Analysis of Temporal Effects in Quality Assessment of High Definition Video

Martin SLANINA, Tomáš KRATOCHVÍL, Ladislav POLÁK, Václav ŘÍČNÝ

Dept. of Radio Electronics, Brno University of Technology, Purkyňova 118, 612 00 Brno, Czech Republic

slaninam@feec.vutbr.cz, kratot@feec.vutbr.cz, xpolak18@stud.feec.vutbr.cz, ricny@feec.vutbr.cz

Abstract. *The paper deals with the temporal properties of a scoring session when assessing the subjective quality of full HD video sequences using the continuous video quality tests. The performed experiment uses a modification of the standard test methodology described in ITU-R Rec. BT.500. It focuses on the reactive times and the time needed for the user ratings to stabilize at the beginning of a video sequence. In order to compare the subjective scores with objective quality measures, we also provide an analysis of PSNR and VQM for the considered sequences to find that correlation of the objective metric results with user scores, recored during playback and after playback, differs significantly.*

Keywords

ITU-R BT.500, subjective test, Video quality, peak signal to noise ration, video quality metric.

1. Introduction

In the recent years, the video processing community has been largely interested in the quality aspects of the video content delivered to the user. There was a number of subjective, user-based test methodologies defined (e.g. [1]–[4]) – they consider the scenario when a user, as the consumer of the video information, assesses the quality as perceived by himself. Even a bigger number of objective, automated measurement procedures were introduced to substitute the expensive and cumbersome subjective scoring sessions (e.g. [5]–[9]). Still, research in both areas is very active as no universal user-based test methodology has been defined yet, neither has been defined a satisfactory automated algorithm to fully estimate the results of a selected user-based subjective test.

The ambition of this paper is in exploring the temporal behavior of the users when content of suddenly changing quality levels is presented, especially in quantifying the reactive times needed to adjust a slider (human interface device) to the desired position and stabilize the scores for a video sequence with close to constant quality. The question to be answered is: Are the users able to instantly follow the quality changes or is there a significant reactive time that needs to be considered?

An approximation of the time behavior of users was considered in e.g. [10] in terms of defining a temporal pooling algorithm. The temporal pooling is a technique for converting the measured quality values sampled at different time instants (typically calculated for each frame) to a continuous quality curve. Finally, the better the curve follows the user's subjective ratings scanned over time, the higher is the performance of the pooling mechanism and of the metric itself. In contrary to [10], we are studying the user behavior in a longer period of time (several seconds) than such temporal pooling algorithms usually address.

Undoubtedly the most common setup of a real video transmission system is the reference-free scenario. In such case, only the material on the output of a video processing system is available with no reference available for comparison. The quality test procedure needs to be tailored to the considered application and scenario in order to capture the phenomena that impact the results of a quality scoring session.

The research presented in this paper analyses the results of a specific user based quality test session in order to describe the temporal behavior of the assessors providing the scores. For this purpose, the basic principles of two standard test methods described in the Recommendation ITU-R BT.500 [1] are used, namely the SSCQS (Single Stimulus Continuous Quality Scale) and SSCQE (Single Stimulus Continuous Quality Evaluation).

Furthermore, we compare different approaches to reaching a single quality level for the whole video sequence. One such approach is based on averaging the user scores captured during the playback of the sequence (when the scores are stable) while the other approach uses scores given in a pause after a sequence when no video is played back.

The paper is organized as follows: Section 2 describes the test setup of the experiment – the video presentation scheme is characterized and the hardware and software used for the testing is described. Furthermore, the video material that was used for the testing is described and the different quality levels are introduced. The test scenario is also de-

scribed. Section 3 presents the findings of the two experiments that were performed. The paper concludes in Section 4.

2. Test Setup

This section describes the technical prerequisites of the experiment we performed and the design of the testing procedure. The hardware and software used for presenting the video sequences and collecting user ratings will be described, the selection of video sequences and their coding will be mentioned and, finally, the setup of the test session will be explained.

2.1 Interface Hardware

As the test setup requires collecting the users' ratings over time, a specific user interface hardware needs to be used. For this purpose, a slider interface was developed based on the guidelines presented in ITU-R Rec. BT.500 [1]. The interface uses a continuous quality scale with numeric values reaching from 0 (worst quality) to 100 (best quality). Furthermore, the scale is divided into five intervals (20 points each), having a quality label assigned as Bad (0 - 19), Poor (20 - 39), Fair (40 - 59), Good (60 - 79), Excellent (80 - 100). Even though these labels were not used in further processing, they served as a quality guideline for the users. The English labels were not in the mother tongue of the observers (all of them were Czech or Slovak), but any translation would be very likely to introduce inaccuracies in the meanings of the labels [18].

The user's rating is processed using the Atmel ATMEGA8A processor and transmitted to a personal computer over USB using the FTDI FT232 interface. As such, with the proper FT232 driver, the device behaves as a serial port peripheral from the programmer's perspective. It replies to a text query with a value corresponding to the slider position.

2.2 Video Presentation Software

The most important part of the video presentation software is its ability to synchronize the video time with the quality scores acquired from the peripheral interface. To achieve this, a special software tool was developed by the authors. It is based on two components - a Java application taking care of the test session setup and slider interface communication and the player component based on the VLC media player (ver. 1.1.5 and its Java-compatible mutation called vlcj) [11]. The vlcj provides an easy-to-use API whose functions can be called directly from the Java application. The user interface for test setup is displayed in Fig. 1.

After the test session is over, the time codes together with the corresponding user scores are stored in .csv files - basically, the data are organized in a text file separated by a semicolon. Consequent processing of such data can be done either using user-defined scripts (written in C/C++, awk, etc.) or in spreadsheet editors with statistical tools. The latter approach was used in our case.

2.3 Computer Configuration

The presentation of the test video sequences was done on a personal computer with an Intel Core2Duo E8400 CPU at 3 GHz, with 2 GB of memory running Microsoft Windows XP Professional. The output was brought via DVI interface to a Philips 240PW9ES LCD monitor. The tests were performed in a lab equipped with several computers of identical configuration and thus several observers performed the rating in parallel. The viewing distance varied between two and three times the height of the screen, chosen by the users to reach comfortable viewing. Although the distance may seem to be too short considering the ITU-R BT.500 preferred viewing distances, experiments show that especially for high definition content, the viewing distance might be decreased in order to strengthen the user involvement compared to standard viewing conditions [1], [12].

Using identical hardware itself does not assure equal viewing conditions. To keep them as close as possible, the monitors were adjusted to the same peak luminance (200 cd/m^2) and we also checked the ratio of inactive screen luminance to peak luminance, which shall be less than or equal to 0.02 according to BT.500. All the monitors reached performance well below this ratio, at the values between 0.001 and 0.005.

2.4 The Video Sequences

The tests required different video contents with different quality levels. As the source video sequences we used short uncompressed video clips in full HD resolution with interlaced scanning at 50 fields per second (1080i). They were retrieved as uncompressed .mov files from the local television broadcasting company CET 21, running the TV Nova channel. In fact, they were subject to lossy compression while being recorded to the HDCam tapes. Still, this compression does not introduce severe video image degradations. Due to the copyright agreement with the content provider, all video sequences were identified with a time code in the bottom part of the image. For research and non-commercial use, the sequences can be retrieved from [21].

Among the available content, we selected five sequences with the most diverse properties - reaching from static (paper) over low motion video (news) to highly dynamic content (hockey). The length of the sequences varied between 6 and 13 seconds, the shortest being the static newspaper sequence. Screenshots of the video sequences are shown in Fig. 2.

To introduce quality degradation, the video sequences were compressed with different video codecs at different target bitrates. The clue in the selection of the appropriate video

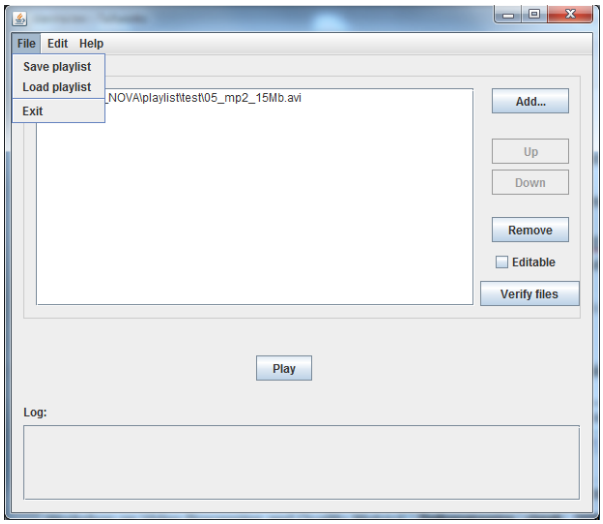

Fig. 1. Java application graphical user interface.

Fig. 2. The video sequences.

compression algorithms was found in the Blu-Ray standard, where three video codecs can be used for the high definition content, namely the MPEG-2 [13] with bitrates between 5 and 15 Mbit/s, H.264/MPEG-4 AVC [14] at 3 to 9 Mbit/s and VC-1 [15] at the same bitrates. To compress MPEG-2, the FFMPEG free software with the in-built encoder was used [16]. For MPEG-4 Part 10, a very well performing open implementation x264 was used [17]. Finally, the VC-1 videos were compressed using the official reference software provided by SMPTE.

For the playback of the videos after compression and decompression, two drawbacks had to be overcome. Firstly, the current implementation of the VLC media player, used in the playback application, is unable to open the VC-1 files created by the reference implementation of the VC-1 encoder. This issue can be easily solved by decompressing the video sequences first and playing them back decompressed. For full HD video sequences, another problem appears in such case as the data rates are quite high – for 1080i25, 8 bits per sample with 4:2:2 chroma subsampling, the required bit rate for raw material is $25 \cdot 1920 \cdot 1080 \cdot 8 \cdot 2 = 829.44$ Mbit/s. The available hard disks have difficulties in providing and guaranteeing such high data rates without special techniques such as RAID arrays. To reduce the required data rate, all the decompressed video sequences were further processed by the MPEG-2 encoder at a bitrate of 100 Mbit/s, which results in fluent playback with no visible quality degradation. The libx264 library with h264 lossless preset was also tested, but resulted bit rates around 200 Mbit/s which caused jerkiness on the hardware used for playback. As we do not compare different codecs (we just want to introduce impairments of different nature), a slight degradation of quality caused by the MPEG-2 recompression is tolerable.

2.5 The Test Session

The first test was aimed at analyzing the user behavior after a sudden quality change of the viewed video. Prior to the test session, observers were instructed to continuously adjust the slider position according to the instantaneously perceived quality. The video sequences were presented one after another, with an 8 second gray screen image between them. To familiarize the observers with the testing procedure, a short training session containing just three sequences (different from those used in the actual experiment) was performed first, but the scores recorded during the training session were discarded. Seventeen observers took part in the experiment, recruited from university students. They were all tested for visual acuity and color blindness prior to the test using the Snellen diagram and the Ishihara chart. In these tests, two of the students failed and could not continue participating in the test. The scores given by fifteen users were thus analyzed.

In the second test, the users were instructed to rate the video quality in the pause between consequent sequences. The position of the slider was scanned at the end of the 8 second gray scale image interval. Again, a short training was performed whose results were discarded. In the second test, the users were also recruited from university students and the same number of users (15) were considered in the analysis.

3. Results

3.1 Rapid Quality Change

The aim of the experiment was to examine the behavior of the users over time, i.e. how long it takes them to react on changing conditions and what time they need for the scores to stabilize at a certain value. The user ratings were recorded

twice in a second as recommended in the SSCQE method in ITU-R BT.500.

After the start of each video sequence, the user naturally needs some time to adapt to the change of quality in the presented content. This obviously happens at the beginning of each video sequence. During the playback of a video sequence, the quality is considered as constant – its changes are very small, compared to the variability among the sequences. This can easily be proved when an objective quality metric is evaluated for each frame of a sequence - the changes are very small during each of the sequences.

The time dependencies of the user ratings are shown in Figs. 3 to 5. Each curve represents the user scores averaged for all observers taking part in the experiment, thus denoted to as MOS - the mean opinion score. The ratings were recorded on a continuous scale reaching from 1 (worst quality) to 100 (best quality), thus the resulting MOS values fall within the same interval. For each time instant t_n in a video sequence, the MOS value can be expressed as

$$\mathrm{MOS}(t_n) = \frac{1}{U}\sum_{u=1}^{U}\mathrm{UQS}(u,t_n) \qquad (1)$$

where U is the number of users considered and $\mathrm{UQS}(u,t_n)$ is the user quality score collected for the time instant t_n from the user u.

Fig. 3 represents the case when the previous sequence was lower quality than the actual sequence. The user's adaptation results in moving the slider towards higher values. For five such cases, we can observe that the user reaction is represented by a sigmoid function with a delay aproximately 6 seconds until the user scores stabilize. The rise time of the curve represents the vast majority of the whole delay - there is approximately one second of user inactivity, followed by 5 seconds of slider adjustment.

The opposite situation is shown in Fig. 4, which represents the case when the actual sequence is coded with lower quality compared to the previously presented video sequence. Again, the curve follows a sigmoid function. In this case, the delay and the rise time of the MOS curves are longer yet comparable. It can also be observed that the users are inactive for about one second and then it takes about six seconds to adjust the quality on the slider. Finally, Fig. 5 represents the case when the actual and the previous sequences have similar quality and the user does not need to significanty move the slider to change his scores. As expected, the change of the user scores over time is not following any trend. Furthermore, the margin values in which the MOS scores change for a given content have no significant difference as the variance of the MOS values is likely to be up to about 20 % of the scale [18].

3.2 Overall Score per Sequence

In the second part of the experiment, we are trying to examine the scores received during the evaluation and put

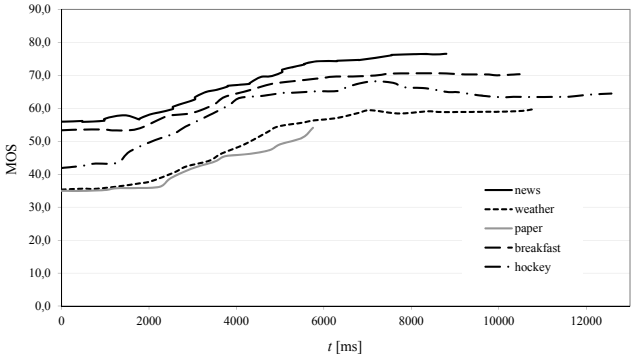

Fig. 3. MOS time dependency for different content when preceded by a lower quality sequence.

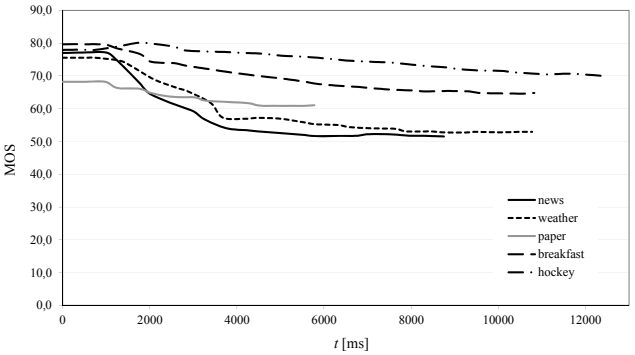

Fig. 4. MOS time dependency for different content when preceded by a higher quality sequence.

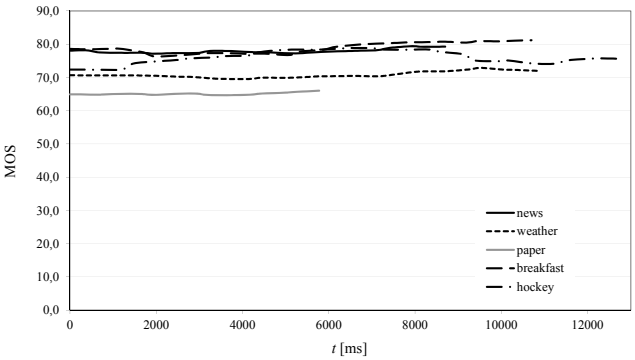

Fig. 5. MOS time dependency for different content when preceded by a comparable quality sequence.

them in correspondence with scores received in a pause inserted between the sequences. The motivation to this task is that we found that the correlation between the scores recorded during the previous test and several objective measurements was very poor.

Based on the findings in Sec. 3.1, in order to neglect the impact of delays in stabilizing the user ratings, we used the mean of the user ratings recorded after fifth second of a sequence to calculate the overall user score for each video sequence.

The user ratings $\mathrm{UQS}(u,t_n)$ are recorded in discrete time instants, twice in a second. The overall user rating for

a sequence taken as an average rating from one user recorded after fifth second of the sequence can be thus expressed as

$$\mathrm{UQS}(u) = \frac{1}{N} \sum_{t_n=5\ \mathrm{s}}^{\max(t_n)} \mathrm{UQS}(u, t_n) \tag{2}$$

where N is the number of samples recorded after the fifth second of a video sequence. Similarly, the mean opinion score MOS taking into account scores from all users for one sequence is calculated as the average of $\mathrm{UQS}(u)$, i.e.

$$\mathrm{MOS} = \frac{1}{U} \sum_{u=1}^{U} \mathrm{UQS}(u). \tag{3}$$

Consequently, the 95 % confidence intervals were calculated according to [1] as:

$$\mathrm{CI} = [\mathrm{MOS} + \delta; \mathrm{MOS} - \delta] \tag{4}$$

where

$$\delta = 1.96 \frac{\sigma}{\sqrt{U}}, \tag{5}$$

U is the number of samples (i.e. the number of observers providing their scores for the sequence under test) and σ is the standard deviation of the collected scores for the sequence. The standard deviation for each sequence is given as [1]:

$$\sigma = \sqrt{\sum_{u=1}^{U} \frac{(\mathrm{MOS} - \mathrm{UQS}(u))^2}{(U-1)}} \tag{6}$$

where MOS is the mean value of the scores collected for the sequence being analyzed while $\mathrm{UQS}(u)$ is the score given by observer u as defined in (2).

3.3 Evaluation During and After each Sequence

The following subsection provides a brief description of two objective video quality metrics - the peak signal-to-noise ratio (PSNR) and the video quality metric defined by the National Telecommunications and Information administration (NTIA) – VQM [8]. The former is a well known, easy to implement and massively used video metric while the latter is an example of a sophisticated comparative metric providing higher correlation with user scores [19].

The peak signal-to-noise ratio for the luma component of each frame in a video sequence is calculated as [19]

$$PSNR = 10 \cdot \log \frac{m^2}{\mathrm{MSE}} \tag{7}$$

where m is the maximum possible luma value of a pixel (255 for 8-bit samples) and MSE is the mean squared error, computed as

$$\mathrm{MSE} = \frac{1}{\mathrm{M} \cdot \mathrm{N}} \sum_{i=1}^{\mathrm{M}} \sum_{j=1}^{\mathrm{N}} \left[I(i,j) - \tilde{I}(i,j) \right]. \tag{8}$$

The constants M, N represent the dimensions of each frame in pixels and the I and $\tilde{I}$ values are the luma samples of the degraded and reference video frames at the position (i, j). In this work, to represent the PSNR of a whole video sequence, we simply calculate the mean over all frames. Note that we are using a reference-based (comparative) quality metric for the objective measurements. In case the video processing involved caused no severe luma offset, spatial offset or temporal offset, we can expect reasonable correlation of the single stimulus user ratings and comparative objective measurement.

The VQM is a rather complex quality metric. It involves preprocessing of the input signals to assure correct spatio-temporal alignment and a thorough analysis of video sequence properties. The core of the metric doesn't work with video frames directly, but breaks the sequence into several spatial and temporal sub-regions, which are processed at once. A more detailed description of the metric is beyond the scope of this paper and can be found in e.g. [8]. The metric is included in the ITU Recommendation ITU-R Rec. BT.1683 [2] and show very good performance for high definition video content [8].

Fig. 6 displays the dependency between the overall MOS for each video sequence according to (3) and the average PSNR of the luma component. The MOS 95 % confidence intervals given by (4) are represented by vertical error bars. It is obvious from the plot that the correlation between MOS and PSNR is very poor – the Pearson correlation coefficient is only 0.25. The usual values of correlation between PSNR and subjective user scores found in literature are between 0.7 and 0.8 [19].

Now, let us replace the PSNR, whose performance is often criticized, for a more complex and more accurate objective video quality metric - the NTIA VQM. The scatter plot diagram showing the dependency between MOS and VQM results is shown in Fig. 7. VQM produces values between 0 and 1, with 0 representing the highest possible quality and 1 representing the worst quality. The descending trend line in the plot is thus expected. The achieved Pearson correlation coefficient in this case is -0.51. This result is much better than for PSNR, but still worse than expected.

There are two possible explanations to the obtained results. Firstly, putting the results of the comparative metric in correspondence with a single stimulus user-based rating may be a too much generalizing approach. Secondly, the user ratings may be biased when performing continuous quality evaluation and the users reflect the relative quality with respect to previous time instants rather than being precise on the absolute scale. To address these findings, another was performed.

In the following, the setup of the experiment is different. The user scores are no longer recorded during the video sequences. Instead, a pause is inserted between consequent video sequences, where only gray image is displayed. In this pause, the users are asked to provide one rating for the whole sequence using the slider. The final slider position is recorded.

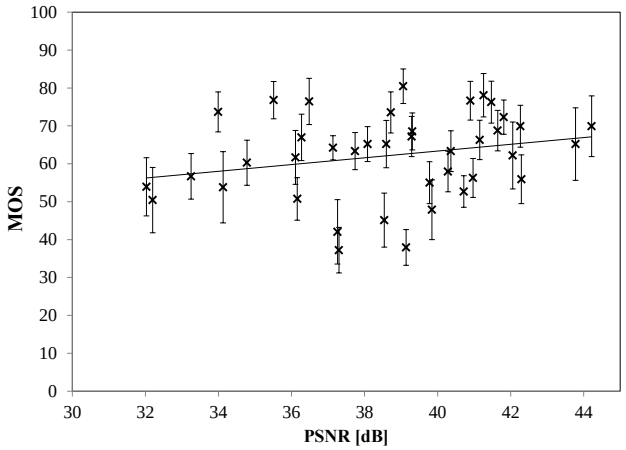

Fig. 6. Scatter plot graph showing the dependency between sequence PSNR and average user score (DMOS) captured at the end of continuous evaluation.

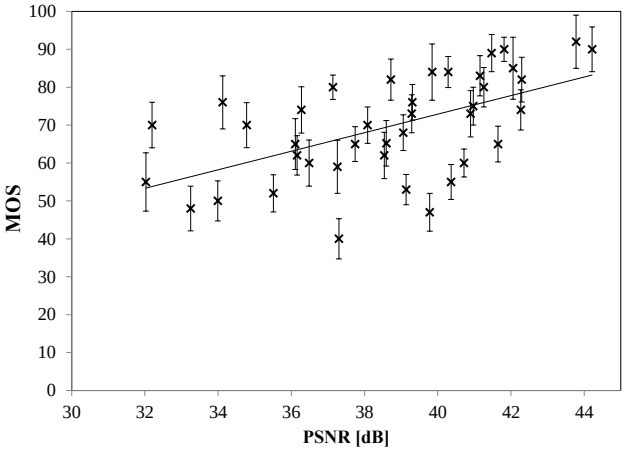

Fig. 8. Scatter plot graph showing the dependency between sequence PSNR and mean opinion score recorded in a gray-image interval after each sequence.

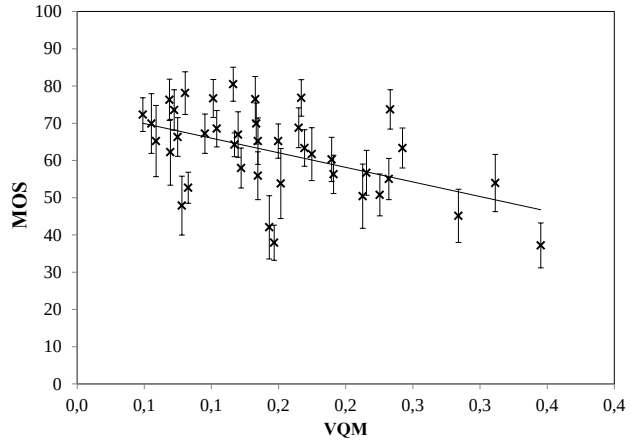

Fig. 7. Scatter plot graph showing the dependency between sequence VQM and mean opinion score captured at the end of continuous evaluation.

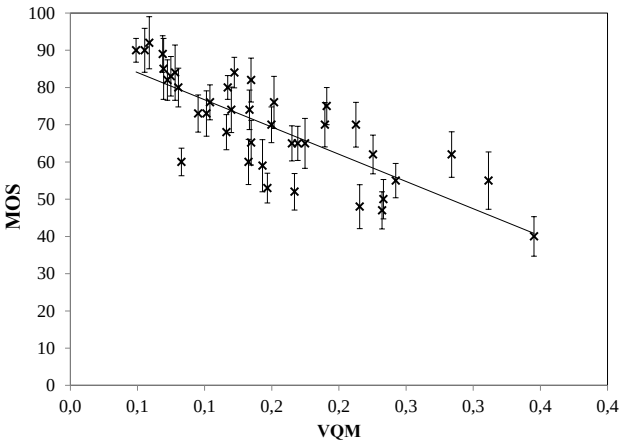

Fig. 9. Scatter plot graph showing the dependency between sequence VQM and mean opinion score recorded in a gray-image interval after each sequence.

The resulting scatter plot diagrams for MOS vs. PSNR and MOS vs. VQM are shown in Figs. 8 and 9. In this setup, we can observe that the correlation is higher for both the PSNR (0.56) and VQM (-0.79). Furthermore, the 95 % confidence intervals are lower for most sequences, which is a consequence of lower standard deviation within the samples. The significant improvement with the change of the experiment and recording user ratings *after* each short video clip proves that the user ratings are likely to be biased when evaluating the video quality in a continuous test. The continuous quality evaluation can be used to detect quality changes in a video sequence rather than gathering absolute quality ratings for different parts of a video presentation. Even though we applied full reference metrics to the video sequences and compared them with the results of single stimulus subjective tests, we succeeded in reaching correlation of the subjective and objective scores close to 0.8 (in the absolute scale).

4. Conclusions

We have shown that when performing subjective quality tests of full HD video sequences, there is a significant temporal impact on the recorded scores that has to be taken into account. The user typically needs one second to start interacting and then several seconds to adjust the desired score.

The consequence of such result is two-fold. Firstly, we have shown that for the continuous quality test sessions such as those described by the SSCQE method in BT.500, the prescribed score scanning interval of 500 miliseconds is sufficient. Secondly, we have shown that the user is unable to instantaneously react to the change of the perceived quality and the delay in which the corresponding value is obtained is in the order of seconds. This fact has to be taken into account when using continuous quality tests as a benchmark of

objective quality evaluation algorithms – full correlation of the objective and subjective scores for a given time instant can hardly be reached.

Furthermore, we have studied the correlation of mean opinion scores calculated from user ratings collected during and after playback of each sequence. We have found that the continuous ratings collected over time tend to be strongly biased. We have also shown that, with a limited accuracy, we can simulate single stimulus user ratings using full reference objective video quality metrics.

Acknowledgements

This work was supported by the Czech science foundation under project number P102/10/1320 and by the Czech Ministry of Education under grant number LD11081. The research published in this submission was financially supported by the project CZ.1.07/2.3.00/20.0007 WICOMT of the operational program Education for competitiveness. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 230126. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. Finally, the authors would like to thank CET 21, s.r.o. for providing the video sequences used during the tests.

References

[1] ITU-R Recommendation BT.500-11. *Methodology for the Subjective Assessment of the Quality of Television Pictures.* Geneva: ITU, 2002.

[2] ITU-R Recommendation BT.1683 *Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference*. Geneva: ITU, 2004.

[3] ITU-T Recommendation P.910. *Subjective Video Quality Assessment Methods for Multimedia Applications.* Geneva: ITU, 2008.

[4] STAELENS, N., MOENS, S., VAN DEN BROECK, W., MARIEN, I., VERMEULEN, B., LAMBERT, P., VAN DE WALLE, R., DE-MEESTER, P. Assessing Quality of Experience of IPTV and video on demand services in real-life environments. *IEEE Transactions on Broadcasting*, 2010, vol. 56, no. 4, p. 458 - 466.

[5] SLANINA, M., ŘÍČNÝ, V. Estimating PSNR in high definition H.264/AVC video sequences using artificial neural networks. *Radioengineering*, 2008, vol. 17, no. 3, p. 103 - 108.

[6] RIES, M., GARDLO, B. Audiovisual quality estimation for mobile video services. *IEEE Journal on Selected Areas in Communications*, 2010, vol. 28, no. 3, p. 501 - 509.

[7] RIES, M., SLANINA, M., GARCIA, D. M. Reference free SSIM estimation for Full HD video content. In *Proceedings of the 21st International Conference Radioelektronika 2011.* Brno (Czech Republic), 2011.

[8] WOLF, S., PINSON, M. Application of the NTIA general video quality metric (VQM) to HDTV quality monitoring. In *Proc. of The Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM).* Scottsdale (AZ, USA), 2007.

[9] GARCIA, M. N., RAAKE, A. Parametric packet-layer video quality model for IPTV. In *10th International Conference on Information Science, Signal Processing and their Applications ISSPA 2010.* Kuala Lumpur (Malaysia), 2010, p. 349 - 352.

[10] MASRY, M., HEMAMI, S. CVQE: A metric for continuous video quality evaluation at low bit rates. *SPIE Human Vision and Electronic Imaging*, 2003, vol. 5007, p. 116 - 127.

[11] *VideoLAN - VLC: Official Site*. [Online] Cited 2011-05-12. Available at: http://www.videolan.org

[12] SAKAMOTO, K., AOYAMA, S., ASAHARA, S., YAMASHITA, K., OKADA, A. Evaluation of viewing distance vs. TV size on visual fatigue in a home viewing environment. In *Digest of Technical Papers, International Conference on Consumer Electronics.* Las Vegas (USA), 2009.

[13] ISO/IEC 13818-2:2000. *Generic Coding of Moving Pictures and Associated Audio Information: Video.* ISO, 2000.

[14] ISO/IEC 14496-10:2005. *Information Technology - Coding of Audio-Visual Objects.* ISO, 2005.

[15] SMPTE 421M. *VC-1 Compressed Video Bitstream Format and Decoding Process.* SMPTE, 2006.

[16] *FFmpeg.* [Online] Cited 2011-05-12. Available at: http://www.ffmpeg.org

[17] *VideoLAN - x264, the Best H.264/AVC Encoder.* [Online] Cited 2011-05-12. Available at: http://www.videolan.org/developers/x264.html

[18] *Huynh-Thu, Q., Garcia, M.-N., Speranza, F., Corriveau, P., Raake, A.* Study of rating scales for subjective quality assessment of high-definition video. *IEEE Transactions on Broadcasting*, 2011, vol. 51, no. 1.

[19] KRATOCHVIL, T., SLANINA, M. Digital video image quality. *Digital Video*. Intech (Croatia), 2010.

[20] SLANINA, M., KRATOCHVIL, T., POLAK, L., RICNY, V. Temporal aspects of scoring in the user based quality evaluation of HD video. In *34th International Conference on Telecommunications and Signal Processing TSP 2011.* Budapest (Hungary), 2011, p. 598 - 601.

[21] KLIMA, M. et al. DEIMOS - an open source image database. *Radioengineering*, 2011, vol. 20, no. 4, p. 1016 - 1023.