Perceived Video Streaming Quality under Initial Buffering and Rebuffering Degradations

Xiangchun Tan, Jörgen Gustafsson, Gunnar Heikkilä
Ericsson Research, Ericsson AB, Luleå, SWEDEN
{xiangchun.tan, jorgen.gustafsson, gunnar.heikkila}@ericsson.com

Abstract

The subjective quality of video streaming has been studied for typical mobile setups, where bad radio conditions give rise to buffering effects (initial buffering and rebuffering). The study included a variety of transport bit rates, frame rates, types of content and buffering/rebuffering conditions. The properties of interruptions having the greatest impact on the perceived quality were found to be the rebuffering length and the rebuffering frequency. If interruptions are unavoidable, a single rebuffering is better than repeated rebuffering events. Initial buffering is better tolerated by mobile users than rebuffering. Results from the study have been used for predicting video streaming quality.

Keywords

video streaming, subjective quality, rebuffering interruption, quality model

1 Introduction

The third generation telecommunication systems make mobile video streaming a reality in the daily life [1][2].

Compared to wire-line internet streaming applications, streaming video in cellular systems is subject to more stringent constraints, such as harsh air interference and limited radio resources. The content providers and network operators need means to measure and monitor the service quality for customer satisfaction. However, video quality assessment for mobile streaming is a fairly young field, and specific measurements and assessment methods are not yet mature [3].

Among the various streaming applications is the on-demand streaming, with which the encoded video contents are transmitted to the mobile terminal from the content server through the core and radio network. Our study will concentrate on this type of streaming.

The quality of a streaming session is affected by a number of factors: first, the quality of the source video and how it is encoded before transmission; second, the interference introduced during the network transmissions; and finally, the decoding and displaying at the end-user terminal. The video data is buffered in the network nodes and in the mobile terminal. In case of bad radio conditions and/or cell changes, throughput outage may occur in radio channels and the mobile terminal buffer may then be emptied so that the mobile user will experience interruptions, such as long start-up times and freezing of the video.

In this article, we present our study of the effect of interruptions on user-perceived streaming quality and delineate the impact of various parameters on the subjective quality. We describe how we simulated interruptions in video streaming and conducted subjective tests for rating the video, audio and overall quality of the degraded material when presented in a mobile-like terminal. We then analyze our results and give examples of quality modelling

The paper is organized as follows: The method of assessing the streaming quality is described in chapter 2, followed by the setup of the subjective test in chapter 3. The results of the subjective test are studied in chapter 4 with regard to the effects of the interruption parameters. In chapters 5 and 6, the features of content type, the subjective video, audio and total quality are investigated respectively, and a quality model is established in chapter 7. Finally conclusions and suggestions are made for further studies in assessing the video streaming quality for mobile applications.

2 Methods

In video streaming, several factors affect the streaming quality: from the network and transport technology point of view, the data rate, packet delay and jitter, and packet loss. Much work has been done on video quality evaluation for internet systems with regard to transport disturbances [4]-[7]. These results, however, cannot be directly applied to mobile applications due to the differences in media encoding, system conditions, and subjective viewing experience.

In a mobile system, the video streaming quality perceived by users is a function of the specific mobile environment and the terminals used. The work done so far on quality evaluation of degraded video sequences has concentrated on artifacts due to transport data bit errors [3][9][10] while paying little attention to interruptions such as initial buffering and rebuffering, which occur in on-demand streaming in a 3G system. We shall focus on the latter type of degradation.

The impact of interruptions on streaming quality for internet applications has in fact been observed in one study [7]; however, these results are rather limited regarding the test ranges and do not give a good understanding of the effects of the influential factors. Methods exist for monitoring interruptions in internet applications [8][11], but the measurement metrics used are not interpreted as subjective quality.

In the quality assessment, the 5-level scale (Excellent, Good, Fair, Poor, Bad) is often used. The subjective quality score graded from a test person is subject to his/her own personal experiences and expectations related to the media content. It is up to the test objectives how to handle these quality diversities.

We carried out subjective tests for streaming with both video and audio on a portable viewing device to mimic the complete experience of mobile streaming and obtain ratings of video and audio quality as well as of the overall quality of the session. We then associated the subjective streaming quality data with the interruption conditions to obtain a prediction model for estimating the streaming service quality under the influence of buffering interruptions. Our objective was to estimate the subjective quality based on the measurements of the service parameters.

3 Test setup

The subjective streaming tests were carefully designed with the following setup and cases. Four typical content types – news, film trailer, music video and sport – were selected. Data on the original video and key test settings are shown in Table 1. The streaming interruptions comprise initial buffering and rebuffering. Initial buffering is characterized by its duration; rebuffering is characterized by the number of rebuffering events and the duration and time of occurrence (location) of each of them. When defining the location of a rebuffering event, the uninterrupted original clip is used as reference.

Table 1 Source streaming material and basic interruption settings

Content	News Film trailer Music video Sport	
Original format	RGB QCIF 25 fps 30 s in length	
Codec	H.263 MPEG -4 H.264 Real video 9/Audio 8	
Bit rate	24, 50, 116 kbps	
Frame rate	5, 8.3, 12.5 fps	
Interruptions	Initial buffering duration T_{init}:	2, 15 s
	Rebuffering, number of events N_{buff}:	1, 2
	Rebuffering, duration of events T_{rb}:	5, 10, 20 s
	Rebuffering, location of events:	10, 15, 20 s

An example of the buffering profile is illustrated in Figure 1.

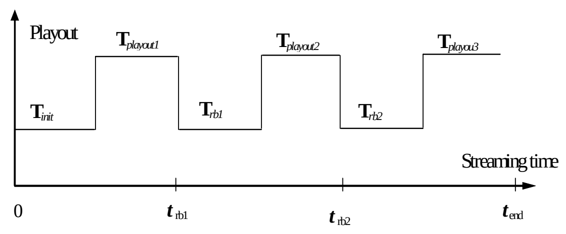

Figure 1 A buffering profile

Obviously, testing all combinations of properties and settings in Table 1 would entail a very large number of test cases. We selected 11 buffering conditions which resulted in the following number of test cases: 12 (11 buffering profiles + clean) times 4 (content types) times 3 (bit rates) plus 4 (originals). The testing time was still found to be excessive, and therefore we screened the test cases by removing part of the lower bit rate stimuli, eventually cutting down the number of test stimuli to 89. See Table 2

Table 2 Design of buffering profiles and stimuli to be used in the subjective test (The superscripts 1 and 2 denote rebuffering starting at location 1 and 2 respectively as shown in Figure 1)

[Initial, Rebuff.]	24 kbps	50 kbps	116 kbps	Original
Reference				4
Clean	4	4	4	
[2, 5]	2		4	
[2, 10]	2	4	3^1	
[2, 10]			4^2	
[2, 20]	4	3	3^1	
[2, 20]			4^2	
[2, 5, 5]		3	4	
[2, 10, 10]		4	4	
[15, 5]		4	4	
[15,10]	1	2	3^1	
[15,10]			4^2	
[15, 5, 5]		4	3	
Total: 89	13	28	44	4

This added up to a viewing time of about two hours. To avoid viewer fatigue and maintain consistency in the subjective quality ratings, the viewing was divided into two sessions held on consecutive days.

Table 3 Subjective viewing conditions

Viewing device iPAQ (Pocket PC h5500)	TFT 65536 colors (16 bits) 240 x 20 pixels Enhanced processor RXA255 128 ROM/48 RAM
Audio headset	Sony MDR-023
Viewing conditions	Room light 3 Lux Screen luminance "Bright" 84 cd/m^2 "Black" 3 cd/m^2

A portable PC (Compaq iPAQ) was used for playback of the video stimuli: see Figure 2. For this purpose the stimuli were converted to MPEG-1 format at a high bit rate (about 700 kbps). Key features of the test device and viewing conditions are listed in Table 3. The watching distance was 4–6 times the screen width. The subjects listened with both ears and the volume was set at the default level, but could be adjusted by the test person.

Figure 2 Pocket PC h5500

The subjective test followed the ITU P.910 standard [12]. The Absolute Category Rating (ACR) method with hidden references and a five-point rating scale (Excellent, Good, Fair, Poor, Bad) were chosen to grade the subjective streaming quality.

The stimuli were randomized and the order of viewing was different for each test person. Before the formal test, a training sequence with best and worst quality was shown to familiarize the test subjects with the quality variations. After presentation of each video stimulus, the subjects were prompted to mark their scores for audio, video and total quality. The scores were set with a stick pen directly on the iPAQ screen.

The viewers were instructed to rate the audio and video quality without considering the interruptions in particular but judge the total quality by perceiving the streaming session as a whole, i.e. taking into account both audio quality, video quality and buffering interruptions (if any).

4 Results

4.1 General

Twenty-six subjects took part in the subjective test. They ranged from young to middle-aged and were regular mobile users with no involvement in the fields of video encoding, streaming or quality assessment.

Since the total streaming quality reflects the overall quality of audio, video and combines the effect of buffering interruptions, the total quality is the parameter used in our buffering analyses in section 4.2.

4.2 Effect of buffering interruptions on streaming quality

Besides the basic interruption settings (Table 1), a number of further buffering parameters are defined in order to elucidate the interruption features and to facilitate observing their effects on the total streaming quality. See Table 4, in which most of the parameters are none-dimensional and thus more versatile in describing the interruptions. The meanings of the basic variables in Table 4 are illustrated in Figure 1 and Table 1. It should be noted that the display time is equal to the sum of the clip length and the rebuffering durations.

Table 4 Definition of buffering parameters

Parameters	Definitions
Streaming display time, $T_{display}$	$\sum_i T_{rb,i} + \sum_j T_{playout,j}$
Rebuffering percentage (%)	$\frac{\sum_i T_{rb,i}}{T_{display}} \times 100$
Rebuffering frequency (/min)	$\frac{N_{buff}}{\sum_j T_{playout,j}} \times 60$
Location of a rebuffering (%)	$\frac{T_{playout1}}{\sum_j T_{playout,j}} \times 100$
Initial buffering (%)	$\frac{T_{init}}{T_{display}} \times 100$

4.2.1 Rebuffering time

The effect of the amount of rebuffering on the total quality can be observed in Figure 3, where the total quality is plotted against the rebuffering percentage (different symbols being used for each of the three bit rates). Non-linear fitting curves were also drawn to elucidate 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.

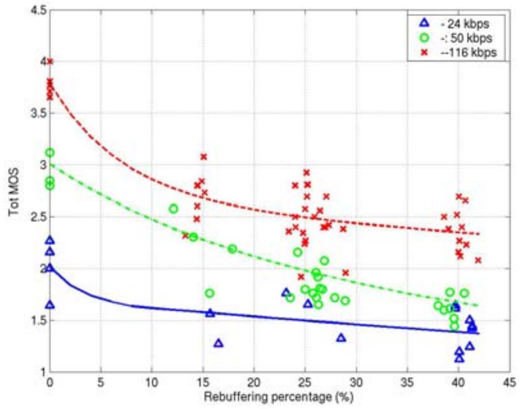

Figure 3 Streaming quality vs. rebuffering percentage

4.2.2 Rebuffering location

The effect of rebuffering location on streaming quality was tested only for the 116 kbps stimuli and the single rebuffering cases. The quality data with three different buffering profiles are presented in Figure 4. The common feature of these profiles is that the rebuffering event may occur in either of two places, one close to the beginning (10 s after start of video) and the other close to the end (20 s after start of video). The total quality is drawn against the relative location parameter in Figure 4. Fitting lines are drawn for each of the three profiles.

It can be seen that, for all profiles, 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 gradually faded, and the more the longer time that elapsed from the event until the scoring. The quality dropped by almost the same amount for all three test profiles. We can infer that there is no cross-interaction between the location factor and the initial buffering time/rebuffering time in the tested interruption ranges.

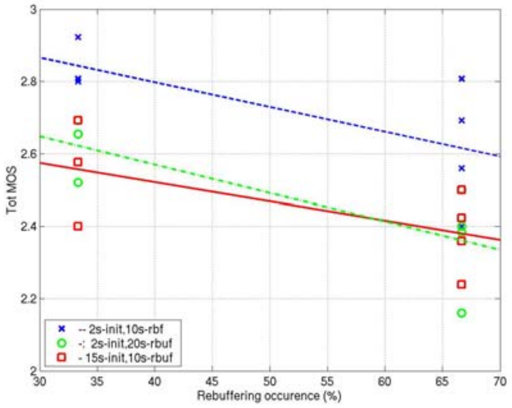

Figure 4 Streaming quality vs. rebuffering location

4.2.3 Rebuffering frequency

When keeping the total rebuffering duration constant, we found that the quality was reduced more if two rebuffering events occurred than if all of the rebuffering was done on a single occasion. See Figure 5; the curves are fitted separately for the encoding rates 50 kbps and 116 kbps for easy observation of the variation trends. We expect that as the number of rebufferings increases, the total quality will converge to 1.

Note that a rebuffering frequency of 2 means that one rebuffering event occurred to the 30-s long uninterrupted video sequences.

We observe that the quality reduction is similar (0.3 MOS) for both bit rates. Furthermore the quality drops faster when N_{buff} goes from 0 to 1 than it does when N_{buff} increases from 1 to 2. This implies that the appearance of the first interruption lowers the quality experience considerably. Additional, separate interruptions decrease the quality further, but not by the same amount. Higher rate stimuli (116 kbps) are subject to a larger reduction than lower rate stimuli (50 kbps) in the initial drop (1.2 MOS respective 0.8 MOS). This can be explained by the fact that for video clips encoded at higher bit rates and thus being of higher quality, viewers are more sensitive to the quality degradation.

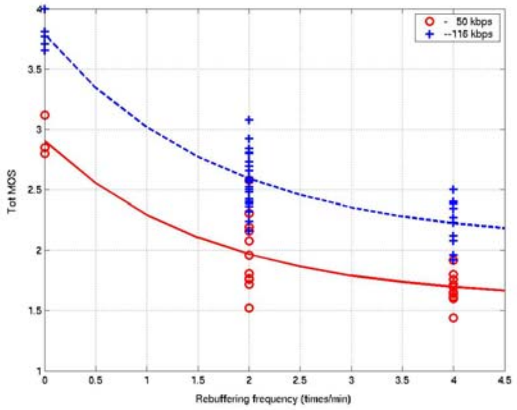

Figure 5 Streaming vs. quality rebuffering frequency. A rebuffering frequency of 2 means that one rebuffering event occurred.

4.2.4 Initial buffering

Like rebuffering, initial buffering affects the total quality negatively. Three groups of stimuli quality are shown in Figure 6, each comprising stimuli of two initial buffering time lengths and differing as to their rebuffering profiles or encoding rate conditions. It is remarkable that less quality reduction occurs at the 50 kbps bit rate than at 116 kbps. This result is similar to the observation in section 4.2.3 that the quality degradation is a combination of the quality of the clip and the interruptions. There was greater diversity in the perception of quality and the degradations were more noticeable for the higher bit rate.

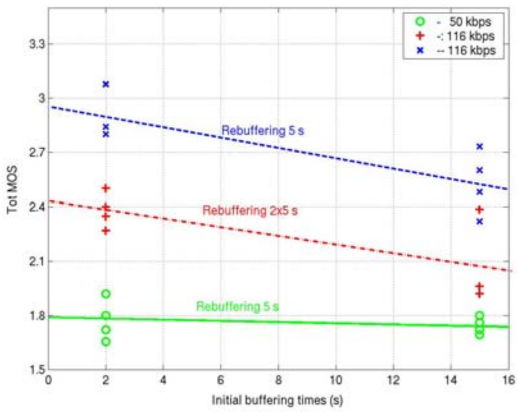

Figure 6 Initial buffering and streaming quality

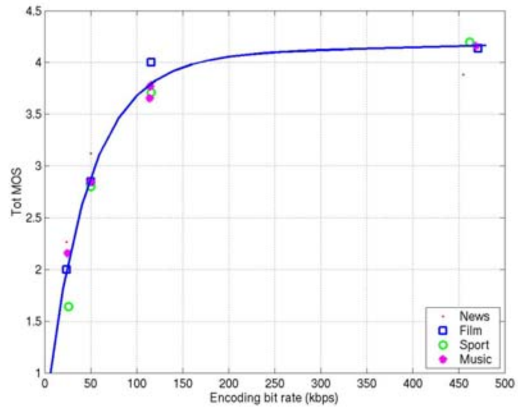

Figure 7 The total streaming quality for clean stimuli of four content types

5 Impact of content type and encoding rate on total quality

In the subjective test, the viewers rated the audio, video and total quality for the various content types. It is interesting to see what effect the content type has on the ratings and how the audio, video and total quality scores are related. Since the video and audio quality was rated for the undisturbed video playout, whereas the total quality was assessed also taking buffering interruptions into account, the two are comparable only under the "clean" conditions (no interruptions). Therefore we take only the clean and reference samples into consideration in this section.

The total streaming quality is shown for the different content types and encoding rates in Figure 7. It can be seen that at a fixed encoding rate, the differing complexity of the content in each case caused a spread in total quality. At lower encoding rates, the sport clips, consisting of video sequences from a football match (with fast-moving players shown alternately up close and at a distance) received lower quality scores than more "stationary" news clips with a reporter talking. However, at the higher encoding rates, the effect of the content type on the perceived quality was less apparent.

6 Audio, video and total quality

Given the subjective quality measurements obtained in our tests, it is natural to investigate how the audio (A_MOS) and video (V_MOS) quality contribute to the total quality (TOT_MOS). We take A_MOS, V_MOS as the independent variables and TOT_MOS as the dependable. A linear regression was carried out resulting in

$$TOT\_MOS_{pred} = 0.4810 \cdot A\_MOS + 0.4554 \cdot V\_MOS + 0.2868$$

The total quality predicted by this formula is compared with the subjective total quality scores in Figure 8. The root mean square error (RMSE) is 0.084 MOS and the correlation coefficient is 0.995.

The regression results indicate that the audio quality and video quality contribute almost equally to the total quality and that no strong bias of the quality components is observed for any of the content types. The reasons might be that the effect was weak and overwhelmed by the mixing of the various conditions of our subjective test, and that the divisions of total bit rates into audio bit rate and video rate were not extensive, but the best possible selections were used instead.

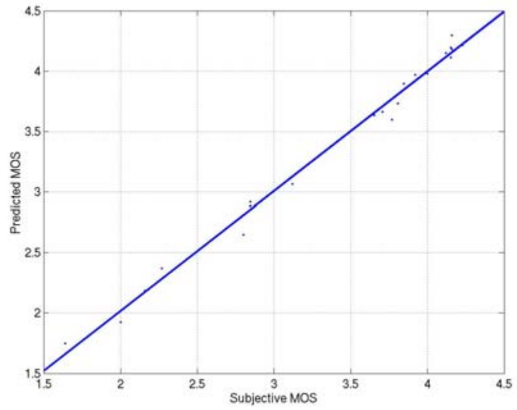

Figure 8 Relationship between total quality and audio and video quality in clean and reference cases

It is noticed that this quality relationship is based on the subjective test in which the video, audio and total quality were perceived simultaneously. This procedure is in agreement with the experience of real-life streaming. This result differs from studies where the tests on the audio and video were carried out separately and which indicate different weights of audio quality and video quality and their interactions in relating with the total quality [13]-[16]. Hence the quality perception and rating methods play a role in the context of the quality and their correlations.

7 Modelling of streaming buffering

The results from the study of subjective perception of the effect of buffering interruptions on the video streaming quality for mobile applications were used for predicting the service quality. A buffering quality model was established by including all these effects:

$$MOS_{Buffer} = f(MOS_{Base}, initial\_buff, rebuff)$$

It is assumed here that the video streaming quality is affected by the quality of the original video and the information loss due to encoding/compression (effects reflected by the parameter MOS_{base} in our model) and by the interference encountered in the transport (represented by the initial buffering and rebuffering interruptions).

The quality predicted by this model is plotted against the subjective quality scores in Figure 9. The data points clustered in the upper right corner are the clean or reference stimuli and the majority of the remaining data points are stimuli subject to buffering interruptions.

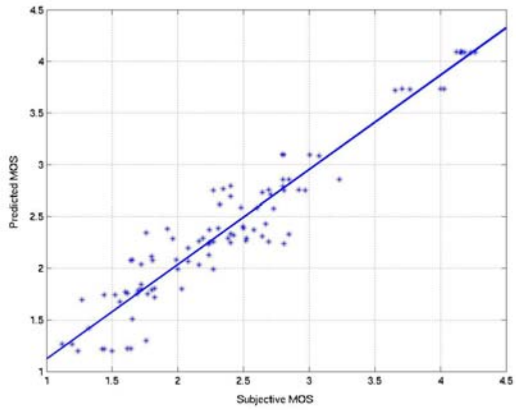

Figure 9 MOS prediction by the rebuffering model

The resulting model confirms that rebuffering has a greater impact on the quality than initial buffering. Our work shows that it is possible to characterize the perceived streaming quality as a function of interruption degradations and of the integrated audio and video quality.

8 Conclusions and discussions

A subjective test of streaming quality has been carried out and the impact of buffering interruptions on the streaming quality observed. The most influential factors are the rebuffering length and rebuffering frequency. Initial buffering is more tolerated by the mobile customers.

It is better to have a single rebuffering than repeated rebuffering events if interruption is unavoidable.

Interruptions late in a video clip cause more irritation than those in the beginning, due to human perception and memory. It is hence very important to maintain good quality toward the end of a streaming session.

Interruptions are more noticeable for high bit rates, and the quality drops quickly when one occurs. For low to medium bit rate streaming, minor degradations might be tolerated.

The streaming quality can be modeled based on the basic clip quality and the streaming interruption parameters.

Further study on interactions of interruptions and other video artifacts/audio impairments are called for as in real-life streaming, both interruptions and video impairments will be encountered. A more robust quality model that handles these video degradation factors would be desirable. This poses more challenges in the design and execution of degradation simulation and subjective tests. The work can further influence formation of measurement strategies, processing and integration of different factors to build up a total end user quality measure to suit the different purposes of quality evaluations. Much can be done in providing quality information to the content providers and system operators: in real time, for a specific service, for specific mobile locations or travel routes, or for an entire cellular network.

References

[1] Video streaming to the mobile phone, *http://www.ericsson.com/mobilityworld/sub/countries/croatia/news_en/n_2004_Video_streaming.shtml.*

[2] Cingular 3G streaming launched, *http://www.3g.co.uk/PR/March2006/2735.htm*, March, 2006.

[3] Myers, M. B., "Predicting and measuring quality of service for mobile multimedia", *Proc. of the 11th IEEE inter. symp. personal, indoor & mobile radio comm. (PIMRC 2000),* pp.1032-1036, London, UK , Sept. 18-21, 2000.

[4] Testing user perceived streaming video quality, *http://www.eurescom.de/~public-webspace/P800-series/P807/index.html*, June, 1999.

[5] M. Clarkpool, J. Tanner, "The effects of jitter on the perceptual quality of video", *Proc. of the 7th ACM inter. conf. on multimedia,* pp. 115-118, Orlando, Florida, USA, Oct. 30 – Nov. 5, 1999.

[6] S. Winkler and R. Campos, "Video quality evaluation for internet streaming application", *Proc. SPIE/IS&T, Human Vision and Electronic Imaging Vol 5007,* pp. 104-115, Santa Clara, CA, Jan. 20-24, 2003.

[7] Z. Wang, et. al., "Study streaming video quality: from an application point of view", *Proc. of the 11th ACM inter. conf. on multimedia*, pp. 327-330, Berkeley, CA, USA , Nov. 02-08, 2003.

[8] A. C. Dalal, E. Perry, "A new architecture for measuring and assessing streaming media quality", *Passive and Active Measurement Workshop Proceedings*, La Jolla, CA, USA, pp. 223-231, Apr. 2003.

[9] Multimedia group Test plan, draft ver.1.11, *VQEG MM-190304002*, Feb. 2006.

[10] S. Winkler and F. Dufaux, "Video quality evaluation for mobile application", *Proc. SPIE/IS&T, Visual communication and Imaging processing, Vol. 5150,* pp. 593-603, Lugano, Switzland, July 8-11, 2003.

[11] Streamcheck, http://www.streamcheck.com, Oct. 2003.

[12] ITU-T P.910, Subjective video quality assessment methods for multimedia applications

[13] M. P. Hollier, et. al, "Multimodal perception", *BT Technol. J.*, vol. 17, pp. 35-46, 1999.

[14] D. S. Hands, "A basic multimedia quality model", *IEEE Trans. Multimedia,* vol. 6, pp. 806-816, Dec. 2004.

[15] S. Winkler, C. Faller, "Audiovisual quality evaluation of low-bit rate video", *Proc. SPIE/IS&T, Human Vision and Electronic Imaging*, pp. 139-148, San Jose, CA, Jan., 2005.

[16] R. M. Hollier, "The significance of Cross-Modal Interaction in Audio-Visual quality perception", *IEEE Signal Processing Society*, Workshop on Multimedia Signal Processing, Copenhagen, Denmark, Sept. 13-15, 1999.