XVIII IMEKO WORLD CONGRESS
Metrology for a Sustainable Development
September, 17 – 22, 2006, Rio de Janeiro, Brazil

METROLOGY PROPERTIES OF HUMAN OBSERVER IN COMPRESSED VIDEO QUALITY EVALUATION

Anna Ostaszewska [1] *Sabina Żebrowska-Łucyk* [2], *Rafał Kłoda* [3]

[1] Warsaw University of Technology, Warsaw, Poland, a.ostaszewska@mchtr.pw.edu.pl
[2] Warsaw University of Technology, Warsaw, Poland, szl@mchtr.pw.edu.pl
[3] Warsaw University of Technology, Warsaw, Poland, r.kloda@mchtr.pw.edu.pl

Abstract: The main purpose of this paper is to present the method of subjective quality evaluation recommended by International Telecommunication Union and the results obtained. Authors propose a bunch of factors describing an individual viewer.

Keywords: Single Stimulus Continuous Evaluation, human audience, subjective quality evaluation

1. INTRODUCTION

In each area of life everybody struggle for the best quality. Hence quality itself is in general difficult to evaluate, as a lot of factors have an influence on it. In most cases it is impossible to work out one parameter that would take into account all factors. The quality of compressed video is the best example, where the additional obstacle is the lack of sufficient knowledge of human perception. That's why subjective methods of assessing still play a great role and there is an urgent need to develop appropriate methods and tools that enable to infer from researches.

The challenge of normalization of those methods was taken up by ITU-T (Telecommunication Standardization Sector of International Telecommunication Union), which in P.911 [1] recommendation specifies four methods of subjective quality evaluation: Absolute Category Rating (ACR), Degradation Category Rating (DCR), Pair Comparison (PC) and Single Stimulus Continuous Evaluation (SSCQE).

2. SSCQE METHOD AND MEASUREMENT STATION

In SSCQE method a series of video sequences is presented once to a viewer. The video sequences may or may not contain impairments. Subjects evaluate the instantaneous quality in real time using a slider with a continuous scale. The SSCQE method yields ratings at regular time intervals and can thus capture the perceived variations in quality as a function of time. The ratings are absolute in the sense that viewers are not explicitly shown the reference sequences. This corresponds well to an actual home viewing situation, where the reference is not available to the viewer either.

In order to carry out researches the authors decided to accept ITU-T P.911 Recommendation [1]. 15 non-expert viewers participated in test sessions.

To prevent fatigue issues, the test was limited to two 15-minute sessions (separated by a 15-minute break). A training process was used to instruct subjects about the task they were to perform during the subjective test and to bound the quality range to be seen in the test.

The source sequences were obtained from Tektronix[1], each of 15 seconds duration, 40 Mbps and 25 fps.

Source sequences were processed to the most popular standard MPEG-2. Taking into account only 13-frame GOP, changing the bit stream (from 2 Mbps to 5 Mbps) and the number of B-frames (from 0 to 2), 30 variants of each sequence were coded.

Subjects were evaluating test sequences by changing the position of a slider (Fig. 1). In this research a slider was a stand-alone hardware device with a hundred-level scale attached (0 stands for the worse quality and 100 - the best). We reset the slider to the middle position at the beginning of each SSCQE session.

SSCQE ratings were sampled at a rate of 2 samples per second and entering the PC directly through the NI 6013 card. Thanks to the software prepared with LabVIEW tools, there was a possibility of displaying all incoming signals, so all data was being monitored by the researcher. All data was synchronized to the timecode for the duration of the test.

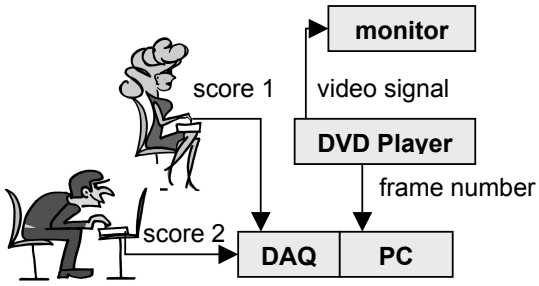

Fig. 1. Diagram of the experimental station used in the SSCQE research

1 ftp://ftp.tek.com/tv/test/streams/Element/MPEG-Video/625/

3. DATA PROCESSING AND RESULTS

The main goals established by authors were to examine: the performance of the SSCQE method across a range of content types and parameters of coding, the stability of human audience in time and properties of individual viewers.

The first step in data analysis was to discard the first six seconds of each clip to provide a period of time for the average viewer response data to stabilize.

In order to check if the method gives results commensurate with those obtained in other experiments [2, 3, 4] the median of all scores given for each sample was calculated.

Additionally to assess the time stability of human audience, authors calculated the standard deviation of scores given in time. The curve revealed no descending or ascending trend. This may lead to the conclusion that the audience (the average subject) did not change its metrology features due to the fatigue.

To investigate individual viewers, for each of them the average, the skeweness, the kurtosis, the standard deviation, the correlation of his/her score with bit rate and the spectrum of scores given in time were calculated.

The averages of scores given by the individual viewer revealed that the upper half of the scale was preferred. All but two individual averages place themselves between 50 and 70. The skeweness was negative for all viewers and so was the kurtosis (platykurtosis). The standard deviation represented differentiation of scores for individual viewer. All subjects had standard deviation between 13 and 33.

While the factors mentioned above would describe the average viewer, authors decided to work out some parameter that would be a significant factor of reliability of an individual viewer. This role would play the correlation of subject's scores with bit rate, which according to the results obtained by ACR and DCR methods is linear for the bit rate in a range of 2 - 4 Mbps. For this range of bit rate correlation was calculated and it appeared that scores given by some observers are increasing with bit rate while the others are not. What is more for 75% of viewers' correlation coefficient is between 0,3 and 0,55. The same time four of viewers had correlation coefficient less than 0,2 and one had 0,77.

In order to find some other differences between observers, the spectrum of individual scores gathered in 10 minutes was plotted. This 10-minute test material consisted of 10 parts of the same content coded with different bitrate. It appeared that those observers, whose scores had weak correlation with bit rate, produced spectrum of scores with harmonics corresponding to the content of the test material. The best correlation with bit rate was observed for the viewers with the lack of any dominant harmonic in the score spectrum.

The other concept of evaluating viewers' accuracy would be to extract individual recordings for the same material coded with the same bit rate, but with different GOP structure (with one, two or without B-frames). Authors assumed that those three curves should be of the same shape and eventually have different constant component. In order to get rid of constant component, each curve was shifted down (an average score for certain coding parameters was subtracted from each point). Then the standard deviation for those three curves was calculated. It became clear that the observer whose scores were correlated with bit rate performed better conformity of those three curves in comparison with the one, whose correlation was really low (Fig. 2).

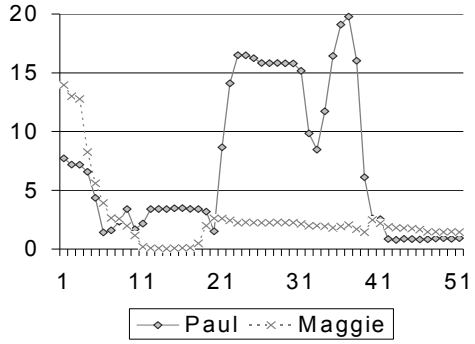

Fig. 2. Standard deviation of scores given by Maggie (the highest correlation coefficient) and by Paul (lowest correlation coefficient) for the three observations of the same sequence

4. CONCLUSIONS

Although described researches were only the preliminary ones, they tend to give future prospects. They seem to enable the assessment of a viewer as a "metrology tool". This would help to decrease the number of viewers and to get more adequate results. Besides the SSCQE recordings can be the source of the knowledge on HVS, on preferences of target audience or even might state a base of psychometric diagnosis of a subject.

The knowledge of viewers' reaction would play a key role in preparing test materials and new systems of quality evaluation, both subjective and objective ones.

REFERENCES

[1] ITU-T Recommendation P.911 (1996), Subjective audiovisual quality assessment methods for multimedia applications

[2] Masry A., Hemami S.: A metric for continuous quality evaluation of compressed video with severe distortions; Signal Processing: Image Communication, 19 (2004) 133-146

[3] Ostaszewska A.: Research on compression influence on the quality of video, MS thesis Faculty of Warsaw University of Technology, 2004

[4] Żebrowska-Łucyk S., Kłoda R., Ostaszewska A.: Metrology tools in subjective quality evaluation, Mechatronics, Conference "Robotics and Biomechanics", Trest, Czech Republic, September 2005