

# Investigating Quality of Experience in the context of adaptive video streaming: findings from an experimental user study

Bjørn J. Villa<sup>1</sup>, Katrien De Moor<sup>1,2</sup>, Poul E. Heegaard<sup>1</sup>, Anders Instefjord<sup>3</sup>

<sup>1</sup>Department of Telematics, Norwegian University of Science and Technology, Trondheim, Norway

<sup>2</sup>iMinds-MICT, Department of Communication Studies, Ghent University, Ghent, Belgium

<sup>3</sup>Vimond Media Solutions AS, Bergen, Norway

Corresponding author: bjorn.villa@item.ntnu.no

## Abstract

Although adaptive video streaming solutions have become very popular over the last years, only a limited number of studies so far have investigated Quality of Experience (QoE) in the context of such services from a real user perspective. In this paper, we present results from a user study (N=32) on adaptive video streaming QoE. Content (i.e., music video clips) was streamed on portable terminal devices (namely iPads), using different bitrate profiles which represent realistic bandwidth conditions. All users conducted the test in two different usage scenarios, representing different social settings. QoE was evaluated in terms of traditional QoE-measures and complemented with a number of alternative, affective-state related measures. The findings indicate differences between the considered bitrate profiles in terms of QoE (traditional and alternative measures) and indicate that a lower, constant bitrate seems to be preferred over the changing bitrate which is - on average - higher. This is the case for both separate usage scenarios. Although we found no significant influence of the social setting when considering traditional QoE evaluation measures, some significant differences between both settings were detected when evaluating QoE in terms of delight and annoyance.

## 1. Introduction

Internet traffic has tremendously increased over the past years and is expected to still increase four times in the period 2010-2015 [1]. Forecasts indicate that by 2015, consumer Internet traffic will be dominated (growing to over 60%) by Internet video and that video in its different forms will represent around 90% of the global consumer traffic. For mobile traffic, the estimated volume of video services amounts up to around 90% in 2015 [2]. This evolution has been driven by the availability of broadband access at affordable prices, in parallel with new types of end user equipment. It also has important implications for the delivery of services to end users. The opportunity to provide advanced services as part of the best-effort Internet traffic class has been utilized by many successful service providers, such as YouTube and Netflix. This type of service delivery is commonly known as the Over-The-Top (OTT) model [3], where any service provider can provide services to any customer – on a global basis.

The OTT service delivery model provides opportunities, but also introduces technical challenges. The best-effort Internet traffic class gives no guarantees in terms of QoS metrics such as e.g., bandwidth. In order for services to perform well under such conditions, the concept of adaptive services has emerged. These services have the ability to change their requirements during service delivery according to experienced network characteristics. Looking at video services in particular, the concept of Dynamic Adaptive Streaming over HTTP (DASH) has been the subject of recent standardization

*This paper was presented at the NIK 2013 conference. For more information see <http://www.nik.no/>*

*This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme (Katrien De Moor). The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 246016."*

efforts [4], and is already in use by major content providers. It has also been argued that adaptive streaming will become even more important in the future: Oyman and Singh [2] refer to a recent study by TDG Research, which forecasted that by 2015, 51% of Internet video will be supported by adaptive streaming solutions. Further on, the results from a survey conducted by the DASH Industry Forum among European Broadcasters [5], show that the majority of them will deploy DASH solutions during 2013 and 2014.

Despite the increasing popularity of DASH based services and its assumed capability to providing better QoE, only a limited number of studies so far have investigated this topic from a real user perspective. This means e.g., considering specific factors that may have an influence, and considering different usage scenarios. In this paper, we therefore aim to 1) investigate QoE in the context of different adaptive video streaming profiles, and 2) explore the possible influence of social setting on QoE in this respect. To this end, we present results from an experimental user study in which three recent music video clips were streamed on iPads as portable terminal devices, and evaluated in terms of QoE by 32 test subjects. Since bandwidth is still seen as one of the most important bottlenecks [2], three different bitrate conditions were used. These conditions correspond with three distinct, realistic bandwidth and thus video delivery *profiles* (i.e., constant high bitrate, constant low bitrate, changing bitrate from high to low). Two different usage scenarios, representing different social settings (i.e., alone, dedicated internet access vs. together with others, notion of sharing internet access with others), were investigated in the test. QoE itself was evaluated in terms of traditional QoE-measures, complemented by a set of alternative, affective-state related measures. The latter were included to link up to the new definition of QoE as '*a degree of delight or annoyance ...*' proposed in [6].

The rest of this paper is organized as follows: section 2 discusses the concepts of adaptive video streaming and Quality of Experience in more detail and give a brief overview of relevant related work. Thereupon, Section 3 discusses the details of the experimental study set-up (i.e., sample description, test procedure, network, service and device configuration, test material and test environment, the usage scenarios and included subjective measures). Section 4 is dedicated to the analysis and discusses the results of the study. Finally, Section 5 shares our conclusions and suggestions for future research.

## 2. Related work

Whereas QoE has been defined by ITU [7] as the '*overall acceptability of an application or service, as perceived subjectively by the end-user*', this dominant definition has been criticized over the last years due to its narrow interpretation (see [8] for an overview). In 2012, a new definition of QoE as '*the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state*' [6] was proposed. The goal is to bring users in the 'zone of delight' and to prevent/reduce user annoyance. In addition, a range of human-, system- and context-related factors that may bear an influence on QoE, are mentioned [6], implying that QoE evaluation requires an interdisciplinary approach (as is the case in our work). This more holistic approach to QoE was developed as part of a bottom-up and community-driven process in the COST Qualinet project on QoE. However, an important challenge is still to translate this broader conceptual understanding of QoE into adequate and extended measurement approaches, which also consider alternative measures of QoE (as indicators of 'delight')

or ‘annoyance’) and which draw on interdisciplinary collaboration. Although some recent studies have taken some first steps to confront the traditional QoE measures with alternative self-report measures (see e.g., [9]), this is still largely unexplored territory. By including both traditional and alternative self-report measures of QoE (see Section 3), the study presented in this paper aims to make a relevant contribution in this respect.

Since DASH solutions are based on the principle of adjusting the quality level on a continuous basis during delivery [4], they imply a need to develop adjusted, reliable QoE evaluation methods, as is argued in [2]. In their work, an overview is given of a set of ‘QoE metrics’ for DASH. These include e.g., average throughput, initial playout delay and buffer level. Although a number of very pertinent integration challenges with respect to QoE optimization for DASH services are discussed and although a QoE evaluation approach is proposed, an important limitation of this work is that QoE issues are not tackled from a real user point of view (i.e., QoE as inherently subjective). QoE is not considered from a more holistic perspective, thus leaving a lot of questions unanswered. In [10], a novel approach to estimate QoE for adaptive HTTP/TCP video streaming is introduced. The proposed no-reference QoE estimation module differs from existing approaches: both video playout interruptions - which may occur due to fluctuations in the bandwidth and retransmissions of packets - and the quantization parameter is considered. The latter indicates the video compression and is therefore directly linked to the bitrate and possible changes in the video quality [10]. This estimation approach is valuable, yet, it also approaches QoE from a very narrow, technical point of view and the output is a prediction of ‘isolated’ QoE in terms of MOS-values.

One of the few papers presenting results from two actual subjective video quality experiments in the context of HTTP adaptive streaming, is the work by Robinson et al. [11], which we will briefly discuss below. They have tested several commercially available solutions for HTTP adaptive video streaming, including Apple HTTP Live Streaming (HLS)[12], Microsoft Smooth Streaming [13] and Adobe Dynamic Streaming [14]. Although these solutions are quite similar in terms of how they operate, they are not interoperable. The main reasons for this are the differences in encoding and client/server message format. The ongoing standardization efforts [15,13] are addressing these issues. In addition to the differences found on the client/server interface, there are also differences in internal functionality on the client side, which impacts the behavior of each solution. As a result, the algorithms controlling when to change quality level, and how often these changes are allowed are different between the solutions. The latter is either decided by video segment duration- or volume. Whereas the segment duration approach is applied in the commercial available solutions from Microsoft, Apple and Adobe, the segment volume based approach is e.g. used in the closed solution used by Netflix [16]. As is discussed in more detail in section 3.3, we used Apple HLS in our study.

The first study of Robinson et al. [11], in which these commercial solutions were used, took place in a controlled lab environment and investigated the impact of different network-related impairments on overall video quality (MOS scores). The results indicate high scores for the three commercial solutions when no impairments are introduced, yet point to a number of differences – which can be linked to the distinct characteristics of the different solutions – when considering the impact of the investigated impairments (namely bandwidth limitations, latency, random and burst packet loss). Moreover, relatively high MOS scores were found up to 2 Mb/s. For the second study, another approach was followed: members from the Alcatel-Lucent youth lab were invited to participate in an online study in which clips containing typical yet

exaggerated impairments in HAS were evaluated in terms of overall acceptability (transformed into MOS scores). The results indicated amongst others, that a constant (or approximately constant) bit rate is preferred over a bit rate that is frequently changing, even when this implies that the average quality is lower [11].

Both the Robinson et al. studies and the work by Singh et al. [10] are valuable contributions. However, QoE is only evaluated in terms of overall video quality perception. In our own study, we aim to extend this traditional approach, by linking up more explicitly to the new definition of QoE and by introducing alternative measures to evaluate QoE for adaptive video streaming. In the next section, we discuss the setup of the study in detail.

### **3. Experimental study setup**

As mentioned, the objective of our study is twofold. First of all, we aim to investigate QoE in the context of different adaptive video streaming conditions. To this end, three different bitrate profiles, representing realistic bandwidth conditions (i.e., constant high, constant low and changing), were used in the experiment. A second objective is to explore the possible influence of specific contextual variables on QoE. More concretely, two usage scenarios were considered in the test representing different social settings (i.e., alone, dedicated internet access vs. together with others, notion of sharing internet access with others).

#### **3.1 Sample description**

Based on the widely used convenience sampling method and following the ITU guidelines concerning the number of observers for subjective quality assessments (i.e., between 15 and 40), 32 test subjects were recruited. The main disadvantage of this type of sampling lies in the fact that the results cannot be generalized to the general population. However, as this is not the aim of our study, the use of this sampling method is justified. 59.4% male and 40.6% female test subjects, participated in the main study (in groups of four). The average age of the participants was 22 (Standard Deviation (S.D.)= 4.6). The youngest participant was 17, the oldest 33. All test subjects had normal or corrected-to-normal vision and reported normal hearing. Whereas a limited number of participants were working or studying in the field of audio/video quality, multimedia processing or a related field, the majority (81.3%) were naive users. Next to basic socio-demographical data, we also collected some additional information about the participants. In particular, we were interested in their use or non-use of tablets in their natural environment. 40.6% owns a tablet and one out of three participants used a tablet on a daily basis or even several times per day during the month before the test (which was used as reference period in the questionnaire). Moreover, 68.8% of the participants watched video content on a tablet during the month preceding the study. In order to link up to the new QoE definition, which also emphasizes the importance of a user's 'current state', we also included a pictorial measure of 'mood' (the validated 'Pick-a-Mood'-scale) in the pre-questionnaire [17]. Whereas 4 respondents (12.5%) evaluated their mood as 'neutral', 81.3% of the test subjects indicated to be in a pleasant state. Two participants labeled their mood as unpleasant (respectively 'tense' and 'bored').

## 3.2 Test procedure

The test procedure consisted of 3 parts and the entire test took around 1 hour. The main experiment was preceded by a pilot test, which was used for optimizing the procedure, the instructions and questionnaires. The procedure was as follows:

In the first part, the participants were given a brief, general introduction to the study (as in [11], without revealing concrete aims or details) and asked to fill in the pre-questionnaire. Thereupon, they were given instructions with respect to the viewing procedure and the operating of the iPad video player. The participants also received a summary of the instructions (with screenshots) on paper. Then, the self-report measures and different scales used in the during-questionnaires (see section 4.1) were explained. Finally, to familiarize the participants with the quality range and evaluation scale, they were shown examples of the worst and the best quality used in the test. Before moving to the next part, the test subjects could ask questions in case something was not clear.

After the introduction and briefing phase, the actual test started (part 2), with the first usage scenario. The participants were informed that they would watch three online music videos in the current setting, namely a situation where they are sitting alone at a desk and have their own, dedicated internet access point next to them. After the watching of a clip, the during-test questionnaire was filled in (considering the entire clip; so one rating per measure per clip) and this procedure was repeated 3 times.

Thereupon, the participants were asked to move to the second usage setting, reflecting a situation in which a user is alone, but watching in the presence of other people and sharing the internet access. Before the next clip viewing round started, the test leader pointed to the different situation (together with other people and sharing the internet access). Again, they watched the three clips, and filled in the during-questionnaire after every clip (considering the entire clip).

## 3.3 Devices and configuration of network and service

We used 9.7" iPads (4<sup>th</sup> gen) as terminal devices in this study. This choice was made based on recent figures for the worldwide tablet market [18], which indicate that Apple is still the number one vendor in terms of market share. The participants were asked to respect an indicated viewing distance of 50 cm. During the playback of the videos, the participants were wearing Bose AE2 headphones, which were set to a medium, comfortable level. The participants were able to change the volume in case needed.

The use of iPads in the experimental part of our work made it necessary to use Apple HLS. The HLS solution is developed for the purpose of streaming audio and/or video from web servers to iOS based clients such as iPhone and iPad. The recommended segment duration value for HLS provided content is 10sec which then gives the highest possible frequency of quality adjustments. A change in quality level is requested by the client whenever it observes changes in available bandwidth, which makes it desirable to either go up or down in quality level. The details of the controlling algorithm on the client side are not disclosed by Apple. In order to make the experiments as realistic as possible, we established a complete network and service platform (cf. Fig 1). The alternative would have been to use locally stored content on each iPad with impairments injected through the encoding process.

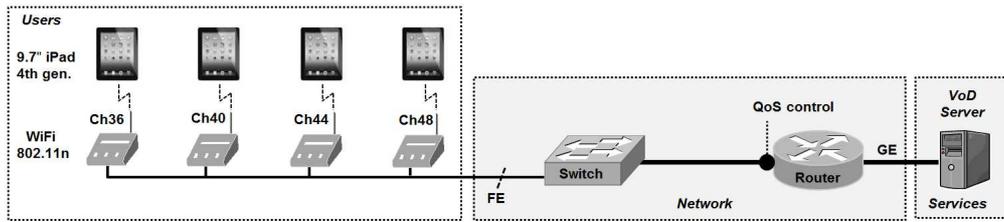


Figure 1 Lab network

Each of the iPad's are connected to the lab network through dedicated 802.11n access points, each of which was configured to use non-overlapping channels (36/40/44/48) to avoid interference. In order to provide QoS control for the video streams during the experiments, bitrate-limiting functions provided by the Cisco router in the network part were used. The QoS control was done individually for each iPad. The role of the QoS control function in the experiments was to enforce the delivery of video streams according to the configured profiles as illustrated in Fig. 2 on the left side. The right side of this figure shows how the actual streams behaved when being controlled by these profiles. As expected, there are some slight variations between the sessions. This is due to the natural lack of synchronization of the video start operation on each tablet. The initial behavior of any adaptive video streams is also visible, as it always starts out low and climbs until it reaches the maximum level possible given by the stream definition and current bandwidth constraint.

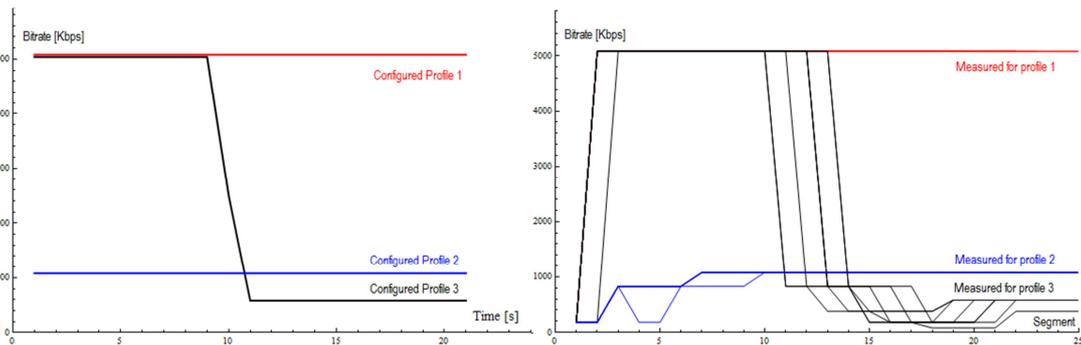


Figure 2 Configured (left) and measured (right) bitrate profiles for video streams

The three bitrate profiles represent different available bandwidth scenarios. The first profile is the case where the video stream is allowed to operate at maximum bitrate level (5Mbps). For the second profile the bandwidth is constrained so that the client decides to request a lower bitrate (1Mbps). In the third profile, a bandwidth constraint is introduced half way into the music video, which makes the client request a bitrate level from the lowest region (600Kbps). In none of the above profiles, the audio quality was affected.

### 3.4 Test material and test environment

The tests took place at the university campus, in a dedicated test room, with semi-artificial lighting. As test material, we used 3 recent music video clips of female singers, with duration of about 210sec. These clips were selected from a larger content pool, based on their relatively similar spatial and temporal information. Each clip was available in 14 different quality levels – encoded as MPEG-2 Constant Bitrate (CBR) transport streams. In Table 1 we present those levels actually used in the experiment (cf. Fig. 2, right side) in more detail. Depending on the spatial / temporal complexity of the video, the encoding process applies the appropriated degree of compression.

Table 1 Quality levels / bitrate profiles used in experiment

<b>Bitrate (Kbps)</b>	5000	1000	750	500	200	100
<b>Resolution (Pixels/Inch)</b>	95010	34204	23753	15202	5938	1518

The sequencing of the clips and order of the bandwidth limitations were presented in such a way that order or other presentation-related effects were avoided.

### 3.5 Usage scenarios

The creation of different Internet access- and social setting scenarios was done by running the experiments in groups of four users, as was already mentioned above. The first usage scenarios we created for the group members was the case where we let them sit alone and with their own access point visible on their desk (see Fig. 3, left). When placed in this scenario, information was given stating that they were now accessing the Internet through a dedicated access.

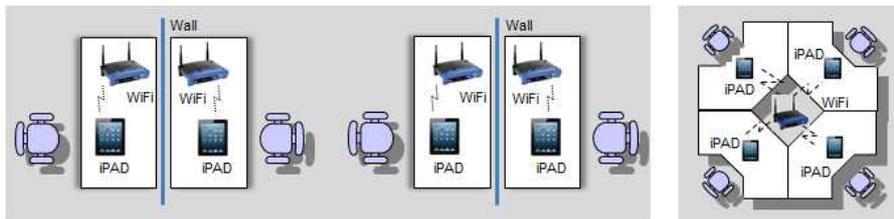


Figure 3 (left) Users separated with dedicated access; (right) Users together with “shared” access

In the second scenario, the idea is to create a notion among the users that they are part of a group which is sharing the access to Internet. This scenario was created by letting the group members sit at the same table where they could see each other, and with a single access point on the desk (Fig 3, right). When placed in this scenario, information was given stating that they were now accessing the Internet through a shared access. The access point placed on the shared desk in this scenario was only powered on, but did not provide service to the iPads. Put differently, whereas the participants were given the impression that they were sharing the bandwidth, each iPad was still connected to the lab network through the dedicated access points.

### 3.6 Subjective measures

Before turning to the results, Table 2 overviews the subjective measures included in the questionnaires after each clip, relevant for this paper.

Table 2 Overview of used subjective measures

<i>Overall video quality</i>	ITU five-point Absolute Category Rating (ACR) scale (due to content length, a double stimulus presentation design was not appropriate)
<i>Acceptability</i>	binary scale (acceptable / not acceptable)
<i>Degree of delight and degree of annoyance</i>	5-point scale with labeled end points, ranging from 0 ('not at all', e.g., feeling not annoyed or not happy at all) to 4 ('extremely', e.g., feeling extremely annoyed or happy). Multiple item constructs, so internal consistency reliability was checked by means of Cronbach's alpha (threshold of .65). New variables were computed for delight ( $\alpha = .89$ , 3 items) and annoyance ( $\alpha = .87$ , 2 items)
<i>Pleasure and arousal</i>	Nine-point SAM-scales, ranging from extremely sad to extremely happy (pleasure dimension) and from extremely calm/bored to extremely excited (arousal dimension). Self-Assessment Manikin (SAM) is a pictorial technique for assessing specific emotions in relation to a stimulus
<i>Expectations</i>	Degree to which the overall video quality was meeting the participant's expectations: measured on a 5-point scale ranging from -2 ('much worse than I expected') to 2 ('Much better than I expected')
<i>Perceived video quality stability</i>	5-point scale, based on two items ( $\alpha = .88$ )

## 4. Results

### 4.1 QoE evaluation

We first of all take a look at the figures for the ‘traditional’ QoE measures included in the study, namely the overall quality ratings and acceptability evaluations. As the descriptives (Average, Avg. and Standard Deviation, S.D.) in Table 3 indicate, the overall quality of the videos is on average evaluated as best in the constant high bitrate profile and as worst in the profile in which the bitrate changed from high to low. Moreover, this is the case in both usage scenarios (Avg. = 4.44 vs. 3.00 in scenario 1, Avg. = 4.34 vs. 2.78 in scenario 2).

Table 3: Descriptives (Avg. & S.D.) of subjective measures, per scenario and bitrate profile

Usage scenario	Scenario 1 (alone, dedicated access)						Scenario 2 (together, shared access)					
Bitrate profile	high		changing		low		high		changing		low	
	Avg.	S.D.	Avg.	S.D.	Avg.	S.D.	Avg.	S.D.	Avg.	S.D.	Avg.	S.D.
<b>Overall quality</b>	4.44	0.88	3.00	0.95	3.75	0.67	4.34	0.60	2.78	0.83	3.59	0.80
<b>Delight</b>	2.46	0.90	1.51	0.79	2.00	0.96	1.74	0.95	1.41	0.81	1.61	0.96
<b>Annoyance</b>	0.22	0.44	0.92	0.88	0.38	0.79	0.45	0.77	1.23	1.03	0.58	0.64
<b>Arousal</b>	7.09	1.20	5.84	1.27	6.56	1.48	6.25	1.41	5.22	1.43	5.84	1.48
<b>Pleasure</b>	4.47	1.93	3.28	1.40	3.69	2.04	3.72	2.00	3.00	1.67	3.03	1.71
<b>Expectations</b>	0.75	0.95	-0.47	0.92	-0.13	0.83	0.59	0.76	-0.69	0.82	-0.06	0.56
<b>Video quality stability</b>	3.23	0.93	0.63	0.68	1.83	1.29	3.33	0.71	0.64	0.64	2.03	0.84

Figures 4a and 4b show absolute numbers and thus give a more nuanced overview of the distribution of the overall quality ratings. As can be observed, in both usage scenarios, the overall quality of the videos shown with high constant bitrate is clearly evaluated as better than the videos shown with a low, but constant bitrate and those with a changing bitrate. Thus, although the changing profile implies a bitrate that is on average higher, our findings indicate that a lower but constant bitrate scores higher in terms of overall quality. To investigate whether these differences depending on the bitrate profile are significant, non-parametric tests were used. For both usage scenarios, we conducted Friedman’s ANOVA with the overall quality ratings as dependent and the different bitrate profiles as independents. Based on these tests, we have no evidence to state that the differences between all three bitrate profiles in terms of overall quality, as shown in figures 4a and 4b, are significant from a statistical point of view.

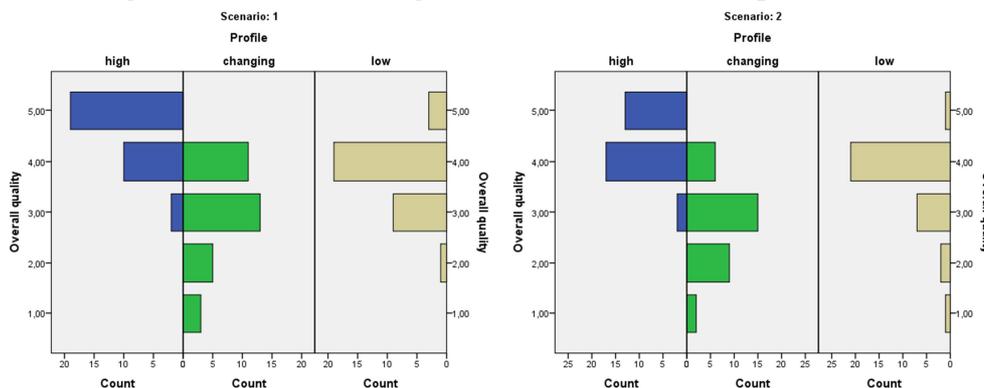


Figure 4a and 4b: absolute ratings for overall quality, for the bitrate profiles and usage scenarios

Purely by looking at the descriptives and figures, it can also be argued that the overall quality ratings for the different bitrate profiles are – in general - relatively high. This is illustrated in the boxplot for the overall quality ratings (Fig 5a): even for the

changing bitrate profile, the median - as indicator of the middle tendency – is still relatively high (namely 3, indicating a ‘fair’ overall quality).

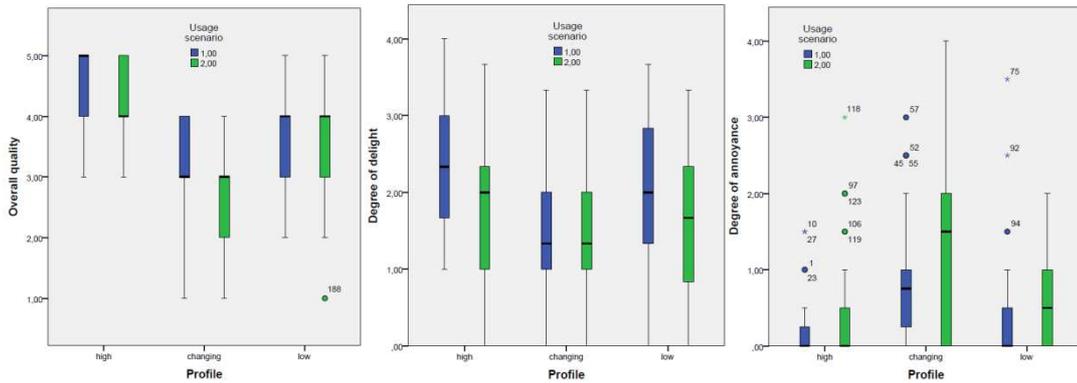


Figure 5a, 5b, 5c Boxplots for overall quality, degree of delight and degree of annoyance

When considering the acceptability of the overall video quality (i.e., whether or not it is good or good enough to watch the video) for the different bitrate profiles and scenarios, we found that only 7.3% of all videos were evaluated as ‘not acceptable’ in the scenario where participants were watching alone and had a dedicated network access (scenario 1). In the second usage scenario, this percentage is a bit higher, but still relatively low (14.6%). These figures indicate that – in both settings – the majority of the videos were evaluated as acceptable to watch, despite the fact that the test subjects noticed some interruptions and assessed the overall video quality differently according to the bitrate profile. Based on these observations, it can be argued that the strong fixation on the evaluations of overall video quality in terms of MOS scores as only measure of QoE, needs to be nuanced and broadened. It is however interesting to further explore which videos were considered as not-acceptable. In this respect, the acceptability ratings support the findings that were discussed above: the percentage of videos that are rated as not-acceptable, is highest within the changing bitrate profile (i.e., 71.4% in scenario 1 and 78.6% in scenario 2). We used Pearson’s chi-square test to investigate the relation between the evaluated acceptability and the three bitrate profiles and found that there is a significant correlation ( $\chi^2(1)= 19.489, p < 0.001$ ) between the bitrate profile (changing vs. constant) and the acceptability of the overall video quality. This seems to represent the fact that, based on the odds ratio, the odds that a video was evaluated as acceptable were 8.1 times higher if the bitrate profile was constant instead of changing.

As was discussed above, we also included a set of ‘alternative’ measures of QoE in our study. The descriptives in table 2 allow us to share some general observations, before further investigating the differences across the different bitrate profiles. On average, the degree of delight (see Fig. 5b), the self-reported arousal and pleasure and the evaluated video quality stability were highest in the constant high bitrate profile and lowest in the changing bitrate profile. Moreover, this is the case in both usage scenarios. Similarly, on average, the overall quality of the clips shown in the high bitrate profile is slightly better than expected, whereas in the changing and low but constant bitrate profiles, the overall quality is evaluated as worse to slightly worse than expected, as is also illustrated in Fig. 6.

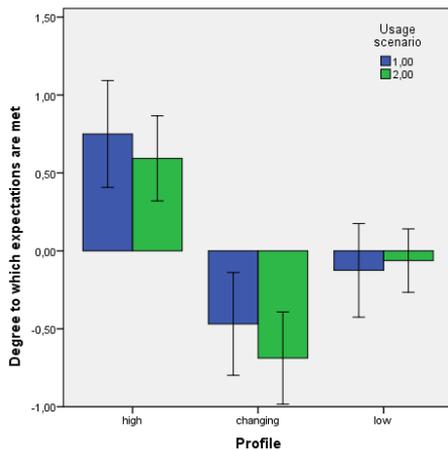


Figure 6 Average expectations (95% C.I.)

Table 4 Friedman's ANOVA results

	Usage scenario 1	Usage scenario 2
	<i>sig.</i>	<i>sig.</i>
Degree of delight	N.S.	N.S.
Degree of annoyance	N.S.	N.S.
Pleasure	< .05	N.S.
Arousal	.05	.05
Expectations	N.S.	N.S.
video quality stability	N.S.	N.S.

With respect to the degree of annoyance (see Fig. 5c), the opposite finding holds true: the annoyance is highest in changing bitrate profile and again, this applies for both scenarios. Again, to investigate whether these observed differences are substantial, Friedman's ANOVA tests were performed with the alternative measures as dependents and the different bitrate profiles as independents. The results, which are summarized in Table 4, point to significant differences between the three bitrate profiles, but only for reported Pleasure and Arousal. To further explore where the differences are situated, the post hoc procedure consisted of the conducting of additional separate Wilcoxon signed rank-tests. For the *first usage scenario*, the differences between the constant high bitrate profile and the changing bitrate profiles for the self-reported Pleasure and Arousal are substantial. The results also indicate a significant difference between the changing and constant low bitrate profile in terms of Pleasure. The reported degree of delight and video quality stability are higher in the constant low bitrate profile than in the changing profile, but we could not find statistical evidence supporting this observation. In the *second usage scenario* the arousal is significantly higher in the constant low bitrate profile than in the constant high bitrate profile.

#### 4.2 Does social setting and internet access sharing influence QoE ?

As can be observed in Table 3, the subjective measures consistently point to a better QoE (e.g., better overall quality evaluation, higher delight, lower annoyance) in scenario 1. Additional analyses were performed to test whether there are significant differences between usage scenario 1 and 2 in terms of QoE (traditional and alternative measures) per profile. As we aim to compare two related, we used Wilcoxon signed-rank test, which can be considered as a non-parametric alternative for the dependent t-test. For the constant high bitrate profile, we found no significant differences in terms of QoE (using both the traditional and alternative measures) between the first and the second usage scenario. Or put differently, in this study, the presence of others and the notion of dedicated vs. shared access did not have a substantial influence on the QoE.

When investigating the differences in scores between the two scenarios for respectively the changing and the constant low bitrate profile however, we found significant differences ( $p < .05$ ) in terms of the self-reported pleasure, arousal and the degree of delight, which were all significantly higher in the first usage scenario. This means that the differences between usage scenario 1 and 2 for the changing and the constant low bitrate profile, are substantial. Finally, the acceptability ratings (compared

using Pearson's chi square test) did not significantly differ between the two usage scenarios, meaning that the presence of others and notion of shared internet access – as operationalized in our study - did not have a significant impact on the participants acceptability evaluations (they were more explicitly and significantly associated with the different bitrate profiles, as was discussed above).

Follow-up research with additional test subjects is needed to validate these findings and observations, as well as to fully understand their implications for QoE-based optimization and differentiation strategies. The role of possible influencing factors (e.g., fatigue and boredom, test setup, content) also requires further investigation.

## **5. Conclusions and future work**

In this paper, we have shared findings from an experimental user study on QoE in the context of adaptive video streaming. The objectives of the study presented in this paper were twofold: we first of all aimed to investigate QoE in the context of different realistic adaptive video streaming profiles. The video streams were delivered according to three different bitrate profiles, representing realistic bandwidth conditions (i.e., constant high bitrate, constant low bitrate, changing bitrate resulting in adaptive video quality). In order to also contribute to the ongoing discussions on the measurement of QoE, both traditional and more alternative measures of QoE (linking up to [5]) were included. Secondly, we wanted to explore the possible influence of specific contextual variables on QoE. More concretely, the test took place in two usage scenarios representing different social settings (i.e., alone, dedicated internet access vs. together with others, notion of sharing internet access with others).

Although our findings are valid for the study presented here and cannot simply be generalized, they point to differences between the three bitrate profiles in terms of QoE (both considering the traditional and the alternative QoE measures). We found indications that a lower, constant bitrate is preferred over the changing bitrate which is - on average - higher. The latter is associated with lower delight, higher annoyance and higher odds to be evaluated as 'not acceptable'. This finding is in line with previous studies such as [11] and confirms that a constant, fluent video playout is important and that variations in video playout are annoying. This is the case for both separate usage scenarios. However, based on our findings from this first study, we cannot claim that these differences are significant from a statistical point of view. Interestingly, the majority of the videos were evaluated as acceptable to watch, and indicate a higher impairment tolerance than when only considering the overall quality ratings. Although we found no significant influence of the social setting when considering traditional QoE evaluation measures, significant differences between both settings were detected when evaluating QoE in terms of delight and annoyance. Moreover, it could be observed that QoE was consistently better in scenario 1.

This study was only a first step and further research is needed to better understand user preferences in the context of adaptive video streaming. In future work, we aim to further zoom in on the possible influence of the social setting (through different setups, e.g., in naturalistic settings with higher ecological validity). Additionally, we aim to take a closer look at the interplay between content likeability, mood and QoE; and to investigate how traditional and alternative measures of QoE relate to each other. In our future work, we aim to also consider other adaptive streaming solutions and terminal devices, additional bitrate profiles (e.g., from low to high bitrate), alternative test setups (including 'fake streaming', with locally stored content and encoded impairments) as well as other types of content.

## 6. References

1. Cisco, Inc. (2011) Cisco Visual Networking Index: Forecast and Methodology, 2010-2015. [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-481360\\_ns827\\_Networking\\_Solutions\\_White\\_Paper.html](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html)
2. Oyman O, Singh S (2012) Quality of experience for HTTP adaptive streaming services. *Communications Magazine*, IEEE 50 (4):20-27. doi:10.1109/mcom.2012.6178830
3. Siglin T (2013) The State of OTT Content Delivery 2013. *Streaming Media Magazine* (2)
4. Sodagar I (2011) The MPEG-DASH Standard for Multimedia Streaming Over the Internet. *Multimedia*, IEEE 18 (4):62-67. doi:10.1109/mmul.2011.71
5. DASH Industry Forum, inc. (2013, May) Survey of European Broadcasters on MPEG-DASH Whitepaper <http://dashif.org/w/2013/05/Survey-of-the-European-Broadcasters-on-MPEG-DASH-Whitepaper-V2.1.pdf>. Accessed June 16, 2013
6. Le Callet P, Möller S, Perkins A (eds) (2012) Qualinet White Paper on Definitions of Quality of Experience, version 1.1, June 3, 2012. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne
7. ITU (2008) ITU-T Rec. P.10 Vocabulary for performance and quality of service, Amendment 2: New definitions for inclusion in Recommendation ITU-T P.10/G.100. International Telecommunication Union, Geneva
8. De Moor K (2012) Are engineers from Mars and users from Venus? Bridging gaps in Quality of Experience research: reflections on and experiences from an interdisciplinary journey. Unpublished doctoral dissertation, Ghent University: Department of Communication Studies,
9. De Moor K, Rios Quintero M, Strohmeier D, Raake A (2013) Evaluating QoE by means of traditional and alternative subjective measures: an exploratory 'living room lab' study on IPTV. Paper presented at the 4th International Workshop on Perceptual Quality of Systems (PQS 2013) Vienna, Austria,
10. Singh KD, Hadjadj-Aoul Y, Rubino G Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC. In: Consumer Communications and Networking Conference (CCNC), 2012 IEEE, 14-17 Jan. 2012. pp 127-131. doi:10.1109/ccnc.2012.6181070
11. Robinson DC, Jutras Y, Craciun V (2012) Subjective Video Quality Assessment of HTTP Adaptive Streaming Technologies. *Bell Labs Technical Journal* 16 (4):5-23. doi:10.1002/bltj.20531
12. Microsoft, inc. (January 2013) IIS Smooth Streaming Transport Protocol, [MS-SSSTR] v20130118. <http://msdn.microsoft.com/en-us/library/ff469518.aspx>.
13. Pantos M (2013, April) Apple HLS (HTTP Live Streaming). IETF draft. <http://tools.ietf.org/html/draft-pantos-http-live-streaming-11>.
14. Adobe systems, inc. HTTP Dynamic Streaming <http://www.adobe.com/products/httpdynamicstreaming/>.
15. ISO, inc. (2012, April) ISO MPEG-DASH Specification (ISO/IEC DIS 23009-1:2012) <http://dashif.org/mpeg-dash/>.
16. Akhshabi S, Begen AC, Dovrolis C An experimental evaluation of rate-adaptation algorithms in adaptive streaming over http. In: *ACM Multimedia Systems (MMSys)*, 2011. ACM, pp 157-168
17. Desmet PMA, Vastenburger MH, Van Bel D, Romero NA Pick-A-Mood; development and application of a pictorial mood-reporting instrument. In: Brassett J, Hekkert P, Ludden G, Malpass M, McDonnell J (eds) 8th International Design and Emotion Conference, Central Saint Martin College of Art & Design (11-14 September 2012), London 2012.
18. IDC Worldwide Tablet Tracker, inc. (2013, May 1) Worldwide Tablet Market Surges Ahead on Strong First Quarter Sales - press release. <http://www.idc.com/getdoc.jsp?containerId=prUS24093213> Accessed June 16 2013