T-V-MODEL: PARAMETER-BASED PREDICTION OF IPTV QUALITY

A. Raake, M.-N. Garcia, S. Möller

Berlin University of Technology
Deutsche Telekom Laboratories
Berlin, Germany

J. Berger, F. Kling

SwissQual AG
Zuchwil
Switzerland

P. List, J. Johann, C. Heidemann

T-Systems Enterprise Services GmbH
Darmstadt
Germany

ABSTRACT

The paper presents a parameter-based model for predicting the perceived quality of transmitted video for IPTV applications. The core model we derived can be applied both to service monitoring and network or service planning. In its current form, the model covers H.264 and MPEG-2 coded video (standard and high definition) transmitted over IP-links. The model includes factors like the coding bit-rate, the packet loss percentage and the type of packet loss handling used by the codec. The paper provides an overview of the model, of its integration into a multimedia model predicting audio-visual quality, and of its application to service monitoring. A performance analysis is presented showing a high correlation with the results of different subjective video quality perception tests. An outlook highlights future model extensions.

***Index Terms*—** Modeling, Multimedia systems, TV broadcasting, video coding, Monitoring

1. INTRODUCTION

When employed for broadcast TV, the video coding and packet-switched transmission over an IP-based network may lead to new types of degradations such as block artifacts, picture freezing or slicing (e.g. in case of packet loss). For a network operator or IPTV service provider it is important to plan the service in an efficient manner so that it yields a high level of perceptual quality, and to monitor the quality it delivers during service operation. For both purposes, it is desirable to use instrumental tools that provide estimates of the audiovisual quality perceived by the user.

Different approaches can be conceived, which can coarsely be classified according to the following set of criteria:

Target service: Service type, e.g. IPTV, VoD, mobile TV; video resolution, e.g. Standard Definition (SD), HD.

Model type: Presence of a reference – Full Reference (FR), Reduced Reference (RR), No Reference (NR).

Application: Codec testing, network planning, verification of Quality of Service classes, monitoring, etc.

Model input: Parametric description of the processing path, i.e. protocol information or planning values; additional payload information from bitstream; reconstructed pictures; combinations of parameters and pictures.

Model output: Overall quality in terms of MOS (Mean Opinion Score) or another index; diagnostic information on quality problems.

An overview of a number of existing or foreseen quality-model standards is given in Table 1, classified according to their in- and output information. Due to the focus on standardization, the table does not contain other relevant quality modeling approaches such as [1, 2, 3].

For service monitoring, RR or NR models are the best choice, since they do not require any reference signal. Hence, they can be applied at different points in the network including the client side. For network planning, only NR models can be used, since no signals are available during planning.

In this paper, we describe the framework for an NR audiovisual quality prediction model ("T-V-model"). It was developed to match the model framework recently outlined by Study Group 12 of the International Telecommunication Union (ITU-T). It foresees a core model (Figure 1), which can be applied in two different contexts: (a) For network planning, by using estimated service characteristics as input (currently referred to as "G.OMVS"), and (b) for service monitoring, by using bitstream information (referred to as "P.NAMS", Figure 2). In addition, VQEG is working on a hybrid model standard combining bitstream and picture-based information.

The paper is structured as follows: Section 2 outlines the model architecture; Section 3 provides a more detailed description of the video quality model and the covered effects. Also provided is a comparison of the model with results we obtained in subjective video quality tests. Finally, Section 4 summarizes our findings and gives an outlook to future work.

2. MODEL OUTLINE

The core part of the T-V-model is outlined in Figure 1. Here, the different input parameters to the model are classified into three categories:

				Model
input information	reference	A-only	V-only	AV quality
pictures	RR/NR			VQEG
”	FR	PEAQ	J.144	VQEG
parameters (estimated)	NR			T-V-Model (≡ G.OMVS)
parameters (from bitstream)	RR/NR			T-V-Model (≡ P.NAMS)
hybrid	FR/RR/NR			VQEG

Table 1. Overview of standardization of audiovisual quality models. J.144: [4]; PEAQ: [5]; G.OMVS: Standard currently developed by Study Group 12 of ITU-T (Opinion Model for Video Streaming applications); P.NAMS: Standard currently developed by Study Group 12 of ITU-T (Non-intrusive parametric model for the Assessment of performance of Multimedia Streaming); VQEG: Ongoing activities within the Video Quality Experts Group.

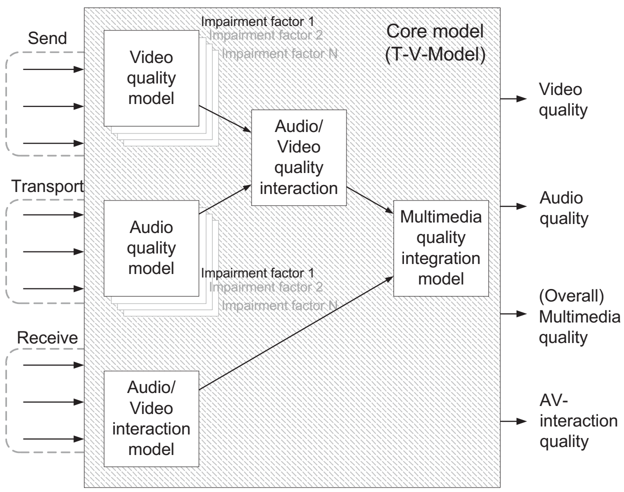


Fig. 1. Overview of T-V-Model core algorithm.

Send: Includes the input quality, “(en)coding information” such as codec type and bitrate, sender-based packet loss concealment options, frame- and key-frame rate, and the video format (e.g. SDTV, HDTV).

Transport: Comprises the “packet information” such as the packet loss rate and pattern, delay jitter, and more detailed media-information (e.g. on the type and location of lost information), the throughput, and also includes transport aspects such as the employed multiplexing (e.g. transport stream usage), and the payload size.

Receive: Combines both the “client information” such as the buffering behavior with the “decoding information”, such as the receiver-based packet loss concealment, the de-jitter-buffer behavior, etc. “Receive” information may also include information on the usage of de-interlacing or rescaling (video) and the audio rendering or electroacoustic properties.

As can be seen from Figure 1, the T-V-Model will consist of four parts: (A) An audio quality model, (B) a video quality model, (C) a model of the interaction between audio and video quality, and (D) a model of audio-visual quality. Note that the model parts (A) and (B) are inspired by the paradigm of the E-model [6], the model currently recommended by the ITU-T for planning voice communication networks. This model is based on an impairment-factor principle, which assumes that different classes of degradations can be transformed onto a quality scale by appropriate transformation rules, and that they are additive on that scale. Note further that model part (C) relates to the effect that certain levels of audio and of video quality have on overall (multimedia) quality; part (D) relates to audio-visual synchronization problems (i.e. “lip-sync”). In the following, the focus is on the video quality model and its performance.

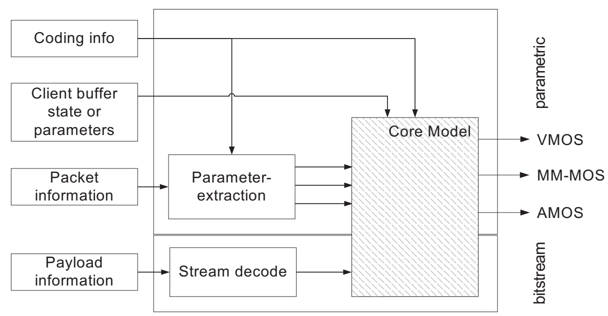


Fig. 2. Monitoring modus (ITU-T SG12’s “P.NAMS”).

3. VIDEO QUALITY MODEL

The basic formula of the video part of the T-V-Model is

$$Qv = Qvo - Ires - Icod - Itra - Idis \qquad (1)$$

Here, Qv is the video quality expressed on a scale ranging from 0 to 100 (100 for best quality). Qvo is the “input quality” reflecting the source quality of the video inserted into the transmission chain. $Ires$ is the impairment resulting from the picture resolution (in case of resolution changes). $Icod$ is the impairment introduced by the employed video coding, $Itra$

Rate (Mbit/s)	Key-Rate (1/s)	Ppl (%) 0	0.06	0.125	0.25	0.5	1	2	4
64	1	M	-	-	-	-	-	-	-
32	1	M, H	-	-	-	M(s)	M(s)	M(s)	M(s)
16	0.5	H	H(f)	H(s)	H(f,s)	H(s)	H(f,s)	H(s)	H(s)
	1	M, H	H(f)	H(s)	H(f,s)	H(s)	H(f,s)	M(s), H(s)	H(s)
8	0.5	H	-	-	-	-	-	-	-
	1	M, H	H(f)	-	H(f,s)	-	H(f,s)	-	H(s)
4	0.5	H	H(f)	H(s)	H(f,s)	H(s)	H(f,s)	H(s)	H(s)
	1	M, H	H(f)	H(s)	H(f,s)	H(s)	H(f,s)	H(s)	H(s)
2	1	H	-	-	-	-	-	-	-

Table 2. Overview of HD test conditions. "H" ≡ H.264; "M" ≡ MPEG2. Ppl denotes the average packet loss percentage. The abbreviations in brackets indicate the packet loss concealment type: "f" ≡ freezing; "s" ≡ slicing. The SD test conditions can be obtained by dividing the bitrates by four.

is the impairment due to transmission (e.g. packet errors and the corresponding sender- and receiver-based error concealment), and $Idis$ is the impairment introduced by the display and processing steps such as de-interlacing or acceleration.

So far, we have conducted separate subjective video tests for HD (1920x1080i) and SD (720x576i). This results in two independent models for the two resolutions. In order to express HD and SD video quality on a common scale, work is underway on a combined SD/HD test. Note that a direct comparison between the two television paradigms is complicated by the different expectations users show [7]. Since no resolution- or display-dependency is implemented in the model yet, we set $Qvo - Ires - Idis = \max(Qv_{subjective})$.

Sixty-four conditions were presented per image resolution. The processing chains comprised several conditions with the MPEG2 and H.264 codecs at different bitrates, with different packet loss rates (uniform distribution) and assuming two types of packet loss concealment (cf. Table 2). As transport mechanism, we used MPEG2-transport streams over RTP over UDP. Six anchor conditions were repeated in each of the four test sessions, covering the entire quality range presented in the test (both in terms of quality levels and possible perceptual effects). In all tests, five source sequences of 16 s duration and of different content were used: A soccer scene, an interview, a movie trailer, a music video, and a movie excerpt. The subjective test ratings were collected from 24 subjects per session using the 11-point quality scale according to [8]. For further processing, the 11-point mean opinion scores (MOS) were transformed onto a 5-point absolute category rating scale (ACR), using $MOS_5 = MOS_{11}/10 \cdot 4 + 1$. The 5-point MOS-values were then transformed to a scale ranging from 0 to 100 according to the transformation rule given in Appendix I of [6].

In a least-square curve-fitting procedure using the subjective test results for the keyframe-rate of $1/s$ as target values, we have identified the following relations for the different impairment factors (both HD and SD):

$$Icod = a_1 \cdot \exp(a_2 \cdot bitrate) + a_3 \quad (2)$$

$$Itra = (b_0 - Icod) \cdot \frac{Ppl}{b_1 + Ppl}. \quad (3)$$

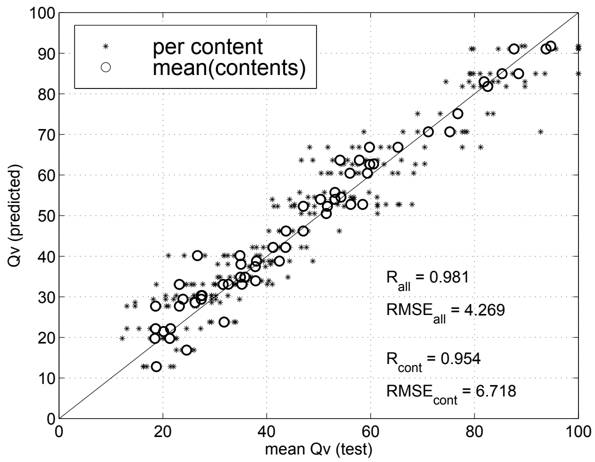
Here, the parameters a_i depend on the employed codec and resolution. The parameters b_i depend on the employed packet loss concealment scheme: For the slicing option used in our tests, we find that $b_1 = f(bitrate)$, which can be approximated by

$$b_1 = c_1 + c_2/bitrate. \quad (4)$$

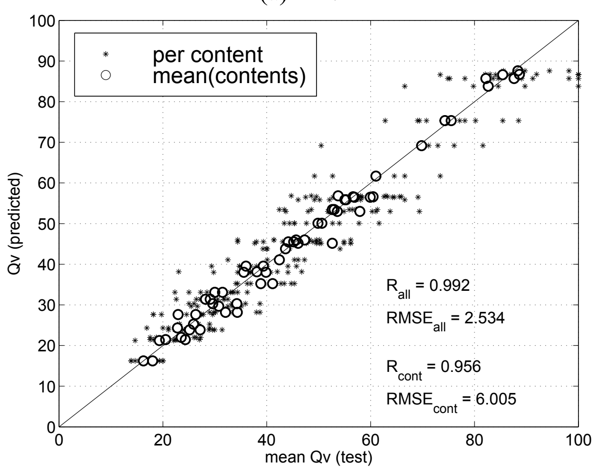
For freezing, the parameter $b_1 = const$ in case of HD, and bitrate-dependent for SD.

Figure 3 shows a comparison between the model predictions and our test results. For model training, only parts of the test data were employed. So far, no handling of the keyframe rate has been included, which reduces the prediction performance. Also provided in the figure are the linear correlations and root mean squared errors of the prediction. Note that two values for each performance measure are provided for both HD and SD: The first set of values with index "all" refers to the test results obtained after averaging over content-types; the second set of values ("$cont$") refers to the test results as a function of content. As can be seen from the graphs and performance measures, there is a measurable content-dependency so far not reflected by the model.

Our approach differs from others, for example: In [3] and related papers, quality is assessed in terms of the visibility of packet loss impairments. Instead, we collected quality judgements from subjects in order to more directly capture the overall effect due to coding and packet loss, and to facilitate the combination of video with audio quality. The approach developed by [1] is more similar to ours, since it also aims at a parametric video quality model. The model predictions they obtain cannot directly be compared with our model, since they used CIF resolution, quantified quality in terms of a degradation category rating (DCR) instead of the absolute ratings we



(a) HD.



(b) SD.

Fig. 3. Comparison of model predictions and test results. Also provided are the correlations and mean squared errors of the predictions, both for the average over content types (top values) and when content type is considered (bottom values).

have collected, and used another codec (H.263). The work described in [9] deals with audiovisual quality, but does not provide a bitstream- or parameter-based model for audio- and video-quality assessment.

4. DISCUSSION AND OUTLOOK

We have presented the framework for a parametric model for assessing the multimedia quality of services such as IPTV or VoD of standard and high definition. Based on a large series of subjective video tests, we have derived a first video model algorithm, which shows very good performance even for data that has not been used for model training. Future versions of the video model part need to include the effect of different content types (i.e. the impact of spatial and temporal complexity in case of transmission impairment due to low-bitrate coding or packet loss) and of key-frame rate. Based on a set of mixed SD-/HD-tests, a combined SD-/HD-model will be developed. Further video tests are underway to determine the quality impact of packet loss of non-uniform distribution. Since our model includes interactions between the input parameters bitrate, packet loss, codec type and packet loss concealment type, we are confident that the additivity assumed on the impairment factor level may be justified (see respective considerations for speech quality in [10]). Both audio- and audiovisual quality tests will serve to determine the modules for audio quality and audio-video quality interaction.

5. REFERENCES

[1] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," *IEEE TRans. Circuits Syst. Video Technol.*, vol. 12, no. 12, pp. 1071–1083, 2002.

[2] K. Brunnström and B. N. Schenkman, "Quality of video affected by packet loss distortion, compared to the predictions of a spatial-temporal model," in *Proc. SPIE, Human Vision and Electronic Imaging VII*, 2002, vol. 4662, pp. 77–80.

[3] A. Reibman and D. Poole, "Characterizing packet-loss impairments in compressed video," in *Proc. International Conference on Image Processing (ICIP 2007)*, 2007, pp. 77–80.

[4] ITU-T Rec. J.144, *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference*, Geneva, 2004.

[5] ITU-R Rec. BS.1387-1, *Method for objective measurements of perceived audio quality*, Geneva, 2001.

[6] ITU-T Rec. G.107, *The E-Model, a computational model for use in transmission planning*, Geneva, 2005.

[7] S. Pechard, M. Carnec, P. Le Callet, and D. Barba, "From SD to HD television: Effects of h.264 distortions versus display size on quality of experience," in *Proc. International Conference on Image Processing (ICIP 2006)*, 2006, pp. 409–412.

[8] ITU-T Rec. P.910, *Subjective video quality assessment methods for multimedia applications*, Geneva, 1999.

[9] S. Winkler and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 973–980, 2006.

[10] Alexander Raake, *Speech Quality of VoIP – Assessment and Prediction*, John Wiley & Sons Ltd, Chichester, West Sussex, UK, 2006.