MEASURING MULTIMEDIA QUALITY IN MOBILE NETWORKS WITH AN OBJECTIVE PARAMETRIC MODEL

Jörgen Gustafsson, Gunnar Heikkilä, Martin Pettersson

Ericsson Research, Sweden

ABSTRACT

The perceived quality of a multimedia stream can be estimated with good accuracy with a parametric no-reference opinion model taking codec, total coded bitrate, packet loss rate, and buffering events as input. A rebuffering event in a multimedia stream severely reduces the perceived quality with up to 1.5 MOS units for a QCIF video. High packet loss rates can reduce the quality with as much as 2.5 MOS units. The effect of these two degradation types has been combined in a non-linear way together with information about codec and coded bitrate to form a parametric no-reference objective opinion model. The model shows good performance with RMSE of 0.33 and correlation coefficient of 0.93.

Index Terms**—** Parametric opinion model, multimedia streaming quality, buffering, packet loss

1. INTRODUCTION

The growth of high-performance mobile and wireline networks enables new video and multimedia services to be deployed around the world. On-demand video streaming is one of the most popular multimedia services in new 3G networks and widely used in wireline broadband networks. End-user perception is crucial for successful deployment of a video service, and for on-demand video streaming the only practical way to measure end-user quality is to use an objective opinion model.

In on-demand streaming data is transported to a video player (located for example in a mobile device) over a network while the video is watched. The most influential quality factors are the coding quality (media production's impact on quality) and potential transport problems. The two main reasons for quality degradation due to transport problems are packet loss and too low data throughput over the network. Packet loss will cause image artifacts and corrupted audio. Too low throughput will eventually cause the video player to interrupt playing and reload data (so called rebuffering).

Much of the quality-evaluation work made for multimedia streaming so far has been targeting streaming over the Internet [1][2][3]. Due to the differences in media encoding and network characteristics these results cannot be directly applied to mobile applications. The work that has been carried out for mobile streaming [4][5] has mainly been focusing on the impact of packet losses when it comes to transport problems, not buffering events.

2. MODEL CONCEPT

An objective opinion model for multimedia streaming uses input parameters to estimate the subjectively perceived quality of the video. A common way to classify objective opinion models is to divide them into *perceptual* and *parametric* models. The input to a perceptual model is the multimedia signal, i.e. audio and video (which can be perceived by a human). The full-reference video quality models for SDTV standardized by ITU-T [6] are examples of perceptual models. Work is also ongoing in VQEG [7] to evaluate a number of lower bitrate perceptual quality models for ITU-T standardization.

The input parameters for parametric models are measurable, but not directly perceivable by the human eyes or ears. Examples of such parameters are packet losses, bitrate, codec information, and client events. An example of a parametric model is the opinion model for video-telephony applications that has recently been standardized by ITU-T [8]. No model has yet been standardized for multimedia streaming, but the work is ongoing.

We have developed a parametric model to estimate the perceived quality of multimedia streaming in a mobile environment. A parametric model has the properties of not requiring much computational resources during operation, being tuned for a specific purpose (such as on-demand video streaming) and providing good prediction performance within its scope. This model type is appealing for test purposes in a mobile network. A parametric model can also be implemented in network nodes to monitor the quality of a service. A parametric model is typically trained with a number of different and typical video content types. Hence a parametric model will for a given set of input parameters give the average score over the content types it is trained for. This property of a parametric model is appealing for test purposes, since the average quality for typical content is what a network operator is interested in. In contrast, for a perceptual model, which gives different score for different content, a network operator must run multiple tests and afterwards calculate an average score for different content.

The basic structure of our model is:

EstMOS = f(MOSBase, initial_buff, rebuf, packet_loss)

The estimate of the subjective Mean Opinion Score (EstMOS) is calculated based on the base quality for a given codec and bitrate (MOSBase), the length of the initial buffering (initial_buff), rebuffering events (rebuf) and packet loss rate (packet_loss).

3. EXPERIMENT RESULTS

3.1 Test Set-Up

The model is based on the results of four subjective tests:

1. Test the effect of rebuffering and initial buffering on perceived quality of video streaming.
2. Test of clean quality (no transport errors for video streaming).
3. Test the effect of packet loss on video streaming.
4. Test the combined effect of rebuffering and packet loss on video streaming.

The video format QCIF was used for all tests. The test sequence length was 30 seconds for test 1, 2 and 4 and 8 seconds for test 3. Four different types of content were used: news, sports, movie trailer and music video. All tests were performed using the *ACR with hidden reference* test method described in ITU-T P.911 [9].

About 24 test persons participated in each test which contained both encoded sequences, with total bitrates ranging from 32 to 256 kbps, and additional un-encoded high-bitrate sequences. A long sequence length (30 seconds) was used to be able to test how rebuffering interruption effected perceived quality. For the rebuffering degradation a progress indicator was used to emulate, as well as possible, a standard video streaming client.

Figure 1. Rebuffering progress indicator

A subjective test has also been performed [10] showing that the single stimulus ACR test method performs equally well as the double stimulus DCR test method when it comes to judgment of multimedia sequences with packet losses and buffering interruptions.

3.2 Test Results

The impact of the amount of rebuffering on the total quality can be observed in Figure 2, where the total quality is plotted against the rebuffering percentage (percent of clip length), with 95% confidence intervals. Non-linear fitting curves are also drawn to highlight the quality-variation trends. The quality decreases as the rebuffering percentage increases. It can be noticed that at the lowest rate, the quality deteriorates only slightly with increased rebuffering length. This is because the quality has almost reached the lower limit. Long rebuffering affects the total quality severely and the quality may drop as much as 1.5 MOS units within our test range.

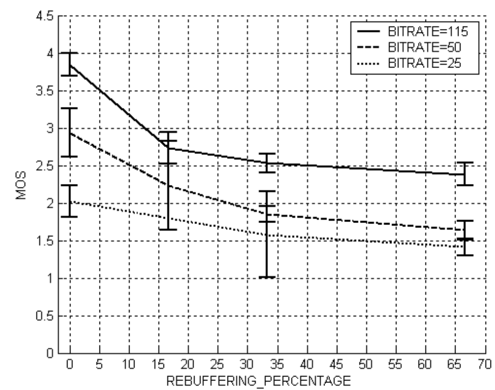

Figure 2 - Streaming Quality vs. Rebuffering Percentage

The impact of rebuffering location on streaming quality was also tested for a single rebuffering event. Two rebuffering interruption locations were tested for a 30 seconds video sequence, one close to the beginning (after 10 s) and the other close to the end (after 20 s).

We found that the rebuffering close to the end caused a quality reduction of about 0.2 MOS compared to that near the beginning. This demonstrates the presence of the recency effect: the subjects' memory of interruption events faded gradually until the scoring. This has been described in more detail in [11].

The effect of packet loss for different bit rates can be observed in Figure 3 (sequence length 8 seconds). As expected the increased packet loss reduces the perceived quality in a non-linear way. In addition, the quality ranking is in agreement with the bit rate ranking. For the same packet loss rate, the 256 kbps sequence exhibits the best quality, whereas the 24kbps shows the worst quality.

The effect of packet loss for different types of content can be observed in Figure 4 below (sequence length 8 seconds). As expected low motion content, such as talking head news, gets better score than high motion video, such as music. In addition, regardless of content type, the non-linear dependency between quality and packet loss rate is maintained.

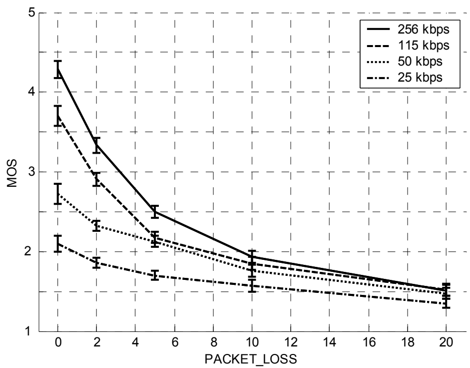

Figure 3 - Streaming quality vs. packet loss rate for different bitrates

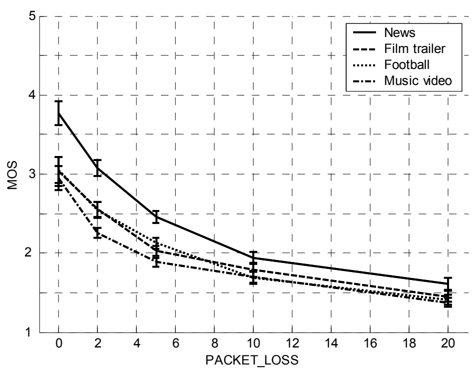

Figure 4 - Streaming quality vs. packet loss rate for different content (average for all bitrates 24 – 256 kbps)

If the effect of packet loss and rebuffering had been purely additive, then the degradation of a sequence could be modeled by first accounting for the rebuffering (assuming 0% packet loss), and then the degradation caused by packet loss (assuming 0% rebuffering). However, the subjective test clearly shows that the combination of packet loss and rebuffering cannot just be added together.

For a 256 kbps sequence without any packet loss, the quality drops from MOS 4.5 to 3.1 when a rebuffering of length 10 seconds occurs in the middle of the clip. When the rebuffering instead occurs together with a 10% packet loss rate, the quality only drops from MOS 1.8 to 1.6. In the first case the relative MOS drop is 1.4 and in the second case 0.2. This clearly shows that the quality degradations are not independent and a model estimating the perceived quality should not decrease the quality by pure subtraction, but rather with a multiplication factor for each degradation type. The effect of initial buffering in combination with packet loss was also investigated. The subjective test showed that the effect of initial buffering on perceived quality is smaller than the effect of rebuffering.

4. ESTIMATION OF MULTIMEDIA QUALITY WITH A PARAMETRIC MODEL

4.1 Model function

Our model for estimating multimedia streaming is based on the results from the four subjective tests. The model calculates an estimate of the MOS for multimedia quality. The calculation is a combination of the clean quality (codec and total bitrate), packet loss and buffering events. We have assumed that the different degradations can be modeled separately. The model function is a weighted product of the clean quality and all the individual degradations. The clean quality function is the estimated perceived quality on a scale ranging from 1 to 5 for a given video codec and bitrate. The buffering reduction and packet loss functions both estimate the quality reduction on a scale ranging from 0 to 1.

The clean quality function and the quality reduction functions for the MPEG4 video codec are shown in Figure 5 below. The solid line shows the experiment results and the dashed line the model estimation.

4.2 Model Performance

Our model performs well for the intended scope with QCIF video format and the video codecs it is tuned for. The RMSE distribution can bee seen in the upper left chart in Figure 6. The confidence-weighted (CW) error, which is the

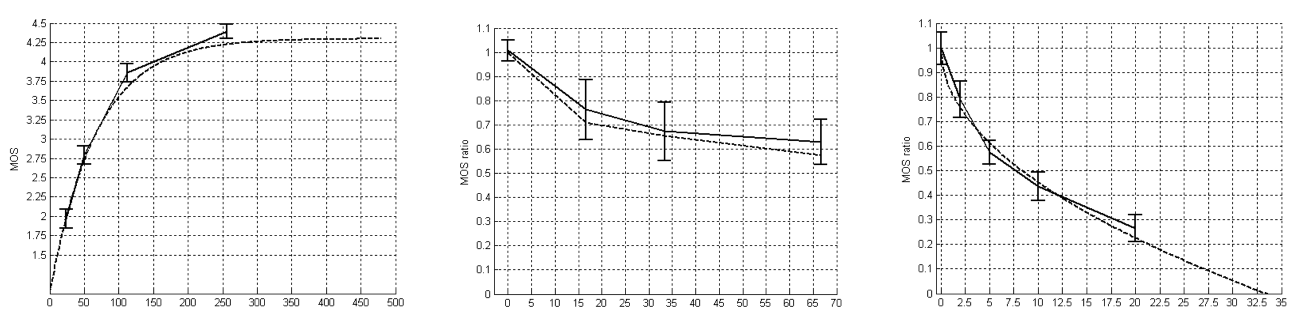

Figure 5 – Clean quality function for MPEG4 codec (left), rebuffering reduction (middle), packet loss reduction (right)

prediction error divided by the 95% confidence interval, can be seen in the upper right chart. For instance, a CW of 0.5 indicates that the error is only half of the confidence interval. The CDF for absolute error and absolute CW error can be seen in the lower part of the figure. The correlation coefficient for the model, when testing against the complete data set, is 0.93. The average RMSE for the model is 0.33 MOS. The RMSE is lower than 0.5 MOS in about 85% of the samples. About 65% of the predictions are also within the 95% MOS confidence interval. These performance results may however be somewhat optimistic since the same data set was used for both training and validation.

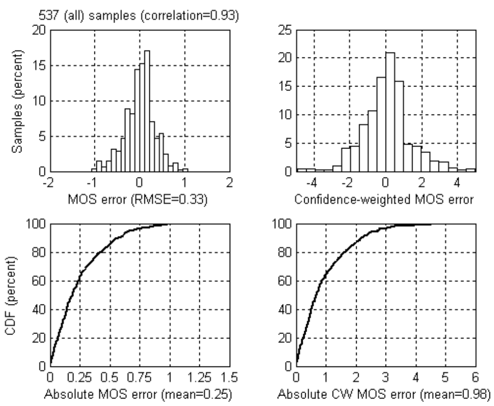

Figure 6 – Model performance

The relation between subjective MOS and the model-estimated MOS can be seen in Figure 7 below.

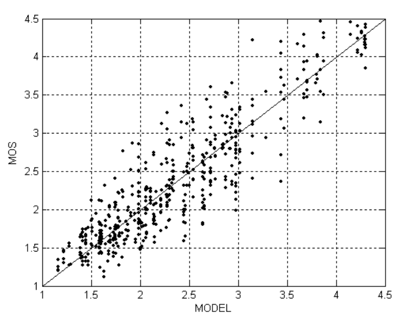

Figure 7 – relation between subjective and estimated MOS for the model

5. CONCLUSIONS

The impact of packet loss and buffering interruptions on perceived quality of multimedia streaming has been investigated. Rebuffering interruptions quickly reduce the perceived quality. A single ten-second interruption can reduce the quality with more than 1 MOS unit. Packet loss is an important factor for reduction of the perceived quality. A packet-loss rate of 4% reduces the quality for a 256 kbps video streaming with about 1.5 MOS units.

A parametric objective-opinion model has been developed. The model takes codec, bitrate, packet loss rate and buffering events as input and estimates the subjective MOS for a multimedia video-streaming sequence. The model shows good performance within its scope with an RMSE of 0.33 and a correlation of 0.93. It has lower computational resource requirements than models analyzing audio and video media. A parametric model has, within its scope, about the same performance compared to a perceptual model. A parametric multimedia model gives the average score for typical video content for a certain transport quality, which is often attractive for service quality monitoring.

6. REFERENCES

[1] M. Clarkpool, J. Tanner, "The Effects of Jitter on the Perceptual Quality of Video", in *Proc. of the 7th ACM inter. conf. on multimedia*, pp 115-118, Orlando, Florida, USA, October 1999

[2] S. Winkler and R. Campos, "Video Quality Evaluation for Internet Streaming Applications", in *Proc. SPIE/IS&T, Human Vision and Electronic Imaging Vol 5007*, pp104-115, Santa Clara, CA, USA, January 2003

[3] Z.Wang, S. Banerjee and S. Jamin, "Sudying Streaming Video Quality: From an Application Point of View", *Proc. Of the 11th ACM inter. conf. on multimedia*, pp. 327-330, Berkeley, CA, USA, November 2003

[4] M.B. Myers, "Predicting and Measuring Quality of Service for Mobile Multimedia", in *Proc. of the 11th IEEE inter. symp. personal, indoor & mobile radio comm. (PIMRC 2000)*, pp. 1032-1036, London, UK, Sept. 18-21, 2000.

[5] S. Winkler and F. Dufaux, "Video Quality Evaluation for Mobile Application", Proc. SPIE/IS&T, *Visual Communication and Image Processing*, Vol. 5150, pp. 593-603, Lugano, Switzerland, July 2003

[6] ITU-T Recommendation J-144 Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference, March 2004

[7] VQEG, http://www.vqeg.org

[8] ITU-T Recommendation G.1070 Opinion model for video-telephony, April 2007

[9] ITU-T Recommendation P.911 Subjective audiovisual quality assessment methods for multimedia applications, December 1998

[10] Ericsson AB contribution C127 to ITU-T Study Group 12 meeting, October 2007 "Performance Comparison of the ACR and the DCR subjective test methods for multimedia with rebuffering interruptions and packet loss "

[11] X. Tan, J. Gustafsson and G. Heikkilä, "Perceived Video Streaming Quality under Initial Buffering and Rebuffering Degradations", MESAQIN 2006 conference