4th International Workshop on Perceptual Quality of Systems (PQS 2013)
2-4 September 2013, Vienna, Austria

Impact of non-QoS related parameters on video QoE

Hajer Gahbiche Msakni, Habib Youssef

PRINCE Research Unit, ISITCom, University of Sousse, Sousse, Tunisia

hajer.gahbiche@gmail.com, habib.youssef@fsm.rnu.tn

Abstract

Nowadays a huge amount of multimedia content is available over different Online Social Networks (OSN) Web sites such as Facebook, YouTube…Therefore, the proposal of mechanisms capable to assessing the user satisfaction, i.e. Quality of Experience (QoE), become imminent. Different work has shown that the QoE may be affected by different QoS parameters. However, these parameters are not the only factors that might affect the QoE. This paper considers two non-QoS parameters: the video length and the different moment of the day. This paper presents a comparative study based on subjective assessment, that consider different lengths of video sequences, i.e. 30 seconds, 1 minute 30s, 3 minutes, 5 minutes and 8 minutes. These videos sequences belong to four different classes: funny, documentary, soccer and film class. This paper considers also different moment of the day, i.e. morning, midday, afternoon and night, in the assessment process.

Index Terms: Quality of Experience, subjective assessment, video sequence's length, moment of the day.

1. Introduction

Every second millions of users enjoy video streaming through social networking application, such as YouTube, Facebook, Twitter, etc. Users create a huge amount of interaction within video and between users. For June 2012; YouTube registered more than 72 hours of uploaded video, and more than 3 billion hours of watched video during a month [1]. It registered also, 500 years of YouTube video that are watched every day on Facebook [1]. These videos have different characteristics with regards to the content, the video length, and the video quality. These different characteristics may affect the acceptability of this content by the user. In this context, a new concept appeared. This concept, take into account the degree of satisfaction of the users, namely Quality of Experience (QoE). The QoE may be affected by different factors [2].

This paper presents a comparative study based on subjective QoE assessment. This work consider different lengths of video sequences, i.e. 30 seconds, 1minutes and 30 seconds, 3 minutes, 5 minutes, and 8 minutes, belonging to four content classes. The second scoop of this paper is the consideration of different moment of the day, i.e. morning, midday, afternoon and night, in the assessment process.

This paper is organized as follow section 2 presents our motivation and related work. Section 3 presents our methodology in the subjective assessment. This section describes the user's profiles, the video content and the assessment process. Section 4 presents and discusses our finding. The last section concludes this abstract.

2. Motivation and related work

Quality of Experience has served as a central research topic during the last five years. Different definition was proposed for the QoE. According to the IUT-T in Rec.P.910, "The QoE represents the overall acceptability of an application or service, as perceived subjectively by the end-user [2]".

According to different works the QoE may be influenced by different factors. In this context, the work in [1] discusses the different factors that influence the QoE. These factors may be classified as Quality of Presentation factors, such as used codec, used equipments, Application QoS parameters …, Quality of Delivery factors, such as type of network, Network QoS parameters,…, and human factors, i.e. age, psychology, users profile ...

The work presented in [4] has shown the impact of the video content and the video rate on the assessed video quality. The video content was selected from social network. This work considered a statistical analysis on the presentation of results. However, this work hasn't specified the length of the considered video sequences.

[5] presented a study based on the length of video sequences, which vary from 30 seconds to 240 seconds. This work considered also different video classes: action scenes, soccer, and documentary. It has shown, through a statistical analysis, that the video duration and the video content affect the MOS scores. Nevertheless, the considered video durations are very short and didn't reflect the duration of existing video sequences, in social network.

The work in [6] considered an everyday-life context for the video assessment process. The strength of this method is that video and audio quality is assessed in the real-life context of the user [6]. This work considered different video content. However, it didn't specify the duration of video sequences, considered in the subjective assessment. Moreover, this work didn't present the moment in which the different tests were done: in the morning, in midday, in the

10.21437/PQS.2013-10

afternoon or in the night. This point is also applicable for all the cited work above, i.e. [4] and [5].

Therefore, in this paper we will present the impact of both factors that may affect the Quality of Experience: the length of video sequence and the different moment of the day. The first factor was set to 30 seconds, 1 minute 30s, 3 minutes, 5 minutes and 8 minutes. The considered moment of the day was the morning, midday, the afternoon, and the night.

3. Subjective QoE Assessment

Subjective assessment is based on humans' viewers. These viewers evaluate a multimedia content in a real environment. The most used metric in subjective assessment is the Mean Opinion Score (MOS) [7]. We followed the general recommendations for subjective assessments methods are described in ITU-R T.500-11 [9] and ITU-T Rec. P.910 [3]. The next subsection describes the general characteristic of these tests.

3.1 Person profile

In this part of the study we considered 56 participants. The age of these participants vary from 21 to 51 years. The participant considered in this part of our study are not accustomed to subjective tests, they are not, also, expert in the field of multimedia.

3.2 Video Clip characteristics: duration and encoding

We considered four video content classes: funny video, documentary, soccer and films video sequences, from YouTube. In each content class we considered 5 different video sequences, which have different content and different duration: 30 seconds, 1minutes 30, 3 minutes, 5 minutes and 8 minutes. The video codec used in this context is MPEG-4 combined with AAC audio codec with 48 KHz and 126 KHz sampling rate. The original video bit rate varies from 1Mbits to 1,5Mbits. We considered, in this part of our study 3 quality levels: Medium Quality (MQ), i.e. 512Kbits, Original video Quality (OQ), and Low Quality (LQ), i.e. 200Kbits. Some of these video sequences can be found in [10].

3.3 Assessment Process

The goal of these subjective tests is to study the impact of video duration and the different moment of the days on the QoE. The different tests were done in laboratory: for these different moments of the days: morning, midday and afternoon, and night. The assessment process was organized as follow. All participants were divided into 4 groups, G1, G2, G3 and G4, according to the different moment of the days: morning, midday, afternoon and night. The first group of participant, i.e. G1, was asked to come the first days at 8 clocks, to evaluate the first video content, i.e. funny video. The second day the same group was asked to come at midday, to evaluate the documentary content. The third days this group was asked to come at 15 clocks, in the afternoon, to evaluate the soccer. And for the last days we asked them to come at 19 clocks, to evaluate the last content. The same process was done with the different others groups of participant. In a way that every day, for the same content, we have a group of participant: in the morning, i.g. G1, in midday, i.g. G2, in the afternoon, i.g. G3, and in the night i.g. G4. All groups evaluated the different content class, i.e. funny (C1), documentary (C2), soccer (C3) and film (C4), with different video qualities. Table 4 present this organization.

	C1	C2	C3	C4
Morning	G1	G2	G3	G4
Midday	G2	G1	G4	G3
Afternoon	G3	G4	G1	G2
Night	G4	G3	G2	G1

Figure 1: Average MOS for different video lengths

In this part of our study, we considered different metrics to assess the video quality, apart the MOS: the Sound Quality (SQ), the Content (C), the Picture Quality (PQ) and the sequence length (L). The participant gave their rating according to 5 different scales: Totally disagree, disagree, neutral, agree, and totally agree, for each of these metrics.

4. Results and discussion

In this section, we present our finding we will show the variation of the MOS and present a statistical analysis of these results. Two statistical test values were considered in this work: F and the p-value, as considered in [5]. In this work, the significance level considered is equal to 0.05. If the p-value is less than 0.05 mean that the considered variable is statistically significant and has an influence on the MOS, otherwise the MOS is not affected by such variable.

4.1 Impact of the video length

As presented previously, we considered different video length. These video lengths vary from 30 seconds to 8 minutes. These video clips belong to different video class and have different video quality. Figure 2 present the variation of the MOS for different lengths. This figure shows that the value of the MOS varies from 3.62 (SD= 0.81) to 3.74 (SD=0.73). Long duration scores (3min, 5 min and 8 min) better than short duration (30 sec, and 1 min 30).

We observed a high significance of the sequence length (L), F= 6,608 and p = 0,000. This may be explained by the increase of human attention through the time. This may lead to a positive evaluation, as it was also, explained by [5].

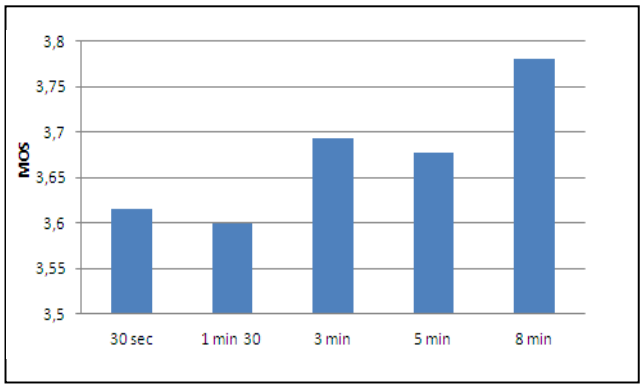

Figure 2: Average MOS for different video lengths

4.2 Impact of the content class

We considered in this work, 4 different content classes: funny, documentary, soccer, and action. The variation of the MOS, across all sequence length and video quality, is presented in Figure 3.

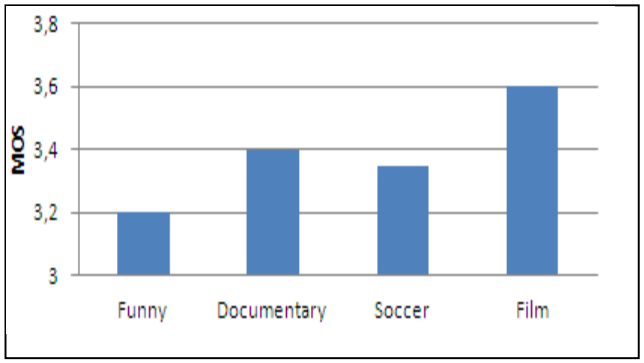

Figure 3: Average MOS for different content class

This figure shows that the film and the documentary categories score better that the funny and soccer. The MOS vary from 3,2 (SD= 0,72) and 3,6 (SD=0,72). This variation may be explained by the fact that we considered some video sequence related to some famous films, such Piranha, Taxi 3, the transporter. Where as in the documentary category we considered video sequences about magic trick, the marshmallow test… We observed, also, a high significance with to the content (C). the F value was equal to 22,74 and the p-value =0,000.

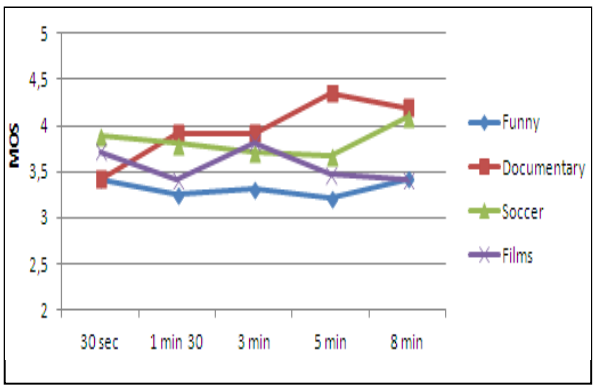

Figure 4: Average MOS for different content class different video duration

Figure 4 present the variation of the MOS, for different content class and different video duration. We observed a high significance between the content (C) and the sequence length (L), p-value=0.000. Figure 4 shows that the value of the MOS, i.e. 4, 1, 4, 2, observed for the documentary and soccer, for long duration are better than other sequences. This may be explained by the choice of the video sequence [10]. The worst value, i.e. 3,4, observed for the film category for 8 minutes, may be explained also by our choice for this sequence. In fact, this sequence presents a snapshot of Agatha Christie Film [10].

The table above summarizes the significance, i.e. through the F-value and p-value, of the different subjective metrics considered in this paper: the Sequence Length (L), the content (C), the Picture Quality (PQ) and the Sound Quality (SQ).

Table 1. Significance of subjective metrics

Subjective metric	F	p-value
Content (C)	22,74	0.000
Sequence Length (L)	6,608	0.000
Picture Quality (PQ)	22,56	0.000
Sound Quality (SQ)	14,85	0.000

This table shows that the different subjective metrics considered on this work are statistically significant and have an impact on the MOS.

4.3. Impact of moment of the day

In this work different moment of the days were also considered: the morning, midday, afternoon and night.
Figure 5 present the variation of the MOS, across different content class, sequence length and video quality, for different moment of the day. This figure shows that high values of the MOS were observed in the afternoon, i.e. 3, 7 (SD=0,71) and in midday, i.e. 3, 6 (SD=0,70).

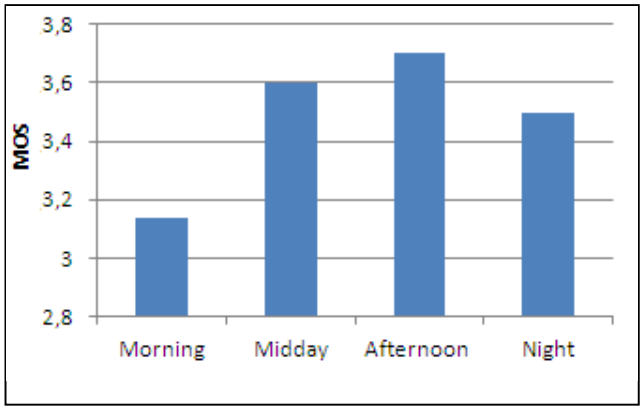

Figure 5: Average MOS for different moment of the day

This result may be explained by the psychological and emotional state of the different participant over these moments of the day. In fact these participant might be well-off and at ease on these moments. The result obtained for the night might be explained by the fact that they are tired, at the end of the day.

Figure 6 presents the variation of the MOS over different moment of the days for different video duration. This variation was obtained across different content class and different video quality. Figure 6 shows that the scores for the afternoon and midday scores better than the results obtained for the morning. The results obtained for long durations (3minutes, 5 minutes and 8 minutes) scores worst than whose obtained for short duration (30 seconds, 1 minutes 30), i.e. MOS<3,5. This might be explained by the decrease of the focus level over the night of the different participant. In this context, we identified a significance of the sequence length over the night, F=2,86 and p-value =0,05.

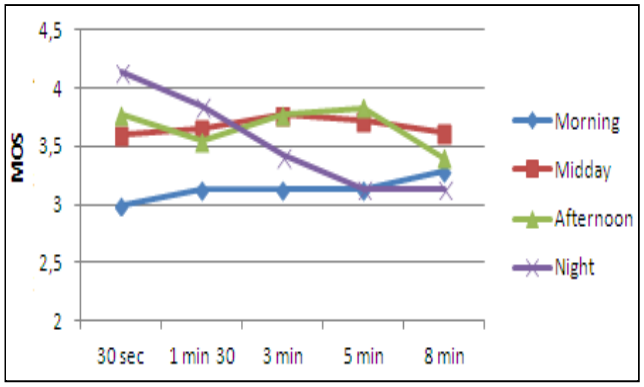

Figure 6: Average MOS for different moment of the day and different sequence lengths

With regard to the first moment of the day, i.e. the morning, the sequence length was not significant on 5% level. In fact, the F value and the p-value were equal, respectively to 3,72 and 0,72. The value of these parameters for the afternoon and the midday are as follow: F=4, 28, p-value =0, 02 and F=2,58, p-value=0,043. This result shows that the sequence length is statistically significant for the afternoon and the midday. Therefore the sequence length has an impact on the MOS, for the different moment of the day, except the morning.

With regard to the variation of the MOS, over different moment of the day and different content class, we have to notice that a video sequence might be characterized by its Spatial Information (SI) and its Temporal Information (TI) [3-4]. These information are respectively related to the different color, edge, etc, in the video sequence, and to the movement, direction, etc, in this video. In our study, the lowest values of these parameters were observed for the documentary, and the funny content class, $(SI, TI)_{Documentary} < (SI,TI)_{Funny}$. Highest values were observed for the films and for the soccer, $(SI, TI)_{Soccer} < (SI,TI)_{Film}$.

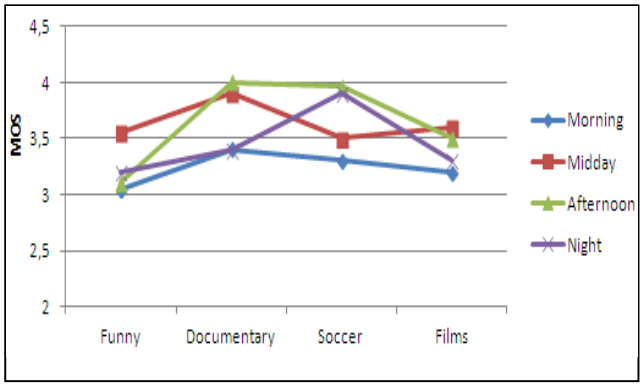

Figure 7: Average MOS for different moment of the day and different content class

Figure 7 presents the variation of the MOS for different content class over different moment of the day. This figure shows that the values of the MOS obtained for the morning, are lower than the other values, i.e. MOS<3.5. However, we have to notice that the value obtained for the documentary class scores better that the other content class, i.e 3,4. This might be explained by the fact that the different video sequences are well-nigh static and have the lowest value of Spatial and Temporal Information.
The content over the morning was not significant at 5% level, (p=0,372).

With regard to the last moment of the day, i.e. the night, Figure 7, shows that the funny and the documentary class scores worst, i.e. 3,2 and 3,4, compared to the other content class. This might be explained by the preference of action films and soccer over this moment of the day. We identified

a significance of the content over the night, F=5, 83 and p-value =0,01.
Over the different other moment of the day, i.e. midday and afternoon, a high significance were obtained toward the content. The obtained value are respectively, F=16, 12, p=0,000 and F=5, 75 and p-value=0,000.

Table 2 summarizes the significance, i.e. through the F-value and p-value, of the different subjective metrics considered, through different moment of the day: the Sequence Length (L), the content (C), the Picture Quality (PQ) and the Sound Quality (SQ).
This table shows that the assessed video quality was impacted by different video characteristics over different moment of the day. For example, in the morning the sound quality was the only significant parameter, at 5% level.
For the second moment of the day, i.e. midday, all the subjective video characteristics considered in this work are statistically significant. These parameters, i.e. content, sequence length, Sound Quality and Picture Quality, have an impact on the video quality.

Table 2. Significance of subjective metrics for different moment of the day

	Morning	**Midday**	**Afternoon**	**Night**
Content (C)	F=1,08 p=0,372	F=16,12 p=0,000	F=5,75 p=0,000	F=5,83 p=0,01
Sequence Length (L)	F=3,32 p=0,72	F=4,28 p=0,02	F=2,58 p=0,043	F=2,86 p=0,053
Sound Quality (SQ)	F=3,52 p=0,02	F=16,45 p=0,000	F=11,84 p=0,000	F=1,3 p=0,264
Picture Quality (PQ)	F=1,75 p=0,177	F=13,72 p=0,000	F=1,64 p=0,185	F=2,08 p=0,141

Over the night, only the content and the sequence length were statistically significant.
For the afternoon, three subjective metrics were significant and have an impact on the video quality.

5. Conclusion

This work studies the impact of video sequences lengths and the different moment of the day on the perceived QoE. Though subjective video assessment, we showed that the video quality may be affected by different factors such as: the content, the picture quality, the sound quality, and the sequence length.
This wok has shown also that the video quality may be affected by different characteristics through different moment of the day. It has shown that over midday and afternoon the users were affected by, almost, the different video characteristics. Whereas for the night, the video quality, as assessed by the user, were affected by the content and the sequence length.
This work can be considered as an extension of the work presented in [5], as it consider different other sequence lengths.
Our future work consists on studying some other factors related to the delivery of the multimedia content through different type of network.

6. References

[1] Chris Atkinson, "Youtube Statics 2012: A year Later: More Jawdropping Growth", June 2012, [On Line] : http://www.reelseo.com/youtube-statistics-growth-2012/

[2] Hajer Gahbiche MSAKNI, Habib YOUSSEF. Provisioning QoE over Converged Networks: Issues and Challenges. In : High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICESS), 2012 IEE 14th International Conference on. IEEE, 2012. p. 891-896.

[3] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications"," 2008.

[4] GARDLO, Bruno, RIES, Michal, RUPP, Markus, et al. A QoE evaluation methodology for HD video streaming using social networking. In : Multimedia (ISM), 2011 IEEE International Symposium on. IEEE, 2011. p. 222-227.

[5] FROHLICH, Peter, EGGER, Sebastian, SCHATZ, Raimund, et al. QoE in 10 seconds: Are short video clip lengths sufficient for Quality of Experience assessment?. In : Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on. IEEE, 2012. p. 242-247.

[6] V.Broeck, W.Jacobs, N.Staelens. "Integrating the everyday-life context in subjective video quality experiments." Quality of Multimedia Experience (QoMEX), Fourth International Workshop on. IEEE, 2012.

[7] ITU-T Recommendation P.10, Vocabulary for performance and quality of service, July 2006.

[8] VAN DEN BROECK, Wendy, JACOBS, An, STAELENS, Nicolas. Integrating the everyday-life context in subjective video quality experiments. In : Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on. IEEE, 2012. p. 19-24.

[9] ITU-T Recommendation BT-500-13,''Methodology for subjective assessment of the quality of Television Picture", 2012

[10] Video Sequences : https://sites.google.com/site/hejergahbiche/videos