Video Quality Evaluation for Mobile Applications

Stefan Winkler[a] and Frédéric Dufaux[b]

[a]Audiovisual Communications Laboratory and [b]Signal Processing Laboratory
Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland

ABSTRACT

This paper presents the results of a quality evaluation of video sequences encoded for and transmitted over a wireless channel. We selected content, codecs, bitrates and bit error patterns representative of mobile applications, focusing on the MPEG-4 and Motion JPEG2000 coding standards. We carried out subjective experiments using the Single Stimulus Continuous Quality Evaluation (SSCQE) method on this test material. We analyze the subjective data and use them to compare codec performance as well as the effects of transmission errors on visual quality. Finally, we use the subjective ratings to validate the prediction performance of a real-time non-reference quality metric.

Keywords: Quality assessment, SSCQE, MPEG-4, Motion JPEG2000, wireless networks, WCDMA

1. INTRODUCTION

As a result of the emergence of broadband wireless networks like third generation mobile telecommunication systems (3G) or WLAN based on IEEE 802.11b (WiFi), combined with a plurality of high performance mobile devices such as laptops, PDA's and cell phones, the transmission of video and images in mobile applications is now becoming a reality.

This paper addresses the problem of evaluating the quality of video sequences encoded for and transmitted over a wireless channel. Quality assessment for television applications has been the subject of extensive work, for instance by the Video Quality Experts Group (VQEG).[11, 13] More recently, we presented results for the evaluation of video quality for Internet streaming applications.[16] However, little work has been carried out so far to address the domain of mobile applications.

Video transmission over wireless is characterized by a specific set of requirements that include low bitrates, small frame sizes, and low frame rates. Furthermore, wireless networks are error-prone. Therefore, the video sequences to be assessed exhibit artifacts resulting not only from compression, but also from transmission errors. Finally, the content is viewed at short distance on a small LCD screen with a progressive display.

In this paper, we describe the test environment for simulating the transmission of video over a WCDMA channel, as it is used for 3G or wireless LAN. We selected source material covering a wide and representative set of content, along with the appropriate compression parameters. The source sequences were encoded using two coding standards well-suited for mobile applications, namely MPEG-4[3] and Motion JPEG2000.[12] We then simulated the transmission errors of a WCDMA channel using representative bit error patterns.

Subjective ratings were obtained for the resulting test sequences using the Single Stimulus Continuous Quality Evaluation (SSCQE) methodology as defined by ITU-R Recommendation BT.500,[5] which permits viewers to rate the time-varying quality of the sequences. We analyze the results of these experiments with respect to inter-subject variability. Furthermore, the ratings are used to compare the performance of the two codecs and to investigate the effects of bit errors on perceived quality. Finally, we combine three existing non-reference metrics for blockiness, blurriness and jerkiness to compute predictions of perceived quality. We show that these predictions can be successfully tuned and evaluated using the subjective ratings obtained.

The paper is organized as follows. Section 2 describes the source material, the simulation environment and the test conditions used to produce the test sequences. In Section 3 we discuss the subjective assessment method, the presentation of the sequences and the viewing conditions. The data obtained in the subjective experiments are analyzed in Section 4. Finally, we evaluate the predictions of a non-reference quality metric in Section 5.

E-mail of corresponding author: stefan.winkler@epfl.ch

This work was done in part while the authors were with Genista Corporation, Tokyo, Japan.

2. TEST MATERIAL

This section discusses the methodology to produce the video sequences to be assessed in the subjective tests. We will first describe the source sequences. We will then introduce the environment to simulate the transmission of video sequences over an error-prone wireless channel. Finally, we will present the source encoders and the parameters used to generate the final test material.

2.1. Source Sequences

The source sequences were chosen to cover a wide range of typical content for mobile applications, such as news, sports and music video clips. Furthermore, they were selected to contain various characteristics such as flat areas, complex textures, object and camera motion, faces and landscapes. Consequently, the scenes span a wide range of coding complexity.

We generated two test sequences by concatenating scenes taken mostly from clips used in previous tests by MPEG[2] and VQEG.[13] Each sequence has a duration of 1 minute. A detailed description of the selected scenes and their compilation is given in Tables 1 and 2. If necessary, these clips were cropped and rescaled to the same frame size.

Table 1. Description of sequence 1.

Scene #	**Name**	**Description**
A	Letters	Letters with different colors flying in all directions over dark background
B	News	Male and female speaker in newsroom, almost still
C	F1 car	Object motion, camera following car, 2 angles
D	Fast food	Texture, people, fast pans, 2 angles
E	Coastguard	Two boats crossing on river, medium motion, water motion
F	Balloons	Amusement park, saturated colors, people, motion

Table 2. Description of sequence 2.

Scene #	**Name**	**Description**
G	Foreman	Talking head, with pan to construction site, geometric shapes
H	New York	Slow city flyover, skyscrapers at sunset, detailed texture
I	Football	Fast camera and object motion, colors
J	Live concert	Dark scene, spotlights, 3 angles
K	Cartoon	Characters dancing through scene, with pan

2.2. Simulation Environment

The purpose of the experimental setup is to simulate the transmission of video sequences over a WCDMA wireless channel. The simulation environment is illustrated in Figure 1. The video source is first compressed by the source encoder. For the tests described in this paper, we have selected two of the most performant and well-suited coding standards for mobile applications, the well-known MPEG-4[3] and the recent Motion JPEG2000.[12] MPEG-4 utilizes block-based DCT with motion compensation, while Motion JPEG2000 is based on an intra-frame wavelet transform. Both MPEG-4 and Motion JPEG2000 include a number of tools to improve their resilience to transmission errors.

By exploiting inter-frame redundancy, MPEG-4 has a higher coding efficiency at the cost of a higher complexity. The dependencies between coded frames and the resulting propagation of errors across consecutive frames also implies a lower error resilience. Conversely, Motion JPEG2000, which is based on intra-frame coding, has a lower coding efficiency at the benefit of a reduced complexity. Additionally, it is more resilient to transmission errors, because consecutive frames are coded independently.

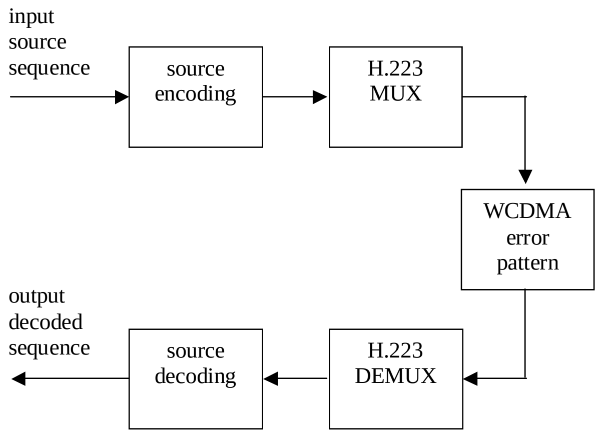


Figure 1. Simulation environment for error-prone WCDMA wireless channel.

After source encoding, H.223[6] is applied to the bitstream for multiplexing, packetization and cyclic redundancy check (CRC). For this task, we use the UCLA/Samsung H.223 Multiplex Simulator.* Transmission errors are simulated using bit error patterns representative of WCDMA.[8] Note that while this setup is a simplification over implementing the complete WCDMA protocol stack and air-interface, this methodology is similar to the one used by 3GPP.[1]

Due to the random nature of transmission errors, the experimental setup results in a statistical process. Consequently, the output video sequence is obtained by running a large number of trials and selecting a representative result. The different trials consist of applying random circular shifts to the same bit error pattern.

Eventually, the output video sequences of this simulation environment exhibit artifacts resulting from both source encoding and transmission errors.

2.3. Test Conditions and Parameters

The encoding parameters were chosen so as to cover a typical range of video qualities for mobile applications. Specifically, we encoded sequences at three bitrates: 64 kb/s, 128 kb/s and 384 kb/s. We used the MoMuSys reference software implementation[10,†] for MPEG-4, and the Kakadu software[‡] for Motion JPEG2000. The sequences were downsampled spatially and temporally as specified in the table below in order to accommodate the low bitrates.

At the transmission stage, we considered the two cases with and without transmission errors. For the case with errors, we selected two distinct error patterns, denoted (I) and (II), with a Bit Error Rate (BER) of 10^{-4}. We found this BER to yield interesting test sequences in the sense that it introduced several visible distortions without completely destroying the video. Contrary to what was observed in the tests for Internet streaming,[16] where transmission errors manifest themselves as packet losses, the bit errors did not lead to dropped frames or delays in the decoded video. Nevertheless, transmission errors are only included for Motion JPEG2000, because the MPEG-4 reference software was not able to recover from them.

The test conditions are summarized in Table 3. Processing the two source sequences with each of the 12 test conditions thus yields a total of 24 minutes of test material.

Figure 2 shows two examples of frames that exhibit both compression and (very severe) transmission error artifacts. Specifically, the two frames result from conditions 3 and 11.

* Available for download from http://www.icsl.ucla.edu/~wireless/.
† Available for download from http://megaera.ee.nctu.edu.tw/mpeg/.
‡ Available for download from http://www.kakadusoftware.com/.

Table 3. Test conditions (MJ2 = Motion JPEG2000, MP4 = MPEG-4).

Condition #	Format	Frame rate	Bitrate	Codec	Bit Error Rate
1	QCIF	4 fps	64 kb/s	MP4	—
2	QCIF	4 fps	64 kb/s	MJ2	—
3	QCIF	4 fps	64 kb/s	MJ2	10^{-4} (I)
4	QCIF	4 fps	64 kb/s	MJ2	10^{-4} (II)
5	QCIF	6 fps	128 kb/s	MP4	—
6	QCIF	6 fps	128 kb/s	MJ2	—
7	QCIF	6 fps	128 kb/s	MJ2	10^{-4} (I)
8	QCIF	6 fps	128 kb/s	MJ2	10^{-4} (II)
9	CIF	8 fps	384 kb/s	MP4	—
10	CIF	8 fps	384 kb/s	MJ2	—
11	CIF	8 fps	384 kb/s	MJ2	10^{-4} (I)
12	CIF	8 fps	384 kb/s	MJ2	10^{-4} (II)

(a) Scene C, condition 3 (b) Scene D, condition 11

Figure 2. Example frames from the test sequences with extreme cases of transmission error artifacts.

3. SUBJECTIVE ASSESSMENT

3.1. Assessment Method

Subjective assessment was based on ITU-T Recommendation P.910[7] and ITU-R Recommendation BT.500.[5] We used Single Stimulus Continuous Quality Evaluation (SSCQE), which is specified in ITU-R Rec. BT.500, as the assessment method for our subjective experiments. In an SSCQE session, a series of video sequences is presented to the 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 quality ratings at regular time intervals and can thus capture the perceived time variations in quality. 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 also not available to the viewer.

Slight modifications of the procedure described in the ITU recommendations were introduced to adapt them to purely PC-based testing.[16] The slider for the SSCQE test was not a stand-alone hardware device, but a graphical on-screen slider that was steered by moving the mouse up and down, i.e. vertical mouse movements were translated directly into slider shifts. We found this to give viewers a good haptic feeling of where they were on the quality scale. People's familiarity with handling a computer mouse is an additional advantage.

We decided not to attach the usual five-level scale of semantic judgment terms ("excellent", "good", "fair", "poor", "bad") on the side of the slider for two reasons: First, none of our videos could be considered "excellent" quality, given that the quality reference for non-experts today is the DVD. Second, studies found that these quality terms may lead to a nonlinear interpretation of the scale by the subjects, i.e. "excellent" and "good" may be considered closer than "poor" and "bad", for example.[14] Therefore, we only put "Good" and "Bad" at the top and bottom end of the slider for general directional guidance.

Furthermore, we decided to make the slider a bright green rectangle ranging from the bottom of the scale to the current slider position, while the rest of the slider was black. In initial tests we found this representation easier to follow from the corner of the eye than a plain gray slider, thereby allowing viewers to check the approximate slider position without having to look away from the video.

In summary, we found the on-screen visual feedback of the slider position in combination with the haptic mouse feedback to be very user-friendly. Another advantage of a software slider is that it can be reset automatically without having to instruct subjects to do so. We reset the slider to the middle position at the beginning of each SSCQE session. During the experiments, the slider position was recorded every 50 ms, on an integer scale from 0 ("bad") to 100 ("good").

3.2. Presentation Structure

Instructions were given to the viewers in written form. After they had read the instructions, a training session was run to demonstrate the task that subjects had to perform as well as the range of quality to be expected.

SSCQE demands constant attention and concentration from the subjects, and we felt that a break would help reduce fatigue. Therefore, subjects were given a short break at half-time, after 12 minutes of SSCQE testing. Including training, the duration of a session was approximately 30 minutes in total. In order to minimize contextual effects, the order of the test sequences was randomized at the clip level such that every subject viewed the test clips in a different order.

3.3. Viewing Conditions

Viewing conditions comply as much as possible with those described in ITU-R Rec. BT.500[5] and ITU-T Rec. P.910,[7] with the necessary modifications of the laboratory setup according to typical user requirements and conditions for the display of video for mobile applications.

Video on a mobile handset is typically viewed by a single person only; this was also the case during our experiments. For our test material, we found subjects to be comfortable at a viewing distance of 3-4 times the height of the video picture, which corresponds to about 30-40 cm in our setup.

Since mobile devices typically have an LCD screen, the monitors used in the subjective assessments are also LCD screens. The specific screen used, a 15" Sony SDM-S51, has the following specifications:

Resolution:	1024×768
Dot pitch:	0.297 mm
Peak luminance:	250 cd/m^2
Contrast ratio:	300:1
Viewing angles:	120° horizontal, 90° vertical
Response times:	10 ms (rise time), 20 ms (fall time)

After calibration and black-level adjustment, the screen properties were measured to be as follows:

Gamma:	2.2
Color temperature:	6400 K
White luminance:	77 cd/m^2
Video surround:	20 cd/m^2

3.4. Viewers

21 non-expert viewers – mostly university students – participated in the test. Prior to the test session, each viewer was screened for the following:

- Normal (20/20) visual acuity or corrective glasses;
- Normal color vision (per Ishihara test);
- Sufficient familiarity with the language to comprehend the instructions.

4. SUBJECTIVE DATA ANALYSIS

4.1. Data Preprocessing

The validity of the subjective test results was verified by screening the observers according to Annex 2 of ITU-R Rec. BT.500. The raw ratings obtained every 50 ms were subsampled onto 500 ms intervals. Subsequently, the Mean Opinion Scores (MOS) and the 95% confidence intervals of the subjective ratings were computed. The first three seconds of data of every test sequence were discarded to remove the influence of large quality changes from one test condition to the next.

Viewer reaction times and slider “stiffness” result in a delay between the display of a video frame and the corresponding slider response from the subject. For comparisons with the video time line as well as PSNR and other metrics (which do not exhibit such a latency), MOS and video/metric data must therefore be time-aligned. This was achieved by computing and applying one global time shift between video and MOS data, which was found to be around 1.5 seconds for our test. The SSCQE scores were thus time-shifted by −1.5 seconds (this is to the left in the plots). Evidence for this time-shift comes from the mappings discussed below in and from previous findings.[4]

4.2. Inter-Observer Agreement

As a quality indicator of the subjective data, the distribution of the 95% confidence intervals is shown in Figure 3. The average size of the confidence intervals is ±7.8 on the 0-100 scale. This indicates a good agreement between observers. For comparison, it was ±8.5 in the Internet streaming experiments,[16] which used the same source material.

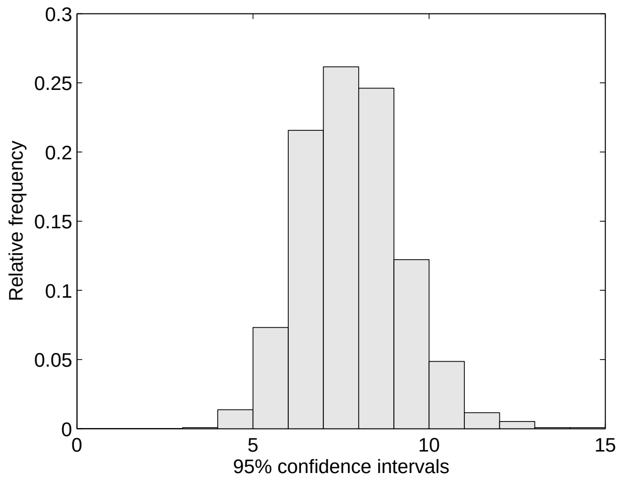


Figure 3. Distribution of 95% confidence intervals.

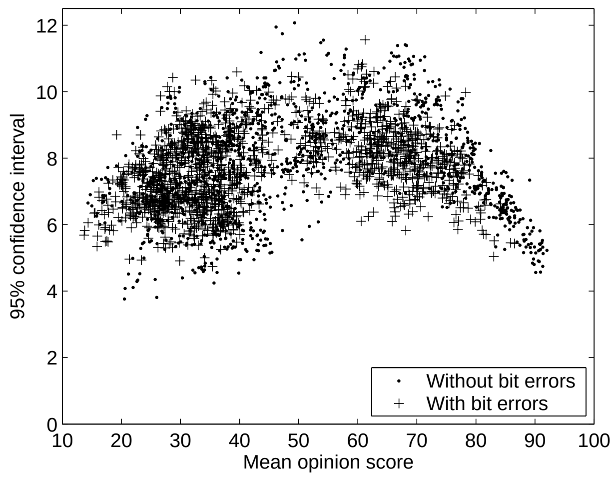


Figure 4. Confidence interval size vs. MOS.

It also is interesting to study the relationship between confidence interval size and MOS, as shown in Figure 4. It is obvious that the agreement between observers is highest at both ends of the scale (especially for the high-quality sequences), whereas the largest confidence intervals occur in the “medium-quality” regime between 40 and 70 on the SSCQE scale.

We expected the confidence intervals to be larger for the test sequences with transmission errors. The bit error artifacts are quite visible, but highly transient, because their effects are usually limited to a single frame at a time. We considered it difficult to respond to these effects in a reliable fashion. Nonetheless, we found no clear evidence of larger inter-subject variation for these sequences in the data (see below for more discussion on transmission error effects).

4.3. Codec Comparison

The data allow us to compare the performance of the two codecs used in the tests. The test conditions without bit errors are compared in Figure 5, where the MOS differences between MPEG-4 and Motion JPEG2000 conditions are shown for the different bitrates.

Overall, the MPEG-4 codec we used has the advantage over the Motion JPEG2000 codec. This is especially true for the lower bitrate sequences at 64 and 128 kb/s, where the motion prediction of MPEG-4 obviously helps to maintain a generally higher quality. The scene-dependence of codec performance is even more evident. MPEG-4 typically outperforms Motion JPEG2000 on scenes with slow to moderate motion (e.g. scenes B, G, K). As soon as there is fast motion, especially camera movement, as for example the continuous high-speed pan in scene D, motion prediction does no longer help in compression. In this case, the Motion JPEG2000 codec has a performance comparable with or even better than the MPEG-4 codec.

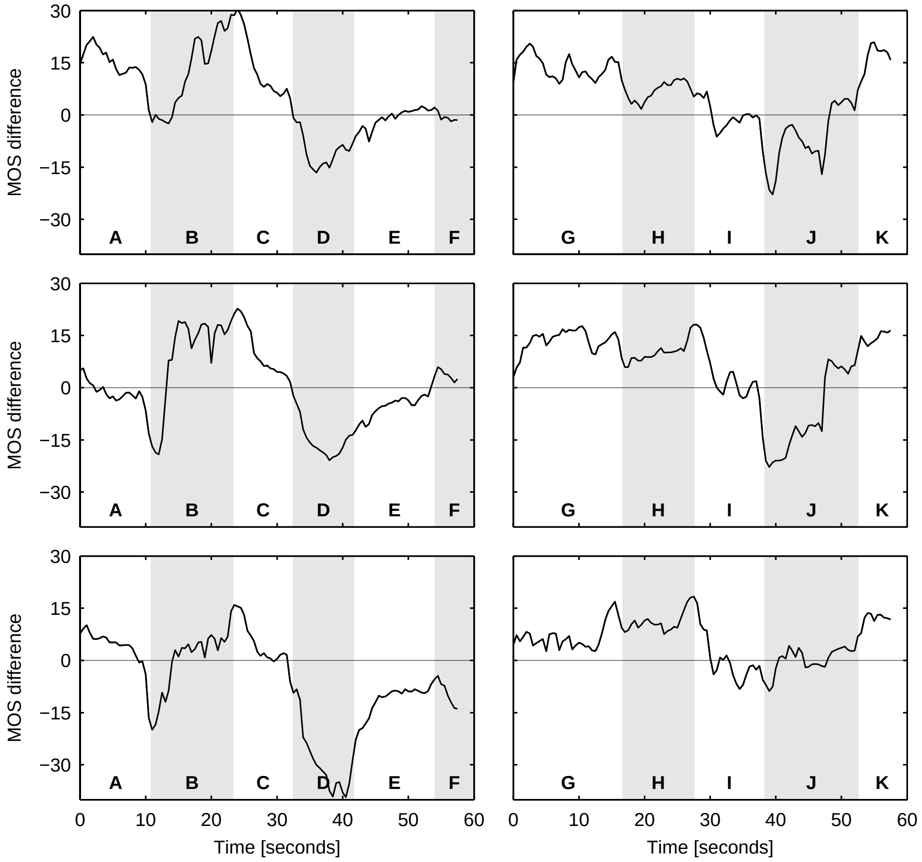
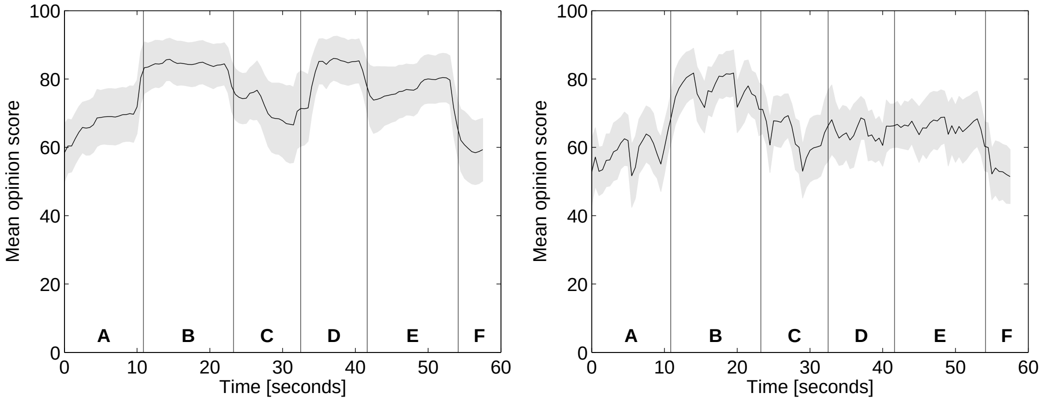


Figure 5. Codec comparison. The MOS differences between MPEG-4 and Motion JPEG2000 conditions are plotted for the two test sequences (left and right column) and the three bitrates 64 kb/s, 128 kb/s and 384 kb/s (top to bottom). Only test conditions without transmission errors are shown, i.e. conditions $1-2$, $5-6$, $9-10$. Positive MOS differences thus indicate that MPEG-4 is better, and negative MOS differences indicate that Motion JPEG2000 is better. The letters indicate the scene numbers from Tables 1 and 2 inside each sequence.

4.4. Transmission Errors

The general difference in perceived quality between test conditions with and without transmission errors is demonstrated with a typical example in Figure 6. Two effects can be observed: one is a global MOS decrease over the entire duration of the sequence; the other and perhaps more interesting one is a more ragged response curve in the case with errors, which is due to observers responding to the clearly visible transmission error artifacts at various instants. Note again that – as mentioned above – confidence interval size does not increase.



(a) Sequence 1, condition 10

(b) Sequence 1, condition 11

Figure 6. Comparison of subjective ratings for a Motion JPEG2000 sequence encoded at 384 kb/s, with and without transmission errors. The gray bands around the MOS values indicate the 95% confidence intervals.

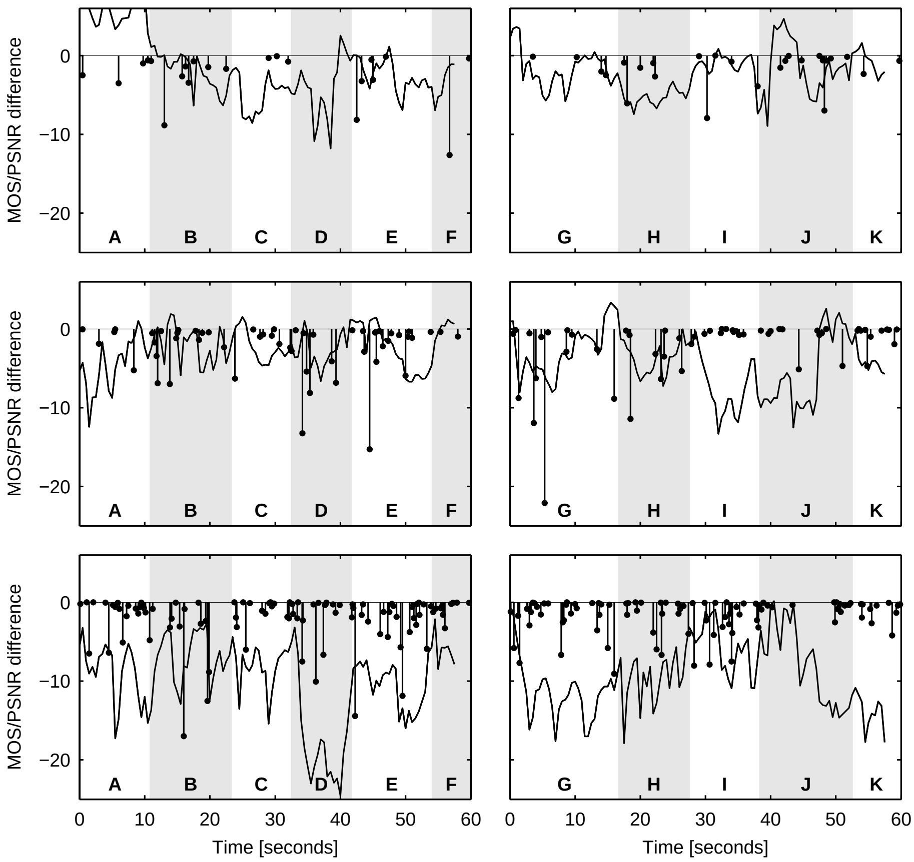


Figure 7. Bit error effects. The MOS differences (curves) between the conditions with and without transmission errors are plotted for the two test sequences (left and right column) and the three bitrates 64 kb/s, 128 kb/s and 384 kb/s (top to bottom). Only the results for bit error pattern I are shown, i.e. conditions $2-3$, $6-7$, $10-11$. Negative MOS differences thus indicate a decrease in quality due to transmission error effects. The PSNR differences (stems and dots, in dB) between these conditions are shown in the same plots. The letters indicate the scene numbers from Tables 1 and 2 inside each sequence.

To investigate the source of these effects, let us have look at the MOS differences between conditions with and without transmission errors, shown in Figure 7. The global MOS decrease is again evident for all sequences. The few cases where MOS actually goes up (e.g. scene A at 64 kb/s) can only be attributed to the limits of statistical significance of the data.

Additionally, the PSNR differences (stems and dots) between the same conditions are shown in the respective plots. As can be seen, only a subset of frames is affected by the bit errors; most frames are identical for the cases with and without transmission errors. Since the bit error rate is the same for all bitrates, there are proportionally more frames concerned as the bitrate increases. For our bitrates of 64, 128 and 384 kb/s, approximately 10%, 15% and 20% of frames are affected, respectively (the increase is not linear because it also becomes more likely for two errors to occur within the same frame).

While PSNR is certainly not a good predictor of the visual quality of these conditions (cf. next section), it can serve as a detector of clearly visible distortions. Unfortunately, the connection between severely distorted frames and viewer response peaks is not obvious in the plots. The bit error artifacts may occur too frequently to allow us to clearly establish such a relationship. It can be observed, however, that the perceived quality degradation increases with bitrate. This can be explained by the temporally higher concentration of erroneous frames at high bitrates. Another reason for this behavior is likely to be the generally bad quality at low bitrates, where the additional distortions due to transmission errors are considered relatively small compared to the compression artifacts already present in the video.

5. MOS PREDICTION

The data obtained in our experiments were used to tune and evaluate the MOS predictions of Genista's *Stream* $PQoS^{TM}$ software.* Its MOS predictions are based on existing non-reference metrics for blockiness,[15] blurriness[9] and jerkiness artifacts. These artifact metrics are computationally light, which makes it possible to compute them in real-time on a standard PC, in parallel to decoding and displaying the video.

Due to the different types of artifacts that are produced by the two codecs used in the tests, individual mappings were determined for each codec separately. For example, the MOS prediction for the MPEG-4 videos relies mainly on the blockiness metric. Tuning was performed on a randomly selected half of the data, and the other half was used for evaluation.

The results over all test sequences are shown in Figure 8, and the prediction performances are summarized in Table 4. The MOS prediction works well, especially considering the fact that it is based on non-reference metrics. The prediction performance is characterized by correlations of around 90%. The prediction error residual (7.4 MOS units on the SSCQE scale from 0 to 100) is comparable in size to the confidence intervals of the subjective data. For comparison, PSNR correlation with the same MOS data is only around 40%.

Table 4. MOS prediction performance.

	Linear correlation	Rank-order correlation	Prediction error
MPEG-4	91%	89%	8.2
M-JPEG2000	93%	89%	7.1
Overall	**93%**	**89%**	**7.4**
PSNR	39%	43%	—

A slight problem that can be noticed in the scatter plot is the separation between the low-quality conditions (64 and 128 kb/s) and the high-quality conditions (384 kb/s), which is somewhat overestimated by the metric for the sequences encoded with Motion JPEG2000. While this difference can also be observed in the viewers' ratings, it is not quite as pronounced. It appears to be difficult for the metrics to cope with both the severe artifacts in the low-bitrate sequences and the rather good quality of the high-bitrate clips at the same time. Furthermore, the transmission error effects are not always measured correctly by the three artifact metrics.

* See http://www.genista.com/ for more information.

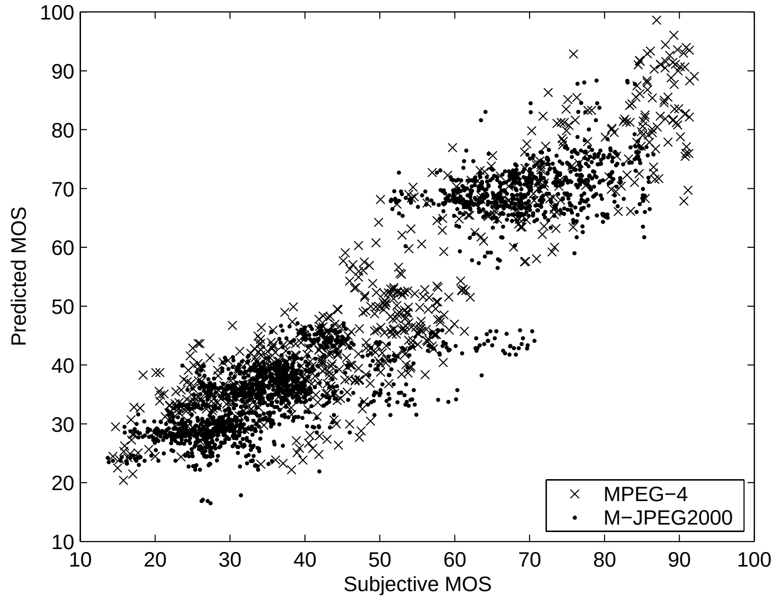


Figure 8. Predicted MOS vs. subjective MOS.

While this scatter-plot comparison and correlation gives a reasonable indication of metric performance, it is probably not the best way to analyze this type of data. Due to the auto-correlation of the time series (i.e. each sample is dependent on the previous and following samples), the values are not independent. This problem is not addressed in existing recommendations and standards. It also makes it difficult to separate the data into tuning and test sets in a meaningful fashion. We are currently examining approaches that are better suited for the comparison of such time series data.

6. CONCLUSIONS

We presented results on quality evaluation for mobile video applications. The test material was selected to be representative of the target application. Test sequences spanning a wide range of content were compressed with two different codecs using typical bitrates and were then subjected to WCDMA bit error patterns. The ratings obtained in the subective experiments proved to be very reliable despite the low quality of the sequences and the highly transient distortions. The codec performance comparison nicely shows the scene- and bitrate-dependent benefits of motion prediction, while the investigation of transmission error effects on perceived quality leaves some open questions. We demonstrated the good MOS prediction performance attainable for this type of material with *Stream PQoS*, a real-time non-reference quality metric. Future work will focus on better analysis methods for time series data as well as improvements of the artifact metrics.

ACKNOWLEDGMENTS

We would like to thank Genista Corporation and all the people involved in the design of the experiments, the metrics and the software described in this paper. We also thank the viewers who participated in our tests. Finally, we acknowledge the support of Prof. Sabine Süsstrunk at EPFL's Audiovisual Communications Lab, whose testing facilities we used for conducting the subjective experiments.

REFERENCES

1. 3GPP: *QoS for Speech and Multimedia Codec: Quantitative Performance Evaluation of H.324 Annex C over 3G*. Technical Report 26.912 (Release 4), 3GPP, 2001.
2. Thierry Alpert, Vittorio Baroncini, Diana Choi, Laura Contin, Rob Koenen, Fernando Pereira, H. Peterson: "Subjective evaluation of MPEG-4 video codec proposals: Methodological approach and test procedures." in *Signal Processing: Image Communication*, vol. 9, no. 4, pp. 305–325, May 1997.
3. Touradj Ebrahimi, Fernando Pereira: *The MPEG-4 Book*. Prentice Hall, 2002.
4. Roelof Hamberg, Huib de Ridder: "Continuous assessment of perceptual image quality." in *Journal of the Optical Society of America A*, vol. 12, no. 12, pp. 2573–2577, December 1995.
5. ITU-R Recommendation BT.500-11: "Methodology for the subjective assessment of the quality of television pictures." International Telecommuncation Union, Geneva, Switzerland, 2002.
6. ITU-T Recommendation H.223: "Multiplexing protocol for low bit rate multimedia communication." International Telecommuncation Union, Geneva, Switzerland, 2001.
7. ITU-T Recommendation P.910: "Subjective video quality assessment methods for multimedia applications." International Telecommuncation Union, Geneva, Switzerland, 1996.
8. ITU-T Study Group 16: *WCDMA Error Patterns at 128 and 384 kbps*. Technical Report Q15-G28, ITU, 1999.
9. Pina Marziliano, Frédéric Dufaux, Stefan Winkler, Touradj Ebrahimi: "A no-reference perceptual blur metric." in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 57–60, Rochester, NY, September 22–25, 2002.
10. MPEG: *AHG Report on Editorial Convergence of MPEG-4 Reference Software*. Technical Report M8041, ISO/IEC JTC1/SC29/WG11, 2002.
11. Ann Marie Rohaly et al.: "Video Quality Experts Group: Current results and future directions." in *Proceedings of SPIE Visual Communications and Image Processing*, vol. 4067, pp. 742–753, Perth, Australia, June 21–23, 2000.
12. David S. Taubman, Michael W. Marcellin: *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2002.
13. VQEG: "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment." April 2000, available at http://www.vqeg.org/.
14. Anna Watson, M. Angela Sasse: "Measuring perceived quality of speech and video in multimedia conferencing applications." in *Proceedings of the ACM Multimedia Conference*, pp. 55–60, Bristol, UK, September 12–16, 1998.
15. Stefan Winkler, Animesh Sharma, David McNally: "Perceptual video quality and blockiness metrics for multimedia streaming applications." in *Proceedings of the International Symposium on Wireless Personal Multimedia Communications*, pp. 547–552, Aalborg, Denmark, September 9–12, 2001.
16. Stefan Winkler, Ruth Campos: "Video quality evaluation for Internet streaming applications." in *Proceedings of SPIE Human Vision and Electronic Imaging*, vol. 5007, Santa Clara, CA, January 21–24, 2003.