Modeling Subjective Quality Evaluations for Mobile Video Watching in a Living Lab Context

Toon De Pessemier, Luc Martens and Wout Joseph
Ghent University / iMinds, Dept. of Information Technology, Ghent, Belgium
{Toon.DePessemier, Wout.Joseph, Luc.Martens}@intec.ugent.be

Abstract—**The ability to understand and quantify the Quality of Experience (QoE, i.e., the users' subjective perception of an application or service) will play a major role in the success of mobile services in the next years. This paper discusses results from an exploratory study in which QoE aspects related to mobile video watching were investigated in a living lab setting. More specifically, we zoom in on usage patterns in a natural research context and on the subjective evaluations regarding different aspects of the mobile video session. Multinomial logistic regression results in a model to estimate the subjective evaluations regarding the perceived distortion and loading speed based on objectively-measured parameters of the video session.**

I. INTRODUCTION

In contrast to traditional Quality of Service (QoS) approaches, which are usually driven by multimedia signal degradation from the signal quality point of view, Quality of Experience (QoE) considers how users perceive and experience multimedia content or services as a whole [1]. Different definitions of QoE exist, but all have similar notion, referring to user satisfaction [2]. By the ITU-T, QoE is defined as "the overall acceptability of an application or service, as perceived by the end-user", which might be influenced by 'user expectations' and 'context' [3]. Since QoE reflects the user-perceived experience rather than the implied impact of QoS, it is considered as a more important metric than QoS [4].

The most reliable way of assessing and measuring the perceived quality of video is conducting subjective quality assessments, in which human observers evaluate a series of video sequences [5]. Through subjective quality assessments, the QoE of mobile video services can be related to objectively-measured parameters [6]. This paper discusses results from an exploratory study in which QoE aspects related to mobile video-on-demand were investigated in a living lab setting. Via this study, the effect of the loading time and packet-loss rate on the subjective evaluation behavior is evaluated by means of a panel of test subjects.

The perceptual effects of packet loss during video streaming are depending on various technical parameters of the video, such as the codec or the type of frame in which packet loss occurs (I, P, or B-frame), as well as on network characteristics such as burst patterns. For MPEG encoded videos for example, packet-loss rates as low as 3% might induce much higher frame error rates (up to 30% of frame errors) and clearly-noticeable distortions for video with an approximate duration of five mintues [7]. Compared to a perfect video, perceptual quality drops by over 50% in the presence of a packet-loss

rate of about 8%. Additional packet loss, up to 22%, induces a further (but weaker) decline in the perceptual quality [8]. In terms of loading time, i.e. the waiting time between selecting a video and the start of the video playback, a loading time of 1.5 seconds is still considered as reasonable for real video applications [9]. Whereas these studies discussed the influence of the loading time and packet loss on the perceptual quality in the context of video watching on desktop computers or television sets (i.e. a fixed setting), this paper is the first to investigate this influence in a mobile context, which can have a serious impact on the QoE.

II. TEST SET-UP

A. Architecture

Figure 1 shows the architecture of the video delivery system used in this living lab experiment, consisting of the client device (i.e. a smartphone running the video player), the video server offering the content, and the technical database storing the objective parameters and subjective evaluations. For this living lab experiment, test subjects could use WiFi networks as well as the commercial cellular data network of Proximus, a Belgian network operator. Through the Proximus network, the client device is connected to a GPRS (General Packet Radio Service), EDGE (Enhanced Data rates for GSM Evolution), UMTS (Universal Mobile Telecommunications System), or HSPA (High Speed Packet Access) network depending on the location of the test subject. Besides, test subjects can opt to connect their device to a WiFi network that is available for them, e.g., at their work or at their home.

For coding the videos, 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 [10]. The AAC LC3 (Advanced Audio Coding, Low Complexity profile 3) compression scheme is used for the audio track.

B. User study

Over a period of one week, 29 test subjects were asked to watch movie trailers on a mobile device in their everyday environment, when and where they wanted (i.e. in a living lab setting), but within a time-span of 1 week (weekend included). The selected trailers, all between 2 and 3 minutes, covered different movie genres (such as comedy, drama, animation, action, and romance) and consisted of recent titles. Figure 2(a) shows a screenshot of the mobile application illustrating the
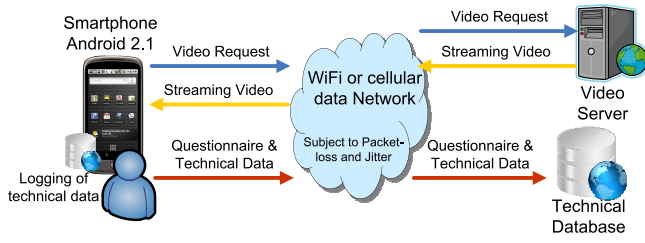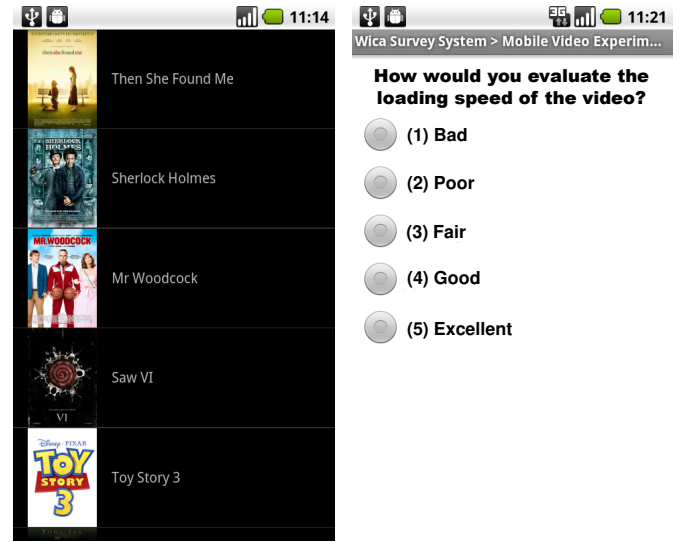
Fig. 1.   The architecture of the video delivery system.

(a) The video selection mechanism   (b) The questionnaire that is shown after each video

Fig. 2.   Screenshots of the video application on the mobile device

list of videos with a thumbnail and the title of the video as additional information. The viewers were able to decide themselves in which order they watched the clips. Every test subject was handed over a Google Nexus One mobile phone, running on Android 2.1 as operating system, to watch the videos.

Immediately after each video session, subjective evaluations were gathered by means of a questionnaire that pops up on the screen of the mobile device. Figure 2(b) shows a screenshot of this digital questionnaire that test subjects were asked to fill in. Test subjects had to assess various aspects of the video, such as the perceived distorting and the loading speed, using a 5-level subjective quality scale. More specifically, the test subjects were asked "How would you evaluate the loading speed of the video?" with a multiple choice answer ranging from 1 = Bad, to 5 = Excellent. The question regarding the perceived distortion was as follows: "Did you perceive visual distortions in the video during playback?" with a multiple choice answer ranging from 1 = Noticeable, very annoying, to 5= Not noticeable.

The questionnaire on the device was complemented with a traditional pen and paper diary consisting of some open questions. The goal of this paper diary was to give users the opportunity to provide more detailed and qualitative feedback regarding the video sessions and their experience. Since inputting text on mobile phones is difficult and tedious, mobile phones are not the optimal tool to gather detailed qualitative feedback. Therefore, we opted for this alternative feedback tool: a small paper diary that can also be used in case of technical problems with the device such as an application crash or a dead battery.

Before the actual test started for every test subject, instruction meetings were organized in groups of five test subjects. Because of the living lab environment, providing assistance during the experiment was more difficult than in the case of a controlled laboratory environment and therefore, special attention was paid to the briefing of the test subjects. After some general information on how to switch on/off, use, charge the device etc., it was explained how to access the test application and how to select and watch the videos. Because test subjects had to evaluate each video via a small questionnaire on the device, it was also shown how to navigate from one question to the next and fill in the questionnaire using the touch screen. At the end of the briefing session, every test subject was given a device, a diary, and an instruction leaflet with practical information, screenshots, and relevant instructions related to the grading scales and univocal interpretation of the questions.

During the video watching, relevant objective video and network parameters were logged, amongst others the video packet-loss rate (i.e. the percentage of video packets lost during transmission to the device) and the loading time of the video (which is measured as the time period between selecting a video and the moment when the video starts playing). These objective parameters are then associated to the subjective evaluations in order to investigate the influence of the technical aspects on the user's perceived quality of the video session.

Given the ordinal nature of the subjective evaluations, the data were analyzed by using multinomial logistic regression. Multinomial logistic regression is used to predict the probabilities of different possible outcomes of a dependent variable (in our case the subjective evaluation), given a set of independent variables which may be real-valued, binary-valued, categorical-valued, etc. (in our case the objective parameters) [11]. The result of multinomial logistic regression is a comparison of the probability of a specific event against the probability of a reference event. In Section III-B and III-C, multinomial logistic regression was used to model respectively the subjectively-perceived loading speed based on the measured objective loading time and the subjectively-perceived distortions based on the measured objective packet-loss rate during video playback.

III. RESULTS

A. Usage Patterns

First, we discuss the test subjects' usage patterns regarding video watching in a natural context. The logged data indicate that the majority of the videos were watched at home (82.7%) or at work (9.7%). Although one might expect that many videos would be watched during travelling, this was not the

case; only 5.2% was watched on the go and 2.4% somewhere else (including at a friend, in a café, or in a museum). In fact, only 8 of the 29 test subjects (i.e. 27% of the test subjects) watched videos during travelling. This confirms the results of a living lab study on mobile TV broadcasting, which showed that most viewing occurred at home [12].

In most cases (61.4%), no other people were around during the watching sessions (radius of approximately 5 meter). 22.8% of the videos were watched by the test subjects in presence of one other person. The presence of other people did not have a significant influence on the overall experience rating and was in 89.8% of the cases not perceived as a disturbing factor. In the remaining 10.2%, the talking of the others and noise made by them or coming from other sources (such as the TV) is often mentioned as disturbing factor.

Both on week days and weekend days, the evening (from 18.00 till 24.00 o'clock) was the most popular watching time (45.4% of the video sessions), followed by the afternoon (from 12.00 till 18.00 o'clock) (25.1% of the video sessions). In absolute numbers, most videos were watched during the week (72.8%), which makes sense since every test subject had one week to finish the test so only two weekend days, but five week days were included in the test period. So test subjects were about equally active during weekend days as during the week days.

B. Loading Speed

One of the quality aspects that test subjects could evaluate was the loading speed of the video. Table I shows the rating options for evaluating the perceived loading speed, the mean of the measured loading time corresponding to each option for the subjective evaluation (i.e. the mean loading time of the videos that received a specific rating), the number of video sessions that received a specific rating, and the fraction of the video sessions that received a specific rating. E.g., 125 videos or 16.6% were rated as excellent in terms of loading speed, and these videos had a mean loading time of 2.9 seconds. The results indicate that the loading speed of the majority of the video sessions (62.4%) is evaluated as 'good' or even 'excellent'. Conversely, for a considerable part of the video sessions (15.4%), the subjectively-perceived loading speed is 'poor' or 'bad'. Therefore, the influence of the objectively-measured loading time on the subjective evaluation of the perceived loading speed is investigated. Besides the loading time, the duration of the video might also influence the subjective evaluation of the loading speed. But since all videos of the experiment had approximately the same duration, this parameter is not included in the analysis.

To model these data by using multinomial logistic regression, the subjective evaluations of the loading speed are selected as dependent, the objectively-measured loading time is an independent (covariate), and the reference event is the evaluation of the loading speed as 'fair'. So for each rating option, the regression model provides a function for the ratio of the probability of obtaining that specific rating, e.g., P(Excellent) and the probability of obtaining the reference

Evaluation of the loading speed	Mean loading time (s)	Number of sessions	Fraction of the sessions
1 = Bad	29.3	59	7.9%
2 = Poor	18.7	56	7.5%
3 = Fair	5.7	167	22.2%
4 = Good	3.5	344	45.8%
5 = Excellent	2.9	125	16.6%
Total	7.1	751	100%

TABLE I
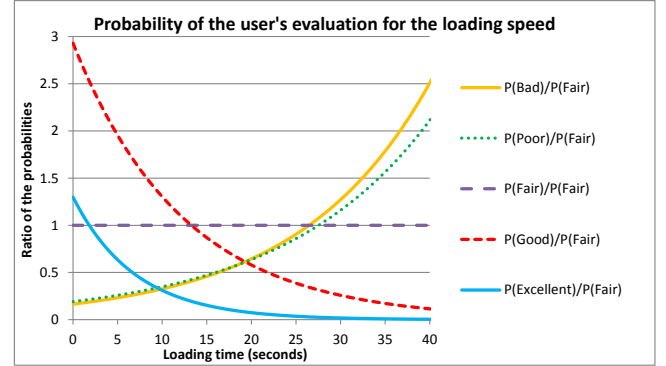SUBJECTIVE EVALUATIONS AND MEAN OBJECTIVE MEASUREMENT OF THE LOADING TIME.

Fig. 3. The probability ratios of the evaluation options for the perceived loading speed.

rating P(Fair), in terms of the objectively-measured loading time.

Figure 3 visualizes these probability ratios for an objectively-measured loading time ranging from 0 to 40 seconds. The graph shows that for short loading times (less than 10 seconds), a high probability exists that users will evaluate the loading speed as 'good' or 'excellent'. Given the large fraction of video sessions evaluated as 'good' (45.8% in Table I), the probability of obtaining 'good' as subjective evaluation is higher than the probability of obtaining 'excellent'. If the measured loading time is more than 13 seconds, users are more willing to evaluate the loading speed as 'fair' than to evaluate it as 'good'. For short loading times, users are not inclined to give low evaluations such as 'bad' or 'poor'. In contrast, after a loading time of approximately 27 seconds, ratings with the label 'bad' or 'poor' are more likely than the reference rating, i.e., 'fair'. And for instance after 40 seconds of waiting time, it is 2.5 times more likely that users perceive the loading speed as 'bad' than that users perceive it as 'fair' (Figure 3).

This model has a likelihood ratio χ^2 of 164.7 with a p-value < 0.0001 and 4 degrees of freedom, which indicates that the model (of Figure 3) as a whole fits significantly better than a model without the loading time as predictor. (The χ^2 statistic is the difference in 2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model [11].)

C. Distortion

Video streaming might suffer from audiovisual distortions if packets are lost during transmission. Therefore, the influence of packet loss on the subjectively-perceived distortion during mobile video watching was investigated for the video sessions of the living lab experiment. Table II shows the rating options for evaluating the perceived distorting during video watching, the mean of the objectively-measured packet-loss rate corresponding to the subjective evaluation (i.e. the mean packet-loss rate of the videos that received a specific rating), the number of video sessions that received a specific rating, and the fraction of the video sessions that received a specific rating.

Table II shows that sessions which received a positive evaluation regarding the perceived distortion ('not noticeable' or 'noticeable, not annoying') are characterized by a low packet-loss rate (mean values of 0.8% and 0.4%). In contrast, low ratings for the perceived distortion ('noticeable, annoying' or 'noticeable, very annoying') are typically due to high packet-loss rates (mean values of respectively 18.9% and 32.5%). Therefore, the influence of this packet-loss rate on the subjectively-perceived distortion during mobile video watching is further investigated.

As already indicated in the introduction, the perceived distortion due to packet loss depends on various technical parameters of the video, such as the codec. So if e.g., another video codec is used, the results can slightly change and the analysis should be repeated. Moreover, different individual packet losses can have a different impact on the perceived distortion due to patterns of subsequent packet losses (bursts) or the type of frame in which packet loss occurs (I, P, or B-frame). However for this analysis, individual packet losses are not investigated in detail, but the effect of a substantial packet loss rate as a whole is investigated.

As in the analysis of the loading speed, the data regarding packet loss were analyzed using multinomial logistic regression resulting in a model to estimate the probability of obtaining a specific rating as a function of the packet-loss rate.

For this analysis, the subjective evaluation of the perceived distortion was selected as dependent, the objectively-measured packet-loss rate is an independent (covariate), and the reference event was the evaluation of the distortion as 'noticeable, slightly annoying'. Figure 4 visualizes the probability ratios of the rating options for a packet-loss rate ranging from 0% to 40% (using a logarithmic scale). Video sessions with a limited packet-loss rate have a higher probability to obtain a positive rating regarding the perceived distortion ('not noticeable' or 'noticeable, not annoying') than to receive the reference rating (i.e. 'noticeable, slightly annoying'). Around a packet-loss rate of 0.25%, the probability of a positive evaluation starts decreasing. When more than 2.6% of the packets are lost during transmission, the probability that users are slightly annoyed by distortions is higher than the probability that users do not notice these distortions (solid blue line versus dashed purple line in Figure 4). Negative evaluations for the perceived distortion are less likely than the reference rating
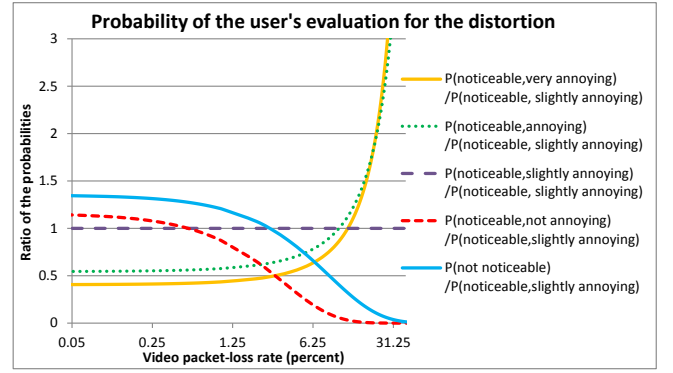

Fig. 4. The probability ratios of the evaluation options for the perceived distortion.

for low values of the packet-loss rate. Conversely, the rating options 'noticeable, annoying' and 'noticeable, very annoying' are more likely than the reference option 'noticeable, slightly annoying' as soon as the packet-loss rate is higher than respectively 10.5% and 12.5%.

The likelihood ratio χ^2 of 149.3 with a p-value < 0.0001 and 4 degrees of freedom tells us that our model as a whole fits significantly better than a model without the packet-loss rate as predictor.

Evaluation of the distortion	Mean packet-loss rate	Number of sessions	Fraction of the sessions
1 = Noticeable, very annoying	32.5%	71	19.1%
2 = Noticeable, annoying	18.9%	67	18.0%
3 = Noticeable, slightly annoying	3.1%	78	21.0%
4 = Noticeable, not annoying	0.4%	68	18.3%
5 = Not noticeable	0.8%	88	23.7%
Total	10.5%	372	100%

TABLE II
SUBJECTIVE EVALUATIONS OF THE DISTORTION AND MEAN OBJECTIVE MEASUREMENT OF THE PACKET-LOSS RATE

IV. CONCLUSION

In this exploratory study, we investigated Quality of Experience aspects related to mobile video watching in a living lab environment. The participants were able to watch the videos when they wanted, where they wanted and user evaluations were gathered by means of questionnaires on the device, complemented with traditional pen and paper diaries. Through this experiment, objectively-measured parameters of the video sessions and subjective evaluations of the test subjects were gathered.

The results illustrate that most videos were watched at home and in the afternoon and evening. In most cases, no other people were around during the watching session. The presence of other people did not have a significant influence on the overall experience rating and was in 90% of the cases not perceived as a disturbing factor.

Multinomial logistic regression results in a model to estimate the subjective evaluations regarding the perceived distortion and loading speed based on the objectively-measured parameters of the video session. Regarding the perceived loading speed, the model showed that when the loading time increases from 10 to 30 seconds, the subjective evaluations gradually evolve from mainly positive to mainly negative. In terms of subjectively-perceived distortion, the probability of receiving a positive evaluation is rapidly decreasing if packet loss occurs during video watching (from a packet-loss rate of around 0.25%) and video sessions with a packet-loss rate of more than 10% are in general evaluated as 'annoying' or even 'very annoying'. These results provide application developers and service providers a tool that clarifies how technical parameters influence the perceived quality of streaming video in a mobile environment.

REFERENCES

[1] U. Reiter, "Overall perceived audiovisual quality - what people pay attention to," in *Consumer Electronics (ISCE), 2011 IEEE 15th International Symposium on*, June 2011, pp. 513–517.

[2] D. Soldani, M. Li, and R. Cuny, *Front Matter*. John Wiley & Sons, Ltd, 2006, pp. i–xxvii. [Online]. Available: http://dx.doi.org/10.1002/9780470034057.fmatter

[3] "Definition of Quality of Experience (QoE)," ITU-T, International Telecommunication Union, Liaison Statement, 2007, ref.: TD 109 rev 2 (PLEN/12).

[4] 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. [Online]. Available: http://doi.acm.org/10.1145/1047936.1047938

[5] 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," *Broadcasting, IEEE Transactions on*, vol. 56, no. 4, pp. 458–466, December 2010.

[6] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching," *Broadcasting, IEEE Transactions on*, vol. 59, no. 1, pp. 47–61, 2013.

[7] J. M. Boyce and R. D. Gaglianello, "Packet loss effects on mpeg video sent over the public internet," in *Proceedings of the sixth ACM international conference on Multimedia*, ser. MULTIMEDIA '98. New York, NY, USA: ACM, 1998, pp. 181–190. [Online]. Available: http://doi.acm.org/10.1145/290747.290770

[8] M. Claypool and J. Tanner, "The effects of jitter on the perceptual quality of video," in *Proceedings of the seventh ACM international conference on Multimedia (Part 2)*, ser. MULTIMEDIA '99. New York, NY, USA: ACM, 1999, pp. 115–118. [Online]. Available: http://doi.acm.org/10.1145/319878.319909

[9] I.-M. Pao and M.-T. Sun, "Encoding stored video for streaming applications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, no. 2, pp. 199–209, 2001.

[10] "H.264 : Advanced video coding for generic audiovisual services," ITU-T, International Telecommunication Union, Tech. Rep., January 2012, online available at http://www.itu.int/rec/T-REC-H.264-201201-I/en.

[11] T. F. Liao, *Interpreting Probability Models, Logit, Probit, and Other Generalized Linear Models*, 1st ed. SAGE Publications, Newbury Park, 1994.

[12] D. Schuurman, K. De Moor, L. De Marez, and T. Evens, "A living lab research approach for mobile tv," *Telematics and Informatics*, vol. 28, no. 4, pp. 271 – 282, 2011, television in a digital era - Usage and policy issues. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0736585310000766