QoE assessment of multimedia video consumption on tablet devices

Conference Paper · December 2012

DOI: 10.1109/GLOCOMW.2012.6477775

4 authors:

Alessandro Floris
Università degli studi di Cagliari
48 PUBLICATIONS **793** CITATIONS

SEE PROFILE

Luigi Atzori
Università degli studi di Cagliari
283 PUBLICATIONS **22,843** CITATIONS

SEE PROFILE

Giaime Ginesu
Università degli studi di Cagliari
27 PUBLICATIONS **347** CITATIONS

SEE PROFILE

Daniele D. Giusto
Università degli studi di Cagliari
233 PUBLICATIONS **1,676** CITATIONS

SEE PROFILE

QoE Assessment of Multimedia Video Consumption on Tablet Devices

Alessandro Floris
Multimedia Communications Lab, CNIT
Pula (Cagliari), Italy
E-mail: alessandro.floris@diee.unica.it

Luigi Atzori, Giaime Ginesu, Daniele D. Giusto
DIEE, University of Cagliari
Cagliari, Italy
E-mail: {l.atzori, g.ginesu}@diee.unica.it, ddgiusto@unica.it

***Abstract*—in this paper we present a study aimed at assessing the Quality of Experience in video streaming when the user is employing a tablet devices. The study has been conducted on a database containing subjective assessment scores relative to 216 streaming sessions of video sequences encoded with H.264/AVC and corrupted by typical wireless channel transmission errors. Four 20sec-long video sequences at Common Intermediate Format (CIF) spatial resolution have been corrupted with 54 combinations of key parameters for the reference application: bitrate, packet loss rate, playout delay and transmission interruption . Subjective evaluations have been collected in compliance with ITU-T Recommendation P.910 through single-stimulus Absolute Category Rating (ACR) from 40 subjects. The videos were reproduced on two different tablet devices: on an Apple iPad 2 and on a Samsung Galaxy Tab GT-P1000. Subjective observations include the perceived quality considering all aspects of the fruition chain, from the coding to the environment and device specific issues. The evaluation results also provided several remarks that can be helpful in designing systems and applications for multimedia contents played back on tablets. The results show a good correlation between subjective observations and some impairments so that a simple QoE index is proposed. A correlation study between subjective assessment and objective metrics is also provided.**

***Keywords*— subjective video quality assessment; QoE video streaming.**

I. INTRODUCTION

Traditionally, the performance of network architectures has been evaluated through Quality of Service (QoS) metrics, while lately the focus is turned onto Quality of Experience (QoE) metrics. In fact, the knowledge of QoS parameters is not enough to assure an overall satisfaction to the user, because each person is different and has diverse quality expectations. For such reasons QoE has become very important in the last years and the definition of related models is a growing research theme within the telecommunication community. To define QoE models, it is necessary to conduct extensive subjective quality assessment by groups of selected subjects under well-established conditions.

Over the past years, handheld devices, like smartphones and tablets, have had an impressive market boom. They evolved under every technological aspect and their performance, from computational power to display resolution, from connection bandwidth to user-interface logic. Accordingly, they are more and more frequently used for applications that were previously performed only on PCs and/or laptops. One of these applications is the streaming of videos over the Internet (mostly for entertainment or learning purposes). In this work we are interested in analyzing the perception of quality in this application scenario, specifically when the video sequences are transmitted over lossy wireless channels. The analysis has been conducted observing the subjective video quality affected by different impairments of the fruition chain, from the coding to handheld device specific issues, including the effects of various transmission errors.

In a real video streaming transmission over lossy wireless channel, the received video sequences are degraded versions of the original ones, due to the lossy coding and the channel transmission errors. Then, the annoyance caused by various distortion effects on the received video sequences may vary greatly among users. For instance, playback interruptions can be more annoying than the degradation due to low-rate video compression to some people, whereas others prefers watching HD quality video at the expenses of some frame freezing occurrences. A subjective quality assessment has then been conducted with the help of 40 subjects that were asked to rate several video sequences affected by various types of distortions. From the analysis of the assessment data, we have extracted a simple QoE index and provided a correlation study between subjective assessment and objective metrics.

In Section II a background on QoE papers is presented. Section III illustrates the test material used for the assessment. Section IV describes the subjective video quality assessment procedure while experimental results are reported in Section V. Finally, Section VI concludes the paper.

II. PAST WORKS

There are extensive studies that have been carried out in the recent past which focus on subjective video quality assessment; however, most of these consider the fruition of multimedia contents through conventional Personal Computer (PC) monitors and only a few deal with smartphones/tablets displays. In [1], forty subjects were asked to rate 78 video sequences generated from 6 different videos at CIF spatial resolution and affected by different Packet Loss Rate (PLR), ranging from 0.1% to 10%. Video sequences were played back on a desktop video display and the method adopted in their subjective evaluation was the ACR (Absolute Category Rating) [2]. The scores assigned by the observers were averaged to obtain the Mean Opinion Score (MOS). In [3], subjective tests were conducted with Quarter CIF (QCIF) videos over two devices: a PC and a mobile handset. The dataset with 90 test conditions was generated with a combination of parameters associated with the access network (block error rates and mean burst lengths), H.264 coding parameters (sender bitrates) and content types. Forty people participated in the assessment and rated each test clip in a discrete 5-level scale from 1 to 5. Final results show that the same conditions gave rise to greater annoyance (lower MOS values) on mobile handsets than on PC displays. In [4], extensive subjective view tests were conducted for assessing the perceptual quality of low bitrate videos which cover 150 test scenarios and include five distinctive dimensions: encoder type, video content, bitrate, frame size and frame rate. By watching the video sequences in a PC monitor, twenty subjects rated all videos. As a result, they found that in general the perceptual quality of a decoded video is affected by the encoder type, video content, bitrate, frame rate and size in a descending order of importance.

Other works focused on the evaluation of the QoE for generic services (not only video consumption) to find a correlation with QoS parameters and/or to study which factors influence the QoE. In [5], the authors recruited 29 Android users with three types of phones and collected their QoE ratings for various Android applications. Qualitative and quantitative methods were used, which consisted in both unobtrusive context data collections on the user's mobile phone executed multiple times per day and explicit user feedbacks on the perceived QoE, performed weekly. The survey was aimed at gathering information about the user experience, the location, the social context and the mobility level. In [6], the authors surveyed existing studies on subjective quality assessment of scalable video coding, comparing and analyzing them with respect to the considered scalability dimensions. They conclude that it exists a subjective preference among the scalability options, which varies mainly according to bitrate conditions, but also depends on several factors such as: content, encoder types, viewing environments and even user's experience and expectation. In [7], statistical classification techniques were used to detect on-the-fly the application that runs inside the thin-client protocols, by passively observing packet size and rate of the thin-client connection. Then, such information was correlated with the Round Trip Time (RTT) to design a threshold-based algorithm able to infer users QoE. To set the threshold values a subjective quality assessment was performed; 41 subjects were asked to act like a thin-client user and run a specific class of applications for a known period of time, during which RTT values of the connection are recorded. At the end of the test each user gave a quality score so that for each class of applications the correlation between the RTT and the users experience was computed. In [8], the authors highlight the importance of the perceived service quality as differentiator between cloud service providers. Then, they introduced a new categorization scheme for cloud services in order to match the end user experience and usage domain which directly influence the QoE requirements. The authors analyzed new challenges for QoE management of cloud applications and services, with the goal to deliver the cloud application to the end user at high quality while minimizing the costs of different players of the cloud computing stack and the underlying network providers. In [9], the authors introduce the HTTP Adaptive Streaming (HAS), a form of Internet video delivery to end users allowing enhanced QoE. The development of QoE evaluation methodologies, performance metrics and reporting protocols plays a key role in optimizing the delivery of HAS services. In particular, QoE monitoring and feedback are beneficial for detecting and debugging failures, managing streaming performance, enabling intelligent client adaptation, and allowing for QoE-aware network adaptation and service provisioning. Finally, in [10] an interesting QoE interaction model is defined. It considers a communication ecosystem that spans across different domains such as technological aspects, business models, human behavior and context, beyond to regard the interactions between this disciplines in terms of business objective, user experience and customer experience.

In the proposed work we considered video consumption on tablet devices while almost all studies on past works considered video consumption on a fixed PC monitor. Watching a media content on handheld devices is very different to watching the same content on a PC monitor at a fixed distance. Furthermore we observed the subjective video quality considering all aspects of the fruition chain, from the coding to the environment and handheld device specific issues, including the effect of various transmission errors. Studies on past works didn't consider all these aspects.

III. TEST VIDEO SEQUENCES

We considered four YUV video sequences at CIF spatial resolution (352×288 pixels), namely *Big Buck Bunny, Elephants Dream*, *Highway* and *Paris*. All the original sequences are available in raw progressive format at frame rate of 24fps [11]. Only 480 frames (20sec) have been used from the overall original sequences; the frame ranges are specified in Table I. 54 corrupted video sequences have been generated for each of the four considered videos in order to simulate the video streaming transmission in a lossy wireless channel. Table II illustrates the parameters related to both co-decoding and transmission errors that were combined to generate the corrupted video sequences. Note that in this context *overflow* indicates a playback interruption due to buffer starvation at the decoder. To transmit the corrupted video sequences from the video server to the tablet devices a dedicated wireless LAN has been used with an adequate available free band. With such "ideal" wireless channel the degradations of the received video sequences should only be those voluntarily included. Consequently, the error parameters could be kept under control, so that more specific results could be obtained from the subjective assessment.

The playout delay and the overflow errors were introduced directly in the raw video sequences. The playout delay causes a delay in the start of the video playback, which has been set to 0, 3, or 6sec. The overflow error consists in the repetition of a random frame of the video sequence for a random time period (average of 1sec and standard deviation of 1sec). Overflow errors simulate the condition of decoder buffer underflows determined by peaks in the network latencies. Overflow errors result in the duplication of the last correct frame at the decoder side, thus appearing as a temporary freezing of the scene. Such duplication can occur 0, 1, or 3 times in the proposed corrupted videos.

After the introduction of these errors, all the raw corrupted versions of the video sequences have been encoded with H.264. Table III illustrates the encoding parameters used to set up the JM18.0 version of the H.264/AVC reference software [12] used to encode the video sequences.

TABLE I. FRAME RANGES FROM THE ORIGINAL VIDEO SEQUENCES.

Video	Total number of frames	Range of frames considered
Big Buck Bunny	14315	4150 – 4630
Elephants Dream	15691	12000 – 12480
Highway	2000	750 – 1230
Paris	1065	0 – 480

TABLE II. VALUES OF THE ENCODING AND TRANSMISSION ERROR PARAMETERS USED TO GENERATE THE TEST SEQUENCES.

Parameter	Acronym	Value		
Bitrate [kbps]	Rate	200		400
Playout Delay [sec]	PD	0	3	6
Overflow [occurrences]	Over	0	1	3
Packet Loss Rate [%]	PLR	0	0.4	3

TABLE III. VALUES OF THE ENCODING PARAMETERS.

Encoding Parameter	Value
Frame rate	24 fps
Number of frames	480
Profile	Extended
Chroma Format	4:2:0
GOP Size	16
GOP structure	IBBPBBPBBPBBPBB
Number of reference frames	5
Entropy coding method	Universal Variable Length Code (UVLC)
Slice mode	Fixed number of macroblocks
Rate control	Enabled
Macroblock partitioning for motion estimation	Enabled
Motion estimation algorithm	Enhanced Predictive Zonal Search (EPZS)
Early skip detection	Enabled
Selective intra mode decision	Enabled

Each frame is divided into a fixed number of slices, where each slice consists of three full rows of macroblocks. Rate control has been enabled to determine two different video quality profiles: low (200kbps) and high (400kbps).

The PLR errors generation was based on the transmitter_simulator.exe software [13, 14]. This software discards a selected percentage of packets from the compressed H.264 bitstreams, to simulate the packet drop introduced by a lossy wireless channel. This is done through the definition of two error pattern files corresponding to PLRs of 0, 0.4% and 3% according to [15]. Once all 54 corrupted H.264 bitstreams had been generated for each video sequence, they were packaged in the MP4 format container, using the OGG Video Converter software [16].

Two websites have been implemented for the display of the video sequences and the scoring procedure: a website for the iPad 2 and a website for the Galaxy Tab. The iPad 2 website makes use of the html5 language which natively supports video sequences in the MP4 format. Both website and corrupted video sequences reside on the University server. The Galaxy Tab website is based on Flash technology, which allows retrieving the video sequences from YouTube servers by specifying the unique video code. In the latter case, all corrupted video sequences have been uploaded in a YouTube channel.

The choice of implementing two different solutions derives from the need for using the best technology for each terminal. Html5 is the optimal choice in case of the iPad 2, since it allows for accessing the video sequences from our servers, thus gaining complete control over the transmission, and since the iPad 2 does not provide support for the Flash technology. As for the Galaxy Tab, Flash presents a good reliability and the YouTube channel provides good throughput performances. The two implementations have been extensively verified before the subjective tests. The final results seem to confirm such choices, since the judgments for the two devices do not show any significant gap on average. Moreover, such difference is transparent to the subjects; in fact, both websites consist in a login page and a video section where the subjects can watch and rate the video sequences.

IV. SUBJECTIVE VIDEO QUALITY ASSESSMENT

In the proposed assessment 40 students, 33 males and 7 females, were asked to rate 216 video sequences corresponding to four different video contents corrupted by various transmission errors. For subjective tests we adopted the specification given in [2].

The single-stimulus Absolute Category Rating (ACR) method has been adopted, which consists in presenting the test sequences one at a time and allowing for their independent evaluation on a category scale. The method specifies that after each presentation the subjects must be asked to evaluate the quality of the sequence just shown. The five-level scale in Table IV for rating overall quality has been used. The scores assigned by the observers are averaged in order to obtain the Mean Opinion Score (MOS).

The technical specifications of the tablets employed for the subjective video quality assessment are:

1) Apple iPad 2
- display: 9.7 inches diagonal, resolution of 1024x768 pixels, 132 pixels per inch (ppi), LED-backlit glossy widescreen Multi-Touch display with IPS technology;
- operating system: iOS 5.0.1
- browser: Safari

2) Samsung Galaxy Tab GT-P1000
- display: 7.0 inches diagonal with a resolution of 1024x600 pixels, 170 ppi, TFT LCD capacitive touchscreen;
- operating system: Android v2.2
- browser: Browser v2.2

Eighteen people watched the videos with the Galaxy Tab while twenty-two used the iPad 2. The assessment was organized in sessions of 1.5 hours, during which each subject voted for a total of 45 minutes, alternating 15 minutes for voting with 15 minutes of break. Three sessions were needed to complete the subjective assessment per subject, and one session was held every week.

TABLE IV. FIVE-LEVEL QUALITY SCALE USED TO RATE THE VIDEO.

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

The video sequences were displayed in full-screen mode for both tablets. A rating page was automatically displayed after the end of each video playback, so that the subjects had to choose a discrete value from 1 to 5 to vote the overall quality of the video sequence just shown. The voting period was not time-limited and after each vote the subjects had to confirm their choice. Upon confirmation, the next video page would be displayed. Such automatic procedure was iterated for all the 216 video sequences. The presentation order of the video sequences was randomized for each subject. A SQL database has been used to store all videos, votes and voting times submitted by each participant.

V. EXPERIMENTAL RESULTS

MOS and variance values have been computed for each video sequence from the analysis of the experimental data. Fig. 1 summarizes the MOS results for the iPad2 and the Galaxy Tab devices for the four video sequences coded at 200kbps and 400kbps respectively. In such plots, the size of each circle is directly proportional to the relative value. MOS and variance values are reported in Tables V-VIII for clarity, where the initial condition of no errors is represented by (0/0/0). MOS values are shown with a background color/shade scale, ranging from white to dark, so to help distinguish higher from lower values more clearly.

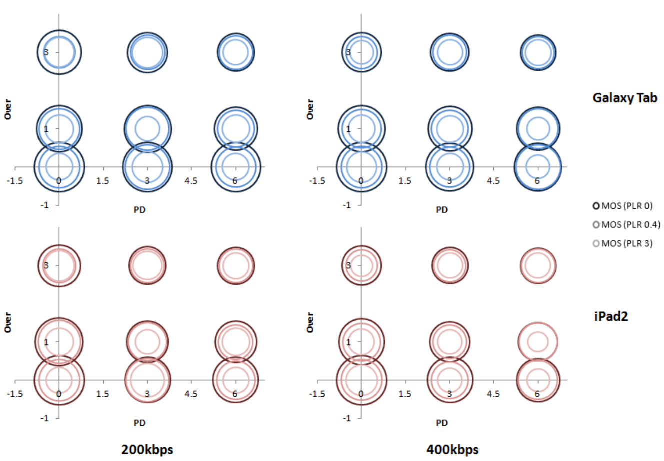

Fig. 1 MOS results for the iPad2 and the Galaxy Tab devices.

As expected, except for few error combinations, MOS results for the 400kbps sequences are higher than those for the 200kbps sequences. The error combinations that caused the higher annoyance to the subjects were those including the greater value of the PLR (3%), resulting in MOS values smaller than two in most cases. On the other hand, the higher scores correspond to sequences with PD errors alone, achieving MOS values greater than three. Then, it can be stated that an initial buffering interval is more tolerable for the subjects than the overflow and the PLR occurring during the video playback.

Even though the MOS results are very similar between the two tablet devices on equal terms of error conditions, it can be noted that, on average, the iPad 2 achieves a better performance when the PD is zero. Conversely, when the PD is greater than zero the Galaxy Tab achieves higher scores than the iPad 2. These results can be explained by the overall faster iPad 2 interface, which builds an expectation of efficiency that is becomes disappointed due to the PD effect.

TABLE V. MOS RESULTS FOR THE FOUR VIDEO SEQUENCES CODED AT 200KBPS.

PD/Over/PLR	iPad 2		Galaxy Tab		Average	
0/0/0	3.3		3.33		3.31	
0/1/0 – 0/3/0	3.18	2.77	3.08	2.94	3.13	2.86
0/0/0.4 – 0/0/3	2.75	1.72	2.75	1.78	2.75	1.75
0/1/0.4 – 0/3/0.4	2.84	2.22	2.74	2.13	2.79	2.17
0/1/3 – 0/3/3	1.85	2.02	1.9	1.97	1.88	2
3/0/0	3.01		3.33		3.17	
3/1/0 – 3/3/0	2.78	2.48	3.14	2.67	2.96	2.57
3/0/0.4 – 3/0/3	2.86	2.22	3.03	2.08	2.95	2.15
3/1/0.4 – 3/3/0.4	2.58	2.24	2.92	2.31	2.75	2.27
3/1/3 – 3/3/3	1.61	1.9	1.61	2.03	1.61	1.96
6/0/0	3.02		3.32		3.17	
6/1/0 – 6/3/0	2.76	2.43	2.92	2.5	2.84	2.47
6/0/0.4 – 6/0/3	2.57	1.74	2.68	1.78	2.62	1.76
6/1/0.4 – 6/3/0.4	2.3	2.23	2.5	2.25	2.4	2.24
6/1/3 – 6/3/3	1.81	1.7	1.82	1.68	1.81	1.69

TABLE VI. MOS RESULTS FOR THE FOUR VIDEO SEQUENCES CODED AT 400KBPS.

PD/Over/PLR	iPad 2		Galaxy Tab		Average	
0/0/0	3.94		3.74		3.84	
0/1/0 – 0/3/0	3.65	3.08	3.51	2.96	3.58	3.02
0/0/0.4 – 0/0/3	3.25	2.45	3.1	2.29	3.17	2.37
0/1/0.4 – 0/3/0.4	2.98	2.52	2.92	2.4	2.95	2.46
0/1/3 – 0/3/3	2.05	1.69	2.04	1.75	2.04	1.72
3/0/0	3.6		3.57		3.59	
3/1/0 – 3/3/0	3.18	2.85	3.32	2.85	3.25	2.85
3/0/0.4 – 3/0/3	2.94	1.95	2.99	1.99	2.96	1.97
3/1/0.4 – 3/3/0.4	2.69	2.56	2.79	2.57	2.74	2.56
3/1/3 – 3/3/3	1.95	2	2.01	1.89	1.98	1.94
6/0/0	3.45		3.56		3.51	
6/1/0 – 6/3/0	3.09	2.82	3.28	2.65	3.18	2.74
6/0/0.4 – 6/0/3	3.11	1.8	3.36	1.78	3.24	1.79
6/1/0.4 – 6/3/0.4	3.07	2.68	2.99	2.43	3.03	2.56
6/1/3 – 6/3/3	1.6	2.01	1.69	1.94	1.65	1.98

TABLE VII. VARIANCE RESULTS FOR THE FOUR VIDEO SEQUENCES CODED AT 200KBPS.

PD/Over/PLR	iPad 2		Galaxy		Average	
0/0/0	1.05		0.74		0.89	
0/1/0 – 0/3/0	0.77	0.69	0.71	0.64	0.74	0.66
0/0/0.4 – 0/0/3	0.81	0.55	0.58	0.47	0.70	0.51
0/1/0.4 – 0/3/0.4	0.63	0.56	0.69	0.60	0.66	0.58
0/1/3 – 0/3/3	0.61	0.64	0.73	0.60	0.67	0.62
3/0/0	0.84		0.69		0.76	
3/1/0 – 3/3/0	0.58	0.65	0.72	0.49	0.65	0.57
3/0/0.4 – 3/0/3	0.87	0.58	0.61	0.63	0.74	0.61
3/1/0.4 – 3/3/0.4	0.57	0.48	0.58	0.55	0.58	0.52
3/1/3 – 3/3/3	0.40	0.54	0.60	0.70	0.50	0.62
6/0/0	0.66		0.72		0.69	
6/1/0 – 6/3/0	0.82	0.66	0.51	0.55	0.66	0.60
6/0/0.4 – 6/0/3	0.71	0.49	0.37	0.64	0.54	0.56
6/1/0.4 – 6/3/0.4	0.54	0.52	0.59	0.67	0.56	0.60
6/1/3 – 6/3/3	0.68	0.59	0.59	0.54	0.64	0.57

TABLE VIII. VARIANCE RESULTS FOR THE FOUR VIDEO SEQUENCES CODED AT 400KBPS.

PD/Over/PLR	iPad 2		Galaxy		Average	
0/0/0	1.03		0.86		0.95	
0/1/0 – 0/3/0	1.01	0.73	0.80	0.69	0.91	0.71
0/0/0.4 – 0/0/3	0.94	0.90	1.09	0.65	1.01	0.78
0/1/0.4 – 0/3/0.4	0.99	0.66	0.77	0.72	0.88	0.69
0/1/3 – 0/3/3	0.52	0.31	0.71	0.58	0.61	0.45
3/0/0	1.06		0.80		0.93	
3/1/0 – 3/3/0	0.79	0.82	0.71	0.80	0.75	0.81
3/0/0.4 – 3/0/3	0.84	0.46	0.74	0.71	0.79	0.59
3/1/0.4 – 3/3/0.4	0.91	0.49	0.78	0.80	0.84	0.64
3/1/3 – 3/3/3	0.57	0.63	0.89	0.57	0.73	0.60
6/0/0	0.98		0.74		0.86	
6/1/0 – 6/3/0	0.89	0.74	0.84	0.98	0.86	0.86
6/0/0.4 – 6/0/3	0.94	0.52	0.72	0.47	0.83	0.49
6/1/0.4 – 6/3/0.4	0.65	0.97	0.83	0.67	0.74	0.82
6/1/3 – 6/3/3	0.48	0.71	0.71	0.72	0.60	0.71

In order to define a QoE index able to evaluate the perceived quality we investigated the suitability of a linear function of the considered impairments as follows

$$Q.I. = a_1 \cdot Rate + a_2 \cdot PD + a_3 \cdot Over + a_4 \cdot PLR + K,$$

with a_1, a_2, a_3, a_4 and K coefficients. Table IX shows the values of the coefficients obtained by fitting the linear model to the MOS results.

The values of the coefficients are very similar for both devices, except for a_2, whose value is almost twice for the iPad 2 with respect to the Galaxy Tab. This confirms that the annoyance generated by the playout delay error is greater in the iPad 2 with respect to the Galaxy Tab.

TABLE IX. VALUES OF THE QUALITY INDEX EQUATION COEFFICIENTS.

Coefficient	iPad 2	Galaxy Tab	Average
a_1	0.00150	0.00096	0.00123
a_2	-0.03767	-0.02045	-0.02906
a_3	-0.13380	-0.15620	-0.14500
a_4	-0.36290	-0.38200	-0.37245
K	2.61650	2.68750	2.65200

Then, the Pearson correlation between the proposed quality index and the MOS is 0.841 for the iPad 2 and 0.8567 for the Galaxy Tab respectively. The MOS values and the Pearson correlation are represented in Fig. 2 and Fig. 3 for the iPad 2 and the Galaxy Tab respectively. By considering the average coefficients reported in Table IX, we achieve correlation values of 0.8388 and 0.8546 between the proposed quality index and the MOS for the iPad 2 and the Galaxy Tab respectively. Then, given the strong correlation between the quality index and the MOS values, the proposed quality index could be used in a video streaming transmission system to estimate the quality perceived at the receiver side by estimating or measuring the values of the error parameters affecting the video transmission.

Finally, a correlation study between the subjective assessment and several objective metrics is provided. Three objective quality metrics have been computed for all the corrupted video sequences: PSNR, SSIM and V_{dist} [17]. PSNR and SSIM are full-reference metrics while V_{dist} is a recent reduced-reference metric. Differently from PSNR, SSIM and V_{dist} implement some Human Visual System (HVS) perception concepts. In order to take into account the overflow and playout delay errors, two measuring modalities are defined:

- overflow occurrence: the objective metric is computed between the decoded "frozen" frames and the original frames that should have been sent in the ideal transmission (no overflow),
- playout delay: since such distortion cannot be measured through the chosen error metrics, we determined percentage values used to decrease the computed error metric in case of sequences affected by the playout delay error. Such values were computed by considering the average MOS difference between sequences affected and unaffected by the playout delay error, and by keeping constant all the other parameters. Table X summarizes such percentage values.

TABLE X. PERCENTAGE VALUES USED TO DECREASE THE OBJECTIVE METRIC MEASURES FOR THE SEQUENCES AFFECTED BY THE PLAYOUT DELAY.

Device	PD = 3 sec	PD = 6 sec
iPad 2	3.136%	4.520%
Galaxy Tab	0.278%	2.454%

Table XI summarizes the Pearson correlation values between the MOS values and the chosen objective metrics over all corrupted video sequences. The measurements related to sequences affected by the playout delay error have been corrected as described in the previous paragraph.

TABLE XI. PEARSON CORRELATION BETWEEN THE MOS VALUES AND SEVERAL OBJECTIVE METRICS.

Device	MOS - V_{dist}	MOS - PSNR	MOS - SSIM
iPad 2	0.300	0.5748	0.6696
Galaxy Tab	0.2021	0.4979	0.6375

As expected in accordance with the recent literature [18], the SSIM has the higher correlation with the MOS while the V_{dist} achieves the worst result.

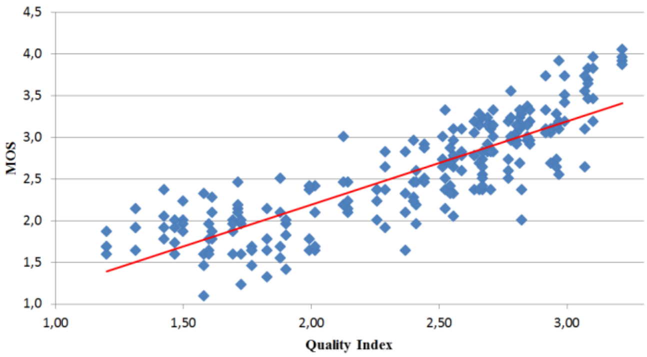

Fig. 2 Pearson correlation between the MOS values and the Q.I. for the iPad 2.

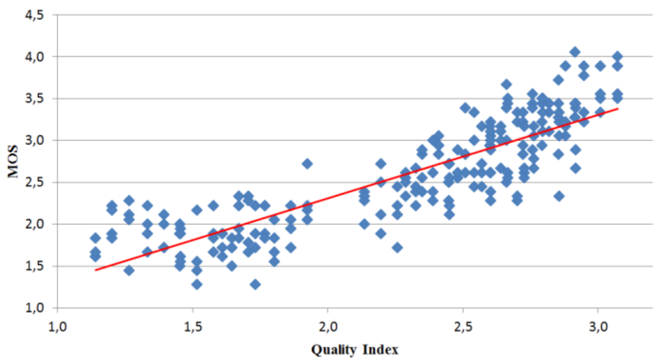

Fig. 3 Pearson correlation between the MOS values and the Q.I. for the Galaxy Tab.

VI. CONCLUSIONS AND FUTURE WORK

A subjective video quality assessment has been conducted on 216 video streams encoded at two different bitrates with the H.264/AVC and corrupted by typical wireless channel transmission errors: decoding buffer overflows, playout delay and packet loss. The peculiarity of the proposed work is that it is meant to assess the QoE in a realistic wireless scenario in the case of tablet device utilization. In fact, two tablet devices have been used: the Apple iPad 2 and the Samsung Galaxy Tab GT-P1000.

The MOS results show that the occurrence of packet loss errors are perceived as the worst degradation types while the sequences affected by high playout delays only provide the highest perceived quality. A simple quality index is consequently proposed. Finally we provided a correlation study between the MOS values and three objective metrics: PSNR, SSIM and V_{dist} showing that the SSIM achieves the highest correlation with subjective evaluations.

The present study is meant to be taken further by extending both the number of considered devices and the number and characteristics of the subjects.

The goal is to thoroughly analyze the QoE perceived by different categories of end users that are watching the same video content in tablet devices that vary in terms of display size and resolution, hardware configuration and operative system.

REFERENCES

[1] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S.Tubaro, and T. Ebrahimi, “Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel”, Quality of Multimedia Experience (QoMEx), 2009, pp. 204-209, 29-31 July 2009.

[2] ITU-T Recommendation P.910, “Subjective video quality assessment methods for multimedia applications”.

[3] A. Khan, L. Sun, J.-O. Fajardo, I. Taboada, F. Liberal and E. Ifeachor, “Impact of end devices on subjective video quality assessment for QCIF video sequences”, Quality of Multimedia Experience (QoMEx), 2011, pp. 177-182, 7-9 Sept. 2011.

[4] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang and M. Etoh, “Cross-Dimensional Perceptual Quality Assessment for Low Bit-Rate Videos”, IEEE Trans. on Multimedia, vol. 10 no.7, pp. 1316-1324, Nov. 2008.

[5] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J.-H. Hong and A.K. Dey, “Factors Influencing Quality of Experience of Commonly Used Mobile Applications”, IEEE Communications Magazine vol. 50 no.4, pp. 48-56, April 2012.

[6] J.-S. Lee, F. De Simone and T. Ebrahimi “Subjective quality assessment of scalable video coding: A survey”, Quality of Multimedia Experience (QoMEx), 2011, pp. 25-30, 7-9 Sept. 2011.

[7] M. Dusi, S. Napolitano, S. Longo and S. Niccolini, “A closer look at Thin-Client connections: Statistical Application Identification for QoE Detection”

[8] T. Hobfeld, R. Schatz, M. Varela and C. Timmerer, “Challenges of QoE Management for Cloud Applications”, IEEE Communications Magazine vol. 50 no.4, pp. 28-36, April 2012.

[9] O. Oyman and S. Singh, “Quality of Experience for HTTP Adaptive Streaming Services”, IEEE Communications Magazine vol. 50 no.4, pp. 20-27, April 2012.

[10] K.U.R. Laghari, N. Crespi and K. Connelly, “Toward Total Quality of Experience: A QoE Model in a Communication Ecosystem”, IEEE Communications Magazine vol. 50 no.4, pp. 56-65, April 2012.

[11] trace.eas.asu.edu/yuv/

[12] http://iphome.hhi.de/suehring/tml/

[13] http://vqa.como.polimi.it/sequences.htm

[14] https://sites.google.com/site/matteonaccari/software

[15] S. Wenger, “Proposed Error Patterns for Internet Experiments”, Q15-I-16.zip, October 1999.

[16] http://www.brothersoft.com/ogg-video-converter-435475.html

[17] L. Ma, S. Li, F. Zhang and K.N. Ngan, “Reduced-Reference Image Quality Assessment Using Reorganized DCT-Based Image Representation,” IEEE Trans. on Multimedia, vol. 13 no. 4, pp. 824 - 829, Aug. 2011.

[18] W. Lin, C.-C. Jay Kuo, “Perceptual visual quality metrics: A survey”, Journal of Visual Communications, 2011.