cases, the network connection provides sufficient throughput to transmit the video and prevent interruptions during video playback. For most video scenes, also a low-bandwidth connection provides sufficient throughput to transmit the low-quality video source. However, peaks in the (variable) bitrate of the video may occasionally introduce rebufferings, which explains why the median of the number of rebufferings is 3 in this case.

On the other hand, the throughput obtained by using a low-bandwidth connection is insufficient for transmitting high-quality video sources fluently. This is confirmed by Table 6, which shows a large difference in the number of rebufferings for the high-quality video sources transmitted over a low-bandwidth connection compared to the other cases (e.g., the median of the number of rebufferings is 75 for high-quality video sources transmitted over a low-bandwidth connection). Also a medium-bandwidth connection provides insufficient throughput to transmit a high-quality video source without requiring rebuffering interruptions during playback (median of 45 rebufferings). Peaks in the video bitrate sometimes exceed the available network throughput. Still, the higher throughput of the medium-bandwidth connection compared to the throughput of the low-bandwidth connection reduces the (median of the) number of rebufferings by about half. Table 6 also shows that the standard deviation of the number of rebufferings during video playback is relatively high. Noise in the communication channel and the variable bitrate of the different videos result in a varying number of rebufferings for each combination of connection type and source quality. Therefore, the influence of these rebuffering interruptions on

the different measures of QoE is investigated in Section IV E.

TABLE 6 DETAILS ABOUT THE MEASURED REBUFFERING AND LOADING TIMES FOR THE DIFFERENT CONNECTION TYPES (LOW, MEDIUM, OR HIGH BANDWIDTH) AND QUALITY VERSIONS OF THE VIDEO SOURCE (LOW OR HIGH QUALITY)

	Low-quality source	High-quality source	Low-quality source	High-quality source	Low-quality source	High-quality source
	Low bandwidth	Low bandwidth	Medium bandwidth	Medium bandwidth	High bandwidth	High bandwidth
Median loading time (seconds)	5.7	6.4	3.0	4.3	1.7	1.9
Median of a single rebuffer time (seconds)	0.9	1.0	0.7	1.0	0.7	0.7
Maximum single rebuffer time (seconds)	6.4	9.0	6.4	8.7	5.5	5.9
Median of the number of rebufferings	3	75	1	41	1	1
Standard deviation of the number of rebufferings	2.8	64.4	2.5	43.5	1.1	0.7
Median of the loading + total rebuffer time (seconds)	8.9	85.2	4.6	48.2	2.2	2.9
Standard deviation of the loading + total rebuffer time (seconds)	4.1	68.2	3.3	46.6	1.9	1.6

In general, the period that video playback is interrupted by a rebuffering is quite short. Many interruptions last only a few hundred milliseconds and are hardly noticeable for the end-user, (the median of this rebuffer time is 1 second or less). However, summing the (possible large amount of) rebufferings and the initial loading time of the video results in a substantial waiting time for the end-user, ranging from 2.9 seconds for the most optimal solution to 85.2 seconds for the worst case. Therefore, we expect this waiting time together with the high frequency of rebufferings and the coupled video interruptions might deteriorate the quality of the user's experience significantly for some cases. The varying number of rebufferings for each combination of connection type and source quality results in a high standard deviation of the sum of the loading time and the total rebuffer time.

Given the high frequency of rebufferings and the short rebuffer times, the user's QoE might be improved by enlarging the buffer size thereby increasing the rebuffer times but reducing the frequency of rebufferings. However, since the

build-in media player of the Android OS was used, changing the frequency of rebufferings or the buffer size was not possible in this experiment.

### C. Subjective measures

We first take a closer look at the evaluation of the content and technical quality for the different technical scenarios. The histogram of Figure 4 visualizes the number of ratings gathered for each possible answer (going from 1 = bad to 5 = excellent) to the question in Table 4 regarding the content (question one). These subjective content evaluations provided by the participants of the experiment are partitioned according to the connection type and the quality of the video source. The large number of positive ratings (4 = good or 5 = excellent) indicates that most users appreciate the content of the video experiment.

Moreover, the histogram illustrates that videos sent over a high-bandwidth connection received almost no negative evaluations regarding the content whereas video sessions using a medium- or low-bandwidth connection received a considerable number of negative assessments. Especially the content of video sessions in which a high-quality video is sent over a low-bandwidth connection (highQ lowB) is poorly evaluated. So, the video sessions which suffered from the most rebufferings due to insufficient throughput of the network connection received the worst evaluation regarding the video content. More than 28% of the content ratings are negative i.e., 1 = bad or 2 = poor (Figure 4). Also the high-quality video sources transmitted over a medium-bandwidth connection (highQ mediumB), which are also characterized by a lot of rebufferings, received a considerable number of negative evaluations regarding the content (12% of the content ratings are 1 = bad or 2 = poor). On the other hand, less than 4% of the video content that is transmitted over a high-bandwidth connection is negatively evaluated by the users (i.e. received a rating of 1 = bad or 2 = poor).

This difference in content appreciation, which is unlikely due to coincidence, indicates an effect of the technical quality of the video playback (and the coupled rebuffering interruptions) on the subjective evaluation of the content of the video. This confirms the results of our previous research [37], which states that the user’s subjective evaluation of the content is a combination of the user’s preferences regarding the content and the subjective evaluation of the technical quality of the video.

Figure 5 shows the histogram of the ratings evaluating the technical quality partitioned according to the connection type and the quality of the video source. This histogram visualizes the test subjects’ answers concerning question two of Table 4. High-quality video sent over a low-bandwidth connection (highQ lowB) received the worst evaluation from the test subjects as the majority of these sessions (71%) are evaluated as “bad” or “poor” on the technical quality. The reason for this poor evaluation may be the high number of rebufferings and the coupled playback interruptions due to the low-bandwidth

connection, as indicated in Table 6. Transmitting such a high-quality video over a medium-bandwidth connection (highQ mediumB), decreases the number of rebufferings by approximately 50% but still results in a suboptimal technical quality, as indicated by the considerable number of videos evaluated as “bad” or “poor” technical quality. However, the majority of the users assesses the quality of these video sessions as “fair” and the evaluations are roughly equally divided between positive and negative.

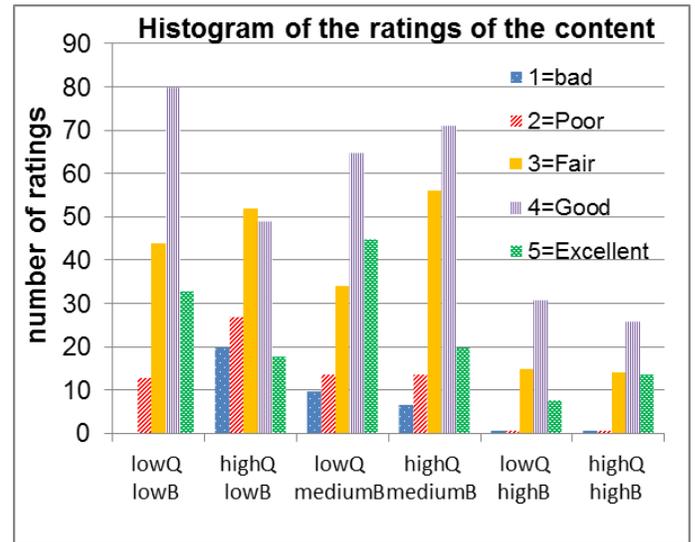


Fig. 4. Histogram of the test subjects’ ratings evaluating the content according to the connection type (low, medium, or high bandwidth (B)) and the quality (Q) of the video source (low or high). 1= very bad, 2= poor, 3=fair, 4=good, 5=excellent.

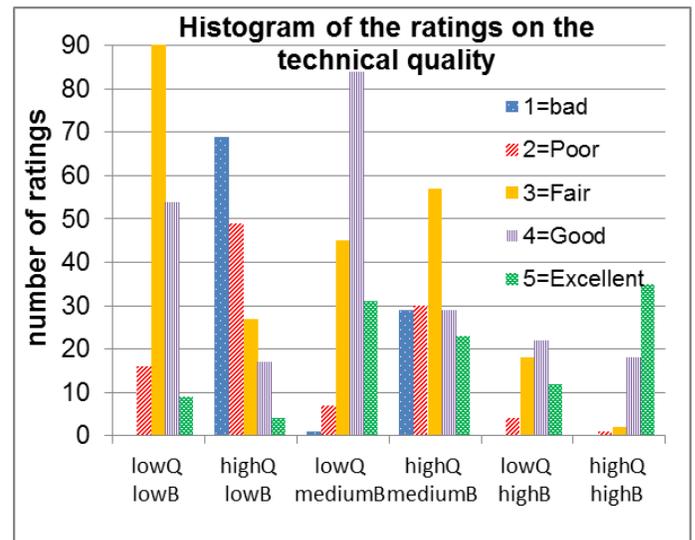


Fig. 5. Histogram of the test subjects’ ratings evaluating the technical quality of the video according to the connection type (low, medium, or high bandwidth (B)) and the quality (Q) of the video source (low or high). 1= very bad, 2= poor, 3=fair, 4=good, 5=excellent.

These two scenarios (high-quality video that is transmitted over a low- or medium-bandwidth connection) are the only scenarios which introduce a large number of rebufferings

during video playback. Accordingly, only these scenarios received a considerable number of very negative evaluations (1=bad) regarding the technical quality from the test subjects. Other scenarios, in which video playback is not or only a few times interrupted, receive in general only neutral or positive evaluations on the technical quality.

Transmitting a low-quality video over a low-bandwidth connection (lowQ lowB) requires an acceptable number of rebufferings (the median of the number of rebufferings is 3 in Table 6). This almost fluent video-playback, together with the low bitrate and resolution induces a fair evaluation of the technical quality (Figure 5), without many extreme positive or negative evaluations. Transmitting such a low-quality video over a medium-bandwidth connection (lowQ mediumB) reduces the number of rebufferings, which is reflected in more positive evaluations.

Video sessions using a high-bandwidth connection (lowQ highB and highQ highB) experience no or a very limited number of rebufferings thereby obtaining a very positive evaluation of the technical quality. E.g., 56% of the low-quality video sources and 97% of the high-quality video sources transmitted over a high-bandwidth connection received a rating of 4 = good or 5 = excellent on the technical quality. As expected, the best results are obtained by transmitting a high-quality video over a high-bandwidth connection (highQ highB). The high resolution and bitrate together with the fluent video playback convince users to evaluate the technical quality of these sessions as “good” or “excellent”.

As indicated in Table 7, the subjective evaluation of the video are positively correlated to each other. All these 2-tailed Pearson correlations are significant at the level of 0.99 ( $p < 0.01$ ). The quality of the video source and the available bandwidth of the communication channel is the common factor that influences the subjective evaluations of all these video characteristics. As a result, the histograms of all these subjective evaluations have similar distributions and do not reveal additional insight; so they are omitted in this paper.

Table 7 also shows that the subjective evaluations of the technical quality and the overall experience are positively correlated (based on the Pearson correlation), proving the consistency of these general subjective parameters. Also the acceptability of the technical quality is in line with these subjective parameters: if the quality is evaluated as ‘acceptable’, the average of the subjective evaluations of the technical quality and overall experience are respectively 3.8 and 3.5 whereas for video sessions assessed as ‘unacceptable’, these averages are respectively 2.1 and 2.0.

TABLE 7 CORRELATIONS BETWEEN THE SUBJECTIVE USER EVALUATIONS.

	Content	Technical quality	Interruptions	Fluidity	Loading	Experience
Content	1.000	0.300	0.245	0.272	0.298	0.593
Technical quality	0.300	1.000	0.744	0.737	0.507	0.595
Interruptions	0.245	0.744	1.000	0.826	0.563	0.598
Fluidity	0.272	0.737	0.826	1.000	0.562	0.625
Loading	0.298	0.507	0.563	0.562	1.000	0.495
Experience	0.593	0.595	0.598	0.625	0.495	1.000

To quantify the influence of the objectively-measured parameters of Table 6 on the subjective evaluations of the video obtained via the questionnaire (Table 4), we calculated the statistical correlations (using the Pearson correlation coefficient,  $\rho$ ). Table 8 lists the correlations between these parameters of the video session (i.e. the number of rebufferings, the waiting time, which is defined as the sum of loading and rebuffer times, and the average RTT) and the subjective user evaluations regarding the aspects of Table 4. All these 2-tailed Pearson correlations are significant at the level of 0.99 ( $p < 0.01$ ). The results show a strong negative correlation between on the one hand the subjective evaluation of the user’s experience and the ratings related to the technical quality of the video session (Quality, Interruptions, Fluidity, and Loading), and on the other hand the number of rebufferings and the time that these rebufferings require (together with the initial loading time of the video). This confirms that the subjective evaluation of the video quality and the coupled QoE are strongly influenced by the duration and amount of interruptions during video playback. A negative correlation ( $\rho = -0.35$ ) is also observed between the average RTT and the evaluation of the experience as well as the ratings related to the technical quality of the video session (Quality, Interruptions, Fluidity, and Loading). So, high round-trip delay times will have a negative influence on the users’ QoE and the subjectively-observed video quality.

Noteworthy is the significant negative correlation ( $\rho = -0.20$ ) between on the one hand the users’ evaluation of the content and on the other hand the number of rebufferings, the time that these rebufferings require, and the average RTT. Although users were supposed to evaluate the content of the video regardless of the loading time, rebuffering interruptions, network characteristics, and technical quality of the video, content ratings are clearly influenced by these technical parameters.

TABLE 8 CORRELATIONS BETWEEN THE OBJECTIVE PARAMETERS OF THE VIDEO SESSION AND THE SUBJECTIVE USER EVALUATIONS

	Number of rebufferings	Waiting time= loading + rebuffer times	Average RTT
Content	-0.203	-0.208	-0.191
Quality	-0.552	-0.562	-0.421
Interruptions	-0.613	-0.622	-0.393
Fluidity	-0.705	-0.712	-0.397
Loading	-0.452	-0.478	-0.441
Experience	-0.510	-0.518	-0.345

#### D. Subjective Technical Quality and Overall Experience.

The correlations of Table 8 prove the influence of the objectively-measured parameters of the video sessions on the subjectively-observed video quality. However, to investigate which technical scenarios show a significant difference in subjective technical quality and overall experience, they are compared pairwise via a statistical test.

A one-way Analysis of Variance (ANOVA) relies on the restrictive assumptions of homogeneity of the variances of the distributions and normality of the distributions of the residuals [35]. Also the commonly-used T-test, a statistical hypothesis test which compares the mean values of 2 groups, relies on the assumption that the samples follow a normal distribution [35]. Since the user evaluations are discrete values, these assumptions may not apply. Therefore, the six video scenarios were compared using the Wilcoxon rank sum test as alternative. The Wilcoxon rank sum test is a non-parametric statistical hypothesis test for assessing whether one of two samples of independent observations tends to have larger values than the other [37]. This way, the subjective ratings of the technical quality and of the overall experience (dependent variables) were compared using the different technical combinations (connection type & source quality) as the grouping variable (independent variable).

Table 9 shows the results of this Wilcoxon rank sum test, performed on the subjective ratings of the general technical quality and the overall experience during video playback using a significance level of 0.05 ( $p < 0.05$ ). The second column specifies which two scenarios (characterized by the connection type and the quality of the video source) are tested for a significant difference in the mean rating specified by the user. For reference purpose, each of these tests received a sequence number in first column of the table. The third and fifth column show the point estimation of the mean difference between the two scenarios (first scenario minus second scenario) for respectively the technical quality and the overall experience. The standard error on this point estimation is indicated between brackets. The p-value (of the fourth and sixth column) is an indication for the significance of the difference between the two scenarios. If the p-value is below 0.05, the evaluations of the scenarios are considered as significantly

different. Tests 1, 2, 3, 4, and 5 compare the subjectively-observed quality and overall experience of video sessions using two network connections with a different bandwidth. For each of these tests, the quality of the video source is identical for the two scenarios (low quality for tests 1 and 2; high quality for tests 3, 4 and 5), whereas the bandwidth of the connection in the second scenario is higher than the bandwidth of the connection in the first scenario. The significant differences in subjective technical quality and overall experience as well as the negative values of the point estimations of these differences prove that users notice the more fluent video playback (i.e., less and shorter rebufferings as well as a shorter loading time) if a higher bandwidth is available for transmission. Only for test 1 and 2, the difference in overall experience was not found to be significant.

The point estimation of the mean difference in observed technical quality and overall experience is respectively -2.530 and -2.190 for test 4. High-quality video sources that are sent over a low-bandwidth connection are characterized by a large number of rebufferings and receive therefore a low evaluation. High-quality video sources transmitted over a high-bandwidth connection on the other hand, deliver a perfect image quality and require no or a very limited number of rebufferings during playback. Therefore, the biggest difference in subjective technical quality and overall experience is measured for these two extreme situations.

Also test 6 compares the observed technical quality and overall experience of video sessions using two network connections with a different bandwidth. However, this test shows no significant differences if a high-bandwidth connection is used instead of a medium-bandwidth connection for the transmission of a low-quality video. Since a medium-bandwidth connection provides already sufficient throughput for transmitting a low-quality video fluently, switching to a high-bandwidth connection brings no further improvement in the observed technical quality or overall experience.

Test 7 shows a significant difference in observed technical quality and overall experience between high-quality and low-quality video sources that are transmitted over a low-bandwidth connection. The negative values of the point estimations of the mean differences indicate that users provide a better evaluation for the low-quality video source. The reason for this is the high number of rebufferings that users experience if a high-quality video source is transmitted over a low-bandwidth connection. This indicates that in this case users prefer a more fluent playback of the video above a higher resolution, frame rate, and bitrate. So, if the available bandwidth of the data connection is low, content providers can optimize the subjectively-observed quality and overall experience of the video session by transmitting a low-quality video instead of a high-quality video to the end-user.

Tests 8 and 9 further compare the playback of a high-quality video source using a low-bandwidth connection, which introduces a large number of rebufferings, with the (almost) fluent playback of low-quality video. Whereas test 7 uses a

low-bandwidth connection for the transmission of the low-quality video, thereby causing an acceptable number of rebufferings (median=3), tests 8 and 9 transmit the low-quality video using a medium- and high-bandwidth connection, which require no or only a very limited number of rebuffering interruptions (median=1). This further decrease in the number of rebufferings leads to a better technical quality and overall experience. This is reflected in the higher absolute values of the point estimations of the mean differences for the two scenarios.

Tests 10, 11 and 12 compare the subjective quality and overall experience of a high-quality video source transmitted over a medium-bandwidth connection with low-quality video transmitted over various connections. According to the results of test 10, the playback of a low-quality video source results in a better subjective quality and overall experience than the playback of a high-quality video source if the transmission channel has a medium-bandwidth connection. Again, the number of rebufferings and the coupled playback interruptions are the reasons that users prefer a low-quality video above a high-quality video if a medium-bandwidth connection is available. This shows again that users prefer a fluent playback of their video, even if this means that they have to sacrifice resolution and frame rate.

In test 11, a high-bandwidth connection is used as communication channel for the low-quality video in contrast to the medium-bandwidth connection of test 10. This high-bandwidth connection causes no further improvement since a medium-bandwidth connection offers already sufficient throughput for transmitting the low-quality video without introducing too many rebufferings. The test of line 12 shows another interesting result. Video sessions using a low-quality video source and a low-bandwidth connection are significantly better assessed than video sessions based on a high-quality video source and a medium-bandwidth connection. Since the throughput of the medium-bandwidth connection is still insufficient for transmitting high-quality videos and thereby requires too much rebufferings, this test confirms the users' preference for fluent video playback above high-quality video sources.

Test 13 compares two opposite cases: low-quality video over a low-bandwidth connection against high-quality video over a high-bandwidth connection. As expected, the high-quality video using a high-bandwidth connection receives an assessment that is much better than the low-quality video sent over a low-bandwidth connection. Although the estimated differences of the mean ratings for these two scenarios are very significant (-1.220 for technical quality and -0.671 for overall experience), this is not the biggest difference that was encountered in the experiment. (Test 4 showed the biggest difference in ratings between the two scenarios.)

Finally, test 14 and 15 represent cases in which sufficient bandwidth is available for video transmission and the number of video rebufferings remains limited. In test 14, sending a low-quality video over a medium-bandwidth connection is

compared with the transmission of a high-quality video over a high-bandwidth connection. Since video playback is fluent for both cases, the only discriminating factor is the quality of the video source. Therefore, the high-quality video (which is sent over a high-bandwidth connection) is evaluated better than the low-quality video (which uses the medium-bandwidth connection). Finally, test 15 compares high-quality and low-quality video sources which are both transmitted over a high-bandwidth connection. Since this connection provides enough throughput for both quality versions, the difference in technical quality and overall experience is merely based on the difference in the quality of the video sources. As expected, the high-quality video source is assessed significantly higher than the low-quality video source.

TABLE 9 RESULTS OF THE WILCOXON RANK TEST PERFORMED ON THE RATINGS OF THE TECHNICAL QUALITY ON A SIGNIFICANCE LEVEL OF 0.05

	Difference of two scenarios	Mean difference in technical quality (Std)	P value technical quality	Mean difference in QoE (Std)	P value QoE
1	lowQ lowB - low Q mediumB	-0.486 (0.105)	<10 <sup>-4</sup>	-0.105 (0.106)	0.2191
2	lowQ lowB - lowQ highB	-0.421 (0.149)	0.0008	-0.189 (0.151)	0.1382
3	highQ lowB - highQ mediumB	-0.899 (0.106)	<10 <sup>-4</sup>	-0.990 (0.107)	<10 <sup>-4</sup>
4	highQ lowB - highQ highB	-2.530 (0.149)	<10 <sup>-4</sup>	-2.190 (0.151)	<10 <sup>-4</sup>
5	highQ mediumB - highQ highB	-1.630 (0.149)	<10 <sup>-4</sup>	-1.200 (0.151)	<10 <sup>-4</sup>
6	lowQ mediumB - lowQ highB	0.066 (0.149)	0.5793	-0.0833 (0.151)	0.6554
7	highQ lowB - lowQ lowB	-1.310 (0.105)	<10 <sup>-4</sup>	-1.520 (0.107)	<10 <sup>-4</sup>
8	highQ lowB - lowQ mediumB	-1.790 (0.106)	<10 <sup>-4</sup>	-1.630 (0.107)	<10 <sup>-4</sup>
9	highQ lowB - lowQ highB	-1.730 (0.149)	<10 <sup>-4</sup>	-1.710 (0.151)	<10 <sup>-4</sup>
10	highQ mediumB - lowQ mediumB	-0.893 (0.105)	<10 <sup>-4</sup>	-0.637 (0.107)	<10 <sup>-4</sup>
11	highQ mediumB - lowQ highB	-0.827 (0.149)	<10 <sup>-4</sup>	-0.720 (0.151)	<10 <sup>-4</sup>
12	lowQ lowB - highQ mediumB	0.407 (0.105)	0.0008	0.532 (0.106)	<10 <sup>-4</sup>
13	lowQ lowB - highQ highB	-1.220 (0.149)	<10 <sup>-4</sup>	-0.671 (0.151)	<10 <sup>-4</sup>
14	lowQ mediumB - highQ highB	-0.738 (0.149)	<10 <sup>-4</sup>	-0.565 (0.151)	0.0001
15	highQ highB - lowQ highB	0.804 (0.183)	<10 <sup>-4</sup>	0.482 (0.185)	0.0009

The other subjective evaluations regarding the technical properties of the video (Interruptions, Fluidity and Loading) show similar results, also pointing to the consistency of test

subjects in their ratings. Almost every test reveals significant differences between the video scenarios. Even the evaluation of the content of the video shows to be significantly different for various couples of video scenarios. For example, the content of the high-quality video transmitted over a high-bandwidth connection is significantly higher assessed than the content of high-quality video transmitted over a low-bandwidth connection. Again, this supports the assumption that the subjectively-observed technical quality and the overall experience are aspects that influence the subjective evaluation of the content.

### E. Acceptability of the Technical Quality

Besides knowing which video scenarios receive a different evaluation regarding the technical quality and overall experience, it is essential to identify the video scenarios with an acceptable technical quality. This means, video sessions have to be classified as “acceptable quality” or “unacceptable quality”. Therefore, users, not informed about the source quality or the connection type, were asked to evaluate the acceptability of the technical quality of each video during the experiment via the last question of Table 4. Table 10 summarizes this acceptability of the technical quality for the different video scenarios.

The high-quality video sources that are sent over a low-bandwidth connection and thereby require numerous rebufferings during video playback are in general evaluated as “unacceptable”. The long (median duration = 1 second (Table 6)) and frequent (median number = 75 (Table 6)) rebufferings are experienced as annoying and often even intolerable, since only 7.8% of these video sessions are evaluated as “acceptable”.

If a medium-bandwidth connection is used to transmit such a high-quality video, still a considerable number of rebufferings is necessary during the video playback (median number = 41 (Table 6)). Despite this high number of rebuffering interruptions, the technical quality of 39.3% of these sessions is evaluated as “acceptable”. The reason for this is that users’ expectations regarding the technical quality of mobile video services might be quite low, hereby expecting and accepting interruptions during video playback.

The sessions in which low-quality video is transmitted over a low-bandwidth connection undergo a limited number of rebufferings (median number = 3 (Table 6)). The combination of this low-quality video source and the small number of playback interruptions is accepted in 85.3% of the cases.

Since low-quality video can fluently be transmitted over a high- or medium-bandwidth connection without requiring rebufferings, the technical quality of these video sessions is almost always acceptable (in respectively 91.1% and 94.6% of the cases).

The highest acceptance rate (96.4%) of this experiment is measured for high-quality video sources, transmitted over a high-bandwidth connection. The fluent playback of this high-quality video provides the most optimal video rendering on the

mobile device but requires a high throughput to prevent rebuffering interruptions.

TABLE 10 EVALUATION OF THE ACCEPTABILITY OF THE OBSERVED VIDEO QUALITY FOR THE DIFFERENT COMBINATIONS OF CONNECTION TYPE AND QUALITY OF THE VIDEO SOURCE

	Total number of ratings	Number of acceptable sessions	Number of unacceptable sessions	Rate of acceptance (%)
High-quality source Low bandwidth	166	13	153	7.8
High-quality source Medium bandwidth	168	66	102	39.3
Low-quality source Low bandwidth	170	145	25	85.3
Low-quality source High bandwidth	56	51	5	91.1
Low-quality source Medium bandwidth	168	159	9	94.6
High-quality source High bandwidth	56	54	2	96.4

To obtain a model to predict the acceptability of the technical quality of the video session, a logistic regression analysis was performed. Logistic regression is used to predict the probability of an event (in this case, the rejection of the video quality) by fitting data to a logistic curve [35]. In contrast to the analysis of the subjectively-observed technical quality and overall experience of Section IV D, the acceptability of the technical quality is modeled via a logistic regression, because of the binary nature of this evaluation.

Because of the significant correlations between the subjective evaluations and the waiting time (i.e. the sum of the loading time and rebuffer times during video playback), we opted for the waiting time as a predictor variable (independent variable) and the acceptability of the technical quality is chosen as the dependent variable. The result of this logistic regression analysis is a model for the probability that the user will not accept the quality of the video,  $p$ . This equation (Eq. (1)) illustrates that the probability of an unacceptable quality increases as the waiting time increases. A critical point is reached if the waiting time becomes more than 39 seconds, since the probability of an unacceptable quality is then higher than 50%.

This model is based on the subjective evaluations of the acceptability of the technical quality of 782 video sessions. The null deviance of this model is 1039 whereas the residual deviance is 547, which is smaller than the 95% quantile of the  $\chi^2$  distribution with 782 degrees of freedom i.e.,  $\chi^2(0.95, 782) = 848$ . This statistical test confirms that the data is distributed according to the proposed logistic regression model [35].

$$p = \frac{e^{-2.4437+0.0626\text{WaitingTime}}}{1 + e^{-2.4437+0.0626\text{WaitingTime}}} \quad (1)$$

Figure 6 visualizes the result of this logistic regression analysis by plotting the probability of an unacceptable technical quality as a function of the waiting time which is varying from 0 to 130 seconds (line diagram). Figure 6 also compares this logistic curve with the subjective evaluations of the acceptability obtained through the questionnaire of the experiment. Therefore, the video sessions of the experiment are classified according to the objectively-measured waiting time. Each of the video classes has a range of 10 seconds in waiting time. Next, the fraction of video sessions that are evaluated as “unacceptable” during the experiment is calculated for each of these classes and visualized in Figure 6 as “measured probability”. The graph shows that the predicted probability obtained via logistic regression is a good fit of the measured probability which is based on subjective evaluations. This is confirmed by the RMSE (Root Mean Square Error) of 0.15, which is calculated based on the difference between the predicted probability and the measured probability.

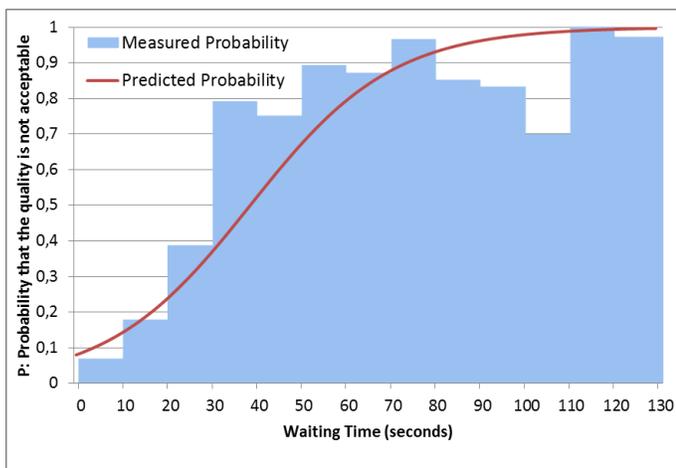


FIG. 6. THE PROBABILITY THAT THE VIDEO QUALITY IS NOT ACCEPTED BY THE USER, AS A FUNCTION OF THE WAITING TIME DURING THE VIDEO SESSION. THE BAR DIAGRAM SHOWS THE MEASURED PROBABILITIES BASED ON THE VIDEO EXPERIMENT. THE LINE DIAGRAM SHOWS THE RESULT OF THE LOGISTIC REGRESSION ANALYSIS.

## V. CONCLUSION

This paper presents the results from an exploratory study on the (acceptability of the) technical quality and overall experience of mobile video sessions. The controlled environment of this experiment allowed us to manipulate the bandwidth of the data connection used to transfer the videos to the mobile device. Three different connection types (low, medium, and high bandwidth) were combined with two levels of audio-visual quality of the video source.

Subjective quality assessments, obtained via a questionnaire, showed to be highly correlated to the objectively-measured parameters of the video session, such as the number of rebufferings, the rebuffer time, and the loading time of the video. Although video interruptions due to rebufferings are experienced as disturbing, users accept a (limited) number of these rebufferings in a mobile context.

Moreover, the subjective evaluations of the video quality learned that the users of our experiment preferred a fluent playback of the video above a higher resolution, frame rate, and bitrate. In comparison with the fluidity of the playback, the users considered the loading time of the video as less critical for having a good experience.

The waiting time during video playback (i.e. the sum of the loading time and the rebuffer times) showed to be determinative for the experience of the user during mobile video watching, and the coupled acceptability of the technical quality. Based on the subjective evaluations of the users and the objectively-measured parameters of the video sessions, this study results in a model for the acceptability of the quality of a mobile video session. This model predicts the probability that users will accept the quality of a mobile video session as a function of the waiting time during video playback. Video sessions with a waiting time below 20 seconds have a high probability (more than 75%) to be accepted by the user, whereas sessions with more than 60 seconds of waiting time are in general (more than 75%) evaluated as not acceptable. This proposed QoE model enables operators to fix performance targets in terms of human perception. Future research should however seek to validate these findings, not only in controlled research settings but also in more ecologically valid usage contexts. The set-up of a complementary living lab or field study, in which the influence of physical as well as social contextual factors can be more closely investigated would be a first step towards a more natural usage environment. Secondly, to take into account the influence of temporal dimensions and effects, a study with a longer time frame (e.g., one to several weeks) could be set up. Finally, it would be very relevant to also look at other types of mobile devices enabling mobile video watching (for instance smartphones vs. tablets) to see if test subjects adjust their expectations and acceptability thresholds depending on the technical context (e.g., screen size).

As such, it could be further investigated which additional factors might affect users’ overall experience and their acceptance or refusal of the produced quality as well as how these factors can be taken into account in order to optimize the experience.

## ACKNOWLEDGMENT

The authors would like to thank the Multimedia Lab (MMLab) research group from Ghent University for their help with the video server.

## REFERENCES

- [1] RapidTVNews, “Live and downloaded mobile video grow strongly,” 2011. [Online]. Available: <http://www.rapidtvnews.com/index.php/2011033011228/video-increasingly-crucial-to-mobile-development.html>,
- [2] Cisco Inc. “Visual networking index: Global mobile data traffic forecast update, 2010–2015,” 2011. [Online]. Available: [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-520862.html](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html).

- [3] K. Mun, "Mobile video calling: improving QoE. Techzine, Technology And Research E-Zine," 2011. [Online]. Available: <http://www2.alcatel-lucent.com/blogs/techzine/2011/mobile-video-calling-improving-qoe/>
- [4] L. Karam, T. Ebrahimi, S. Hemami, T. Pappas, R. Safranek, Z. Wang, and A. Watson, "Introduction to the special issue on visual media quality assessment". *IEEE J. Sel. Topics Signal Process.* vol. 3, no. 2, pp. 189-192, 2009.
- [5] S. Chikkerur, S. Sundaram, M. Reisslein, and L.J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Trans. Broadcasting.* vol. 57, no. 2, pp. 165-182, 2011.
- [6] U. Reiter, "Overall perceived audiovisual quality - what people pay attention to," *IEEE 15th International Symposium on Consumer Electronics.* pp. 513-517, 2011.
- [7] L. A. Rowe, and R. Jain, "ACM SIGMM retreat report on future directions in multimedia research," *ACM Trans. Multimedia Comput. Commun. Appl.* vol. 1, no. 1, pp. 3-13, 2005.
- [8] S. David, L. Man, and C. Renaud, "QoS and QoE management in UMTS cellular systems," Wiley, 2006.
- [9] ITU-T, "Definition of Quality of Experience (QoE)," International Telecommunication Union, Liaison Statement, Ref.: TD 109 rev 2 (PLEN/12), 2007.
- [10] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "Assessing Quality of Experience of IPTV and video on demand services in real-life environments," *IEEE Trans. Broadcasting.* vol.56, no.4, pp.458-466, Dec. 2010.
- [11] S. Balasubramaniam, J. Mineraud, P. McDonagh, P. Perry, L. Murphy, W. Donnelly, and D. Botvich, "An evaluation of parameterized gradient based routing with QoE monitoring for multiple IPTV providers," *IEEE Trans. Broadcasting.* vol.57, no.2, pp.183-194, June 2011.
- [12] K. Piamrat, C. Viho, J.-M. Bonnin, and A. Ksentini, "Quality of experience measurements for video streaming over wireless networks," *Third International Conference on Information Technology: New Generations*, pp. 1184-1189, 2009.
- [13] R. Mok, C. Chan, and R. Chang, "Measuring the quality of experience of HTTP video streaming". In *Proc. IFIP IM IEEE conf. (Pre-conf Session)*, 2011.
- [14] M. Claypool, and J. Tanner, "The effects of jitter on the perceptual quality of video," In *Proc. of the 7th ACM international conf. on Multimedia.* ACM, New York, NY, 1999.
- [15] K. Seshadrinathan, R. Soundararajan, A.C. Bovik, and L.K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Processing.* vol. 19, no. 6, pp. 1427-1441, 2010.
- [16] F. Agboma, and A. Liotta, "QoE-aware QoS management," In *Proc. of the 6th International Conf. on Advances in Mobile Computing and Multimedia*, Linz, Austria, pp. 111-116, 2008.
- [17] T. Hoßfeld, R. Schatz, T. Zinner, M. Seufert, and P. Tran-Gia, "Transport Protocol Influences on YouTube Videostreaming QoE," University of Würzburg Institute of Computer Science Research Report Series. [Online]. Available: <http://www3.informatik.uni-wuerzburg.de/TR/tr482.pdf>, 2011.
- [18] W. He, K. Nahrstedt, and X. Liu, "End-to-end delay control of multimedia applications over multihop wireless links," *ACM Trans. Multimedia Comput. Commun. Appl.* vol. 1, no. 1, 2008.
- [19] A. Perkis, S. Munkeby, and O.I. Hillestad, "A model for measuring quality of experience," In *Signal Processing Symposium Proc. of the 7th Nordic*. pp. 198-201, 2006.
- [20] R. Silva, P. Carvalho, P. Sousa, and P. Neves, "Enabling heterogeneous mobility in android devices," *Mobile Networks and Applications*, Springer Netherlands. vol. 16, no. 4, pp. 518-528, 2011.
- [21] O. Bradeanu, D. Munteanu, I. Rincu, and F. Geanta, "Mobile multimedia end-user quality of experience modeling," In *Proc. of the International Conf. on Digital Telecommunications (ICDT '06)*. pp. 49-49, 2006.
- [22] O. Teyeb, T.B. Sørensen, P. Mogensen, and J. Wigard, "Evaluation of packet service performance in UMTS and heterogeneous networks," In *Proc. of the ACM Workshop on Quality of Service & Security for Wireless and Mobile Networks*, pp. 95-102, 2006.
- [23] X. Tan, J. Gustafsson, and G. Heikkilä, "Perceived video streaming quality under initial buffering and rebuffering degradations," MESAQIN conference, Prague, June 5-6, 2006.
- [24] J. Gustafsson, G. Heikkilä, M. Pettersson, "Measuring multimedia quality in mobile networks with an objective parametric model," In *Proc. of the International Conference on Image Processing, 2008. ICIP 2008.*, pp.405-408, 12-15 Oct. 2008.
- [25] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen. "Initial delay vs. interruptions: between the devil and the deep blue sea," *QoMex 2012: Fourth International Workshop on Quality of Multimedia Experience.* July 4-7, 2012, Yarra Valley, Australia.
- [26] K. De Moor, I. Ketyko, W. Joseph, T. Deryckere, L. De 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*, Springer Netherlands. vol. 15, no. 3, pp. 378-391, 2010.
- [27] 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'*, vol. 10, pp. 99-131, 2008.
- [28] Le Callet, P., Möller, S. and Perkis, A. (Eds.), "Qualinet white paper on definitions of quality of experience (QoE)", 2012. Available at: <http://www.qualinet.eu/>.
- [29] Jumisko-Pyykkö, S., Malamal Vadakital, V.K. and Hannuksela, M. M., "Acceptance threshold: A bidimensional research method for user-oriented quality evaluation studies," *International Journal of Digital Multimedia Broadcasting*, 2008, Article ID 712380.
- [30] Speedtest.net, "The global broadband speed test," [Online]. Available: <http://www.speedtest.net/>, 2011.
- [31] UMTS World, "UMTS network capacity planning," [Online]. Available: <http://www.umtsworld.com/technology/capacity.htm>, 2003.
- [32] Wireshark, "The world's foremost network protocol analyzer," <http://www.wireshark.org/>, 2011.
- [33] Q. Huynh-Thu, M.-N. Garcia, F. Speranza, P. Corriveau, A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Trans. Broadcasting.* vol.57, no.1, pp.1-14, March 2011.
- [34] ITU-T, Rec. P.911. "Subjective audiovisual quality assessment methods for multimedia applications," International Telecommunication Union - Telecommunication sector, 1998.
- [35] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, "Applied linear statistical models", McGraw-Hill/Irwin, 5th edition, 2005.
- [36] J.J. Higgins, "Introduction to modern nonparametric statistics. Stamford, CT: Thomson Learning, 2004.
- [37] T. De Pessemier, K. De Moor, I. Ketykó, W. Joseph, L. De Marez, and L. Martens, "Investigating the influence of QoS on personal evaluation behaviour in a mobile context", *Multimedia Tools and Applications*, Springer Netherlands, vol. 57, no. 2, pp. 335-358, 2012.