Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2011, Article ID 629284, 14 pages
doi:10.1155/2011/629284

Research Article

Impairment-Factor-Based Audiovisual Quality Model for IPTV: Influence of Video Resolution, Degradation Type, and Content Type

M. N. Garcia, R. Schleicher, and A. Raake

Deutsche Telekom Laboratories, Berlin Institute of Technology, 10587 Berlin, Germany

Correspondence should be addressed to M. N. Garcia, marie-neige.garcia@telekom.de

Received 2 November 2010; Accepted 2 March 2011

Academic Editor: Khaled El-Maleh

This paper presents an audiovisual quality model for IPTV services. The model estimates the audiovisual quality of standard and high definition video as perceived by the user. The model is developed for applications such as network planning and packet-layer quality monitoring. It mainly covers audio and video compression artifacts and impairments due to packet loss. The quality tests conducted for model development demonstrate a mutual influence of the perceived audio and video quality, and the predominance of the video quality for the overall audiovisual quality. The balance between audio quality and video quality, however, depends on the content, the video format, and the audio degradation type. The proposed model is based on impairment factors which quantify the quality-impact of the different degradations. The impairment factors are computed from parameters extracted from the bitstream or packet headers. For high definition video, the model predictions show a correlation with unknown subjective ratings of 95%. For comparison, we have developed a more classical audiovisual quality model which is based on the audio and video qualities and their interaction. Both quality- and impairment-factor-based models are further refined by taking the content-type into account. At last, the different model variants are compared with modeling approaches described in the literature.

1. Introduction

In order to achieve a high degree of user satisfaction for current and upcoming video services like video on demand (VoD), internet protocol television (IPTV), and mobile television (MoTV), perceived quality needs to be estimated both in the network planning phase and as part of the service monitoring. Quality assessment can be achieved using audiovisual subjective tests or by instrumental methods, which yield estimates of audiovisual quality as perceived by the user. If properly conducted, quality tests with human subjects are the most valid way to assess quality, since it is about human perception. However, since subjective tests are time consuming, costly, and do not allow to assess the quality during real-time service operation, instrumental assessment methods are often preferred. Those instrumental methods are based on audiovisual quality models.

Several studies on audiovisual perception have been conducted starting in the 80s (summarized in Kohlrauch and van de Par [1]). However, the first audiovisual quality models to be found in the literature appeared as late as in the 90s. At this time, they addressed either analog degradations, such as audio and video noise—this is the case for Bellcore's [2, 3] and Beerends' models [4]—or compression artifacts, such as blockiness—this is the case for France Telecom's [5], NTIA-ITS' [6, 7], and Hands' [8] models. For an overview of audiovisual quality models covering analog and compression degradations, see [9]. The interest in modelling audiovisual quality is currently rising again, reflected, for instance, by standardization activities such as the Multimedia Phase II project of the Video Quality Expert's Group (VQEG), which intends to evaluate audiovisual quality models for multimedia applications (unfortunately, to the knowledge of the authors, no citable document describing the VQEG Multimedia Phase II has been published at the time of writing this paper). In addition, Ries et al. [10] and Winkler and Faller [11] have recently developed audiovisual quality models for mobile applications, but the reported model

versions do not yet cover the effect of transmission errors. This latter point is problematic since, in the case of the time-varying degradation due to transmission errors, the impact of audio and video quality on the overall audiovisual quality as well as their interaction might differ from the case of compression artifacts. In [12], Belmudez et al. address the impact of transmission errors in addition to compression and frame rate artifacts but for interactive scenarios and small video resolutions, which is not suitable for our application. None of the above-mentioned models addresses HD video, a format for which we expect video quality to play a more important role than for smaller formats. As a consequence, we have developed a new audiovisual quality model which covers all IPTV-typical degradations—mainly audio and video compression artifacts and packet loss—and which is applicable to both SD and HD. Based on the quality perception tests conducted during model development, we have analyzed the influence of the degradation type and of the audiovisual content on the quality impact of audio and video.

For modeling audiovisual quality, we will follow a new approach in which audiovisual quality is computed from audio and video impairment factors instead of audio and video qualities, as it is done in most previous studies. The impairment factors are the quality-related counterpart of technical degradations, that is, the transformation of technical degradations onto a perceptual quality scale in terms of impairments. In the following, we will use the term "impairment-factor-based"—or "IF-based"—for the model based on impairment factors, and the term "quality-based"—or "Q-based"—for the model based on audio and video qualities. The concept of impairment factors is based on the findings by Allnatt for broadcast TV [13], yielding the assumption that certain kinds of impairment factors may be considered as additive on an appropriate (perceptual) quality rating scale. This impairment factor principle has been adopted by the so-called E-model, a parameter-based network planning quality model standardized by the ITU-T [14] for speech services. More recently, it has been used in the so-called T-V-Model developed by our group [15, 16] for predicting video quality in the case of network planning and quality monitoring of IPTV services. NTT followed a similar approach in [17], but their model has been developed for interactive multimedia services such as video telephony, yielding psychological factors not applicable in the case of IPTV, such as "feeling of activity".

The remainder of this paper is structured as follows. Section 2 details the audio, video, and audiovisual subjective tests we conducted to obtain the data the models are based on. Test results are analyzed in Section 3, and the audiovisual quality models developed using the results are presented in Section 4. In this section, the impairment-factor-based models are evaluated against both known (training) and unknown (evaluation) subjective test data, and are compared with quality-based models trained on the same subjective data. The performances of our models are compared with the performances of other quality-based models as they are reported in the literature. Finally, in Section 5 we conclude and give an outlook on future modeling steps. This paper extends the work presented in [18] by providing a deeper insight on the comparison of the models' performance, by addressing the SD resolution in addition to the HD one, by sharpening the analysis of the degradation-type impact on audiovisual quality, and by analyzing the quality impact of the audiovisual content type.

Table 1: Audiovisual content descriptions.

ID	Video	Audio
A	Movie trailer	Speech on music
B	Interview	Speech
C	Soccer	Speech on noise
D	Movie	Classical music
E	Music video	Pop music with singer

2. Experimental Design

Audio, video, and audiovisual subjective tests have been conducted using audio-only, video-only, and audiovisual sequences, respectively. The source material consists of five audiovisual contents of 16 *s* duration each. Video-only and audiovisual tests were conducted separately for the two video resolutions SD and HD. The audiovisual contents are representative of different TV programs. The video contents differ in their amount of details and complexity of structures and movements, and the audio contents in terms of audio category and genre. The resulting audiovisual content types are described in Table 1.

In order to simulate typical IPTV degradations, the five source contents were processed offline according to the test conditions listed in Table 2. This results in 49 audio test conditions for each of the five audio contents, leading to 245 audio sequences to be rated by the subjects; 36 video test conditions for each of the five video contents, leading to 180 video sequences; 49 audiovisual conditions for each of the five audiovisual contents, leading to 245 audiovisual sequences. Apart from the audio-only test, all numbers are given separately per video resolution. As it is typical of IPTV services, we have used an MPEG2-TS/RTP/UDP/IP packetization scheme. Here, seven MPEG2 transport stream (TS) packets are contained in one RTP packet, and each contains either audio or video. For our tests, multiplexing was done for the already decoded files, instead of using ecologically valid multiplexing at TS-level. Note that this choice was made to ensure that the resulting model will be valid in a variety of situations with different levels of audio and video degradations. This is especially reflected in the combinations of loss rates, where different settings have been used for audio and video.

Listening and viewing conditions were compliant to ITU-T Recommendation P.800 [21], and Recommendations ITU-R BT-500-11 [22] and ITU-T P.910 [20], respectively. To ensure that the processed, but uncompressed, material could be played out without playback artifacts, professional high-performance systems were used for audio and video presentation. Between 23 and 29 subjects participated in

Table 2: Test conditions used in the audio-only, video-only and audiovisual tests. CBR: constant bit rate, PLR: uniform packet loss rate, processing was done generating different loss traces; PLC: packet loss concealment.

Parameters	Video	Audio
Video-/audio-only test		
Format	HD (1920 × 1080 pixels)	wav (48 kHz, 16 bit, stereo)
	SD (720 × 576 pixels)	
Codec	H.264; MPEG2	MPEG-2 AAC LC[a] (aac); MPEG-1 LII (mp2)
		MPEG-4 HE-AACv2[b] (heaac); MPEG-1 LIII (mp3)
CBR	H.264: {2, 4, 8, 16, 32} Mbps (HD)	aac: {48, 64, 96} kbps
	{0.5, 1, 2, 4, 8} Mbps (SD)	heaac: {32, 48, 64} kbps
	MPEG2: {4, 8, 16, 32, 64} Mbps (HD)	mp2: {48, 96, 192} kbps
	{1, 2, 4, 8, 16} Mbps (SD)	mp3: {64, 96, 128} kbps
PLR	{0, 0.02, 0.06, 0.125, 0.25}% (freezing)	{0, 1, 4, 8}% (frame loss ≡ 1 frame per packet)
	{0, 0.125, 0.25, 0.5, 1, 2, 4}% (slicing)	codec-built-in (for details, see [19])
PLC	Freezing[c]; slicing[d]	
Audiovisual test		
Format	HD (1920 × 1080 pixels)	wav (48 kHz, 16 bit, stereo)
	SD (720 × 576 pixels)	
Codec	H.264	MP2; AAC
CBR	{2, 4, 8, 16} Mbps (HD)	aac: 48 kbps
	{0.5, 1, 2, 4} Mbps (SD)	mp2: {48, 96, 192} kbps
PLR	{0, 0.02, 0.06, 0.25}% (freezing)	{0, 1, 4, 8}% (frame loss ≡ 1 frame per packet)
	{0, 0.125, 0.5, 4}% (slicing)	
PLC	Freezing; slicing	codec-built-in (for details, see [19])

[a]AAC: advanced audio coding; LC: low complexity.
[b]HE-AAC: high efficiency advanced audio coding.
[c]In case of packet loss, the picture freezes until the next intact I-frame arrives; the frames in between are skipped in our case.
[d]A slice typically corresponds to a certain area of the image that—if affected by loss—the decoder fills with data from the same, previous, or following video frame.

each test, and each subject was allowed to participate in only one test (audio, video, or audiovisual). An absolute category rating (ACR) was used for collecting subjective quality judgements. The subjects rated the quality using the continuous 11-point quality scale recommended in ITU-T Recommendations P.910 [20] and shown (attributes "schlecht", "dürftig", "ordentlich", "gut", and "ausgezeichnet" correspond to "bad", "poor", "fair", "good", and "excellent" in the English version of the scale) in Figure 1. The uncompressed original audio and video were used as hidden references in the tests, but the scores for the hidden reference were not subtracted from the scores, that is, no hidden-reference removal was applied.

3. Subjective Test Results

For each of the five subjective tests (one audio, two video, two audiovisual), the scores were averaged over subjects, yielding mean opinion scores (MOS), were linearly transformed to the 5-point ACR MOS scale by aligning the numbers of the scales, and further transformed to the 100-point model scale using the conversion defined in ITU-T Recommendation G.107 [14].

Note that in the following, and unlike [18], we do not average the ratings across all contents but per content. This choice is motivated by two reasons: (a) the audiovisual quality model is to be applied on audiovisual sequences with various contents, and a predicted quality value per sequence is required; we thus want to capture the quality variation due to content; (b) the audiovisual quality model developed for all contents, that is, with one set of coefficients valid for all contents, is to be compared to an audiovisual quality model with different sets of coefficients for each content.

In order to have a first impression of the quality impact of audio and video on the overall audiovisual quality, we conducted a correlation analysis, correlating the audio quality Qa, the video quality Qv, and their interaction $Qa \cdot Qv$ with the audiovisual quality Qav (see Table 3, column "All"). It can be observed that for both SD and HD, the interaction term is predominant (SD: correlation = 0.94; HD: correlation = 0.92). The video quality seems to have more impact on the overall audiovisual quality than the audio quality, especially for HD (SD: video correlation = 0.75, audio correlation = 0.51; HD: video correlation = 0.80, audio correlation = 0.47). This finding is expected, and it shows that the impact of video quality increases with the video format.

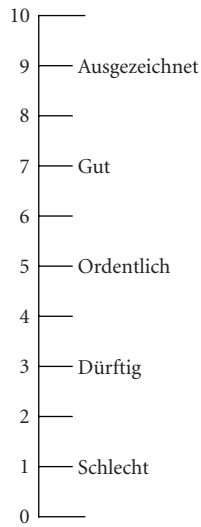

Figure 1: 11-point quality scale (ITU-T Recommendations P.910 [20]) used in audio, video, and audiovisual subjective tests.

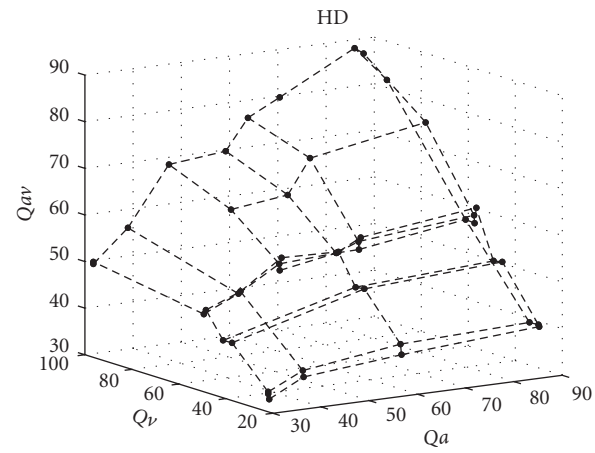

Figure 2: Audiovisual quality (Qav) as a function of audio quality (Qa) and video quality (Qv).

Moreover, the individual impact of the audio and of the video quality on the overall audiovisual quality depends on the quality of the other modality (video and audio). This is reflected by the slopes of the edges in Figure 2: the audio quality Qa has a decreasing influence on the overall HD audiovisual quality Qav for decreasing HD video quality. In turn, the HD video quality Qv has a less strongly declining influence on the overall HD audiovisual quality with decreasing audio quality. Similar observations have been made for SD. Note that, for the sake of clarity, Figure 2 shows the ratings averaged over all subjects and over all contents, that is the per-condition ratings instead of the per-sequence ratings.

Using the results for all contents might hide that for some contents, the above statements are not valid anymore. As a consequence, we computed the same correlations as above, but used ratings per content (see Table 3, columns "A" to "E"). For the contents "A" to "D", the same observation as for "all" contents can be made. For content "E" (music video), the quality impact of audio seems to be higher than for the other contents, and closer to the quality impact of video (SD: correlation $(Qav, Qa) = 0.61$; HD: correlation $(Qav, Qa) = 0.57$). This observation especially applies to SD, confirming the impact of the video format.

One more aspect to be considered is how the degradation type influences the quality impact of audio and video on the overall audiovisual quality. In our case, the employed degradation types were audio and video compression, audio frame loss, and video packet loss. We want to know, for instance, if for a given level of audio and video qualities we obtain different audiovisual quality values for audio compression than for audio packet loss, even though both have resulted in the same audio-only quality in the audio test. This aspect will be discussed further in the following modeling section.

4. Modeling

From now on, we will refer to the quality impact of audio and video degradations as impairment factors. Therefore, we define

(i) $I\text{cod}_X$: the quality impact of video ($X \equiv V$) or audio ($X \equiv A$) compression,

(ii) $I\text{tra}_X$: the quality impact of video- or audio-transmission errors, that is, video packet or audio frame loss.

4.1. Impairment Factors and Quality Models. As mentioned in the introduction, it is assumed that certain kinds of impairment factors may be considered as additive on an appropriate rating scale. Following this assumption, the audio- and video-only quality models are decomposed as follows (for details on the audio and video quality models see [15, 16, 23]):

$$Q_X = Qo_X - I\text{cod}_X - I\text{tra}_X, \tag{1}$$

where Q_X is the predicted audio or video quality, and Qo_X is the base quality level the transmitted audio or video signal can reach for the respective target service. In our experiments, Qo_X is the maximum quality rating obtained in the audio- or video-only subjective tests. $I\text{cod}_X$ thus is derived from subjective tests for transmission error-free conditions as follows: $I\text{cod}_X = Qo_X - Q_X$. Using all conditions, we obtain $I\text{tra}_X$ by computing $I\text{tra}_X = Qo_X - I\text{cod}_X - Q_X$.

4.2. Content and Quality Models. The influence of the content on the perceived quality plays a role at different levels. For instance, in the video-only case, it is well known that the quality impact of the bitrate is highly content dependent [24–28], especially at low bitrates. This result can

Table 3: Correlation of the audio and video quality, and their interaction with the overall quality (SD and HD).

	Qav	All	A	B	C	D	E
HD	*Qa*	0.47	0.49	0.46	0.45	0.45	*0.57*
	Qv	0.80	0.87	0.80	0.86	0.84	*0.69*
	$Qa \cdot Qv$	0.92	0.94	0.94	0.94	0.95	*0.92*
SD	*Qa*	0.51	0.57	0.52	0.48	0.48	*0.61*
	Qv	0.75	0.73	0.77	*0.81*	0.79	*0.67*
	$Qa \cdot Qv$	0.94	0.94	0.95	0.92	0.96	*0.96*

be captured by developing video quality models that are explicitly taking video content characteristics into account. In the present work, we focus on the influence of the audiovisual content on the balance between audio and video quality (see Section 3), and on how this variation can be captured in the audiovisual quality modeling. The impact of the video content on video quality has been addressed, for example, in [24–28], and respective models for audio quality are currently under study.

4.3. Audiovisual Quality-Based Model. Similarly to other studies [2–8, 10, 11, 17], we now model the audiovisual quality *Qav* based on the audio quality *Qa*, the video quality *Qv* and the interaction between *Qa*, and *Qv*

$$Qav = \alpha + \beta \cdot Qa + \gamma \cdot Qv + \zeta \cdot Qa \cdot Qv. \quad (2)$$

This model is called a "quality-based" model, or "Q-based" model. The coefficients α, β, γ, and ζ of (2) vary from one research to the other, depending on the application, the resolution of the video, and the audiovisual content. By applying the quality-based model on SD and HD ratings averaged over all subjects, we obtain the coefficients displayed in Table 4, rows "all". The content-based audiovisual quality model with different coefficients per content is obtained by applying the quality-based model to ratings averaged over all subjects for each content separately. The obtained coefficients are listed in rows "A" to "E" of Table 4.

The regression coefficients are compared taking into account their 95% confidence intervals: if the confidence intervals of two regression coefficients overlap, the regressions coefficients are considered to not be different. If the confidence interval of a coefficient overlaps the value zero, the regression coefficient is considered as nonsignificant, that is, not different from zero.

In our case for HD, and similarly to [8] for high-motion video, the dominance of the video quality over the audiovisual quality leads to $\beta = 0$. For SD, $\beta = 0$, and $\gamma = 0$, confirming that audio quality and video quality are more balanced for this resolution. This is in accordance with the observations made on the correlation values shown in Table 3, Section 3.

When modeling the per-content data (coefficients of rows "A" to "E"), we observe that the model pattern depends on the content. Indeed, for HD, γ is significantly different from zero for all contents except content E (music video). This result was expected from the observation we made on the correlations between audio and video qualities (see Section 3): the impact of audio and video quality is more balanced for content E than for the other contents. Similarly, we had observed in Section 3 that the audio and video quality was more balanced for SD than for HD. This balance is less respected in case of content C (soccer), for which the correlation between video and audiovisual qualities is higher than for the other contents. This is translated into a nonzero value of γ found in the regression analysis.

Table 4: Regression coefficients of the quality-based model for HD and SD, across all contents (rows "all") and per content (rows "A" to "E").

	α	β	γ	ζ
HD all	**28.49**	**0**	**0.13**	**0.006**
HD A	24.57	0	0.28	0.006
HD B	27.50	0	0.11	0.006
HD C	24.37	0	0.21	0.005
HD D	27.85	0	0.17	0.005
HD E	32.59	0	**0**	0.007
SD all	**30.99**	**0**	**0**	**0.006**
SD A	32.77	0	0	0.006
SD B	30.21	0	0	0.006
SD C	25.83	0	**0.15**	0.005
SD D	32.06	0	0	0.006
SD E	30.83	0	0	0.006

4.4. Audiovisual Impairment-Factor-Based Model. The advantage of the quality-based model variant is that it can easily be used with audio and video quality models coming from other laboratories provided they are based on similar types of network conditions and services, and deliver quality estimates on the same scale. The flipside to this advantage is that the quality-based model does not allow for a fine-grained diagnosis of the cause for nonoptimum quality. Indeed, using (2), we only know if a low audiovisual quality *Qav* is caused by a low audio quality *Qa*, a low video quality *Qv*, or both. For diagnostic purposes, we can compute the audio and video impairment factors $Icod_X$ and $Itra_X$ and thus, using (1), know what the audio (*Qa*) and video (*Qv*) quality impact due to audio and video degradations is. However, we do not know if these degradations have a similar

Table 5: Regression coefficients of the IF-based model for HD and SD, for all contents (row "all") and per content (rows "A" to "E").

	$Qavo$	c_{ac}	c_{vc}	c_{avcc}	c_{at}	c_{vt}	c_{avtt}	c_{avtc}	c_{avct}
HD all	**94.33**	**0.466**	**0.713**	**−0.008**	**0.652**	**0.712**	**−0.007**	**−0.009**	**−0.007**
HD A	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
HD B	94.33	0.539	0.814	−0.010	0.752	0.727	−0.009	−0.017	−0.008
HD C	94.33	0	0.786	0	0.685	0.724	−0.007	−0.012	0
HD D	94.33	0.416	0.851	−0.007	0.601	0.724	−0.007	−0.013	−0.007
HD E	94.33	0.560	0.519	−0.009	0.711	0.667	−0.008	−0.011	−0.009
SD all	**82.90**	**0.387**	**0.511**	**−0.004**	**0.539**	**0.507**	**−0.005**	**−0.006**	**−0.006**
SD A	82.90	0.333	0.411	0	0.471	0.523	−0.004	0	−0.008
SD B	82.90	0.510	0.521	−0.006	0.677	0.522	−0.004	−0.012	−0.007
SD C	82.90	0	0.657	0	0.567	0.462	−0.002	−0.010	0
SD D	82.90	0.324	0.472	−0.004	0.559	0.492	−0.005	−0.005	−0.004
SD E	82.90	0.309	0.398	0	0.613	0.484	−0.006	−0.007	0

impact in an audiovisual perception context. If we insert (1) for both audio and video into (2), we obtain the following

$$\begin{aligned} Qav = {} & (\alpha + \beta \cdot Qo_A + \gamma \cdot Qo_V + \zeta \cdot Qo_A \cdot Qo_V) \\ & - (\beta + \zeta \cdot Qo_V) \cdot I\mathrm{cod}_A - (\beta + \zeta \cdot Qo_V) \cdot I\mathrm{tra}_A \\ & - (\gamma + \zeta \cdot Qo_A) \cdot I\mathrm{cod}_V - (\gamma + \zeta \cdot Qo_A) \cdot I\mathrm{tra}_V \\ & + \zeta \cdot I\mathrm{cod}_A \cdot I\mathrm{cod}_V + \zeta \cdot I\mathrm{tra}_A \cdot I\mathrm{tra}_V \\ & + \zeta \cdot I\mathrm{tra}_A \cdot I\mathrm{cod}_V + \zeta \cdot I\mathrm{cod}_A \cdot I\mathrm{tra}_V. \end{aligned} \tag{3}$$

Identical coefficients in (3) imply a similar impact on audiovisual quality. This is, for example, the case for all interaction terms between impairment factors $I\mathrm{cod}_X$ and $I\mathrm{tra}_X$, which are all multiplied by the same coefficient ζ. Thus, this model assumes that all interaction terms between impairment factors have the same weight for audiovisual quality. Similarly, (3) suggests that for each modality (audio and video), the individual terms $I\mathrm{cod}_X$ and $I\mathrm{tra}_X$ with equal X (audio or video) have the same impact on audiovisual quality.

To verify the validity of this assumption, which we will call assumption "A", we express the audiovisual quality directly as a function of the impairment factors, leading to the following model:

$$\begin{aligned} Qav = {} & Qavo \\ & - c_{ac} \cdot I\mathrm{cod}_A - c_{at} \cdot I\mathrm{tra}_A \\ & - c_{vc} \cdot I\mathrm{cod}_V - c_{vt} \cdot I\mathrm{tra}_V \\ & - c_{avcc} \cdot I\mathrm{cod}_A \cdot I\mathrm{cod}_V - c_{avtt} \cdot I\mathrm{tra}_A \cdot I\mathrm{tra}_V \\ & - c_{avtc} \cdot Itra_A \cdot I\mathrm{cod}_V - c_{avct} \cdot I\mathrm{cod}_A \cdot I\mathrm{tra}_V. \end{aligned} \tag{4}$$

As for Qo_X in (1), $Qavo$ is the base audiovisual quality level. During the modeling, $Qavo$ is fixed to the maximum audiovisual quality rating obtained in our subjective tests. The name convention for the coefficients is as follows: the subscripts a, v, c, and t stand for *audio*, *video*, *coding*, and *transmission*, respectively. When c and t are both present in the coefficient name, the first of those two letters is related to audio, the second to video. As an example, c_{avct} represents the coefficient of the interaction between the *audio coding* impairment $I\mathrm{cod}_A$ and the *video transmission* impairment $I\mathrm{tra}_V$.

Note that the interactions between $I\mathrm{cod}_A$ and $I\mathrm{tra}_A$, and between $I\mathrm{cod}_V$ and $I\mathrm{tra}_V$, are implicitly taken into account by including them in $I\mathrm{tra}_A$ and $I\mathrm{tra}_V$ (see Section 4.5). As a consequence, (4) does not explicitly contain the interaction terms $I\mathrm{cod}_A \cdot I\mathrm{tra}_A$ and $I\mathrm{cod}_V \cdot I\mathrm{tra}_V$.

If the regression coefficients c_{ac} and c_{at}, or c_{vc} and c_{vt}, or c_{avcc}, c_{avtt}, c_{avtc}, and c_{avct} are significantly different, we can not validate assumption "A", that is, the respective impairments have the same impact on overall quality. As for the quality-based model, the regression coefficients are compared taking into account their confidence intervals.

Applying multiple regression analysis using the results of the audio-only, the video-only, and the audiovisual subjective tests with (4), we obtain the regression coefficients shown in Table 5, row "HD all" for HD, row "SD all" for SD. Due to processing issues (only one video file, present in both the video and audiovisual tests, was corrupted. However, it was crucial for computing the $I\mathrm{cod}_V$ value of several video files with transmission errors and having the same bitrate using the equation $I\mathrm{tra}_V = Qo_V - I\mathrm{cod}_V - Q_V$ as shown in Section 4.1.) the coefficients for the impairment-factor-based model could not be developed for content A, HD resolution.

Regression coefficients and their confidence intervals are displayed in Figures 3 and 4 for, respectively HD and SD. Significance-related information for the regression coefficients are shown in Table 6, rows "HD all" and "SD all". Coefficients not significantly different from zero and coefficients significantly different from the other coefficients are indicated in the columns "Nonsign. coeff." and "sign. diff. coeff.-pairs".

It can be seen, in both Figures 3 and 4 and Table 6, that all regression coefficients are significantly different from zero. A remarkable behavior can be observed in the case

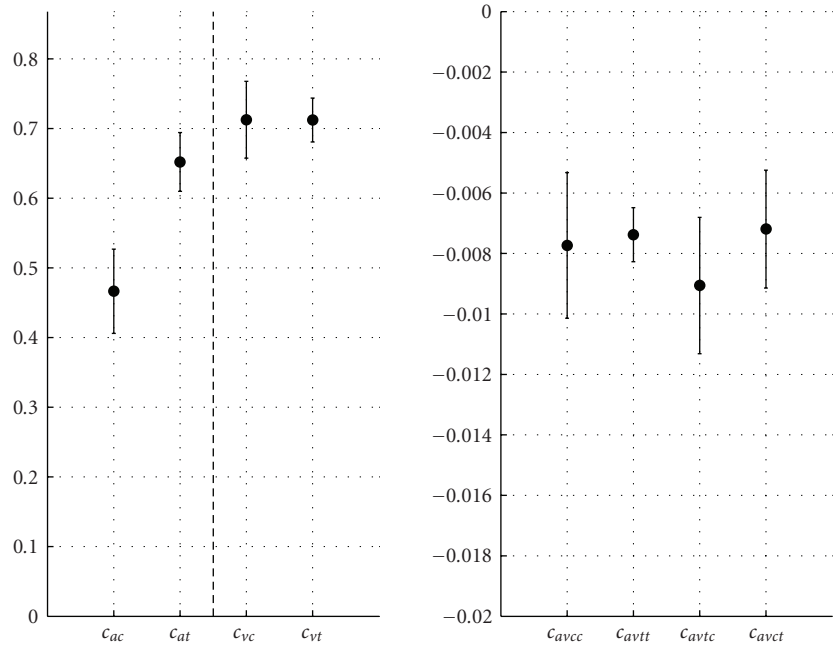

Figure 3: Regression coefficients and 95% confidence intervals. HD IF-based model.

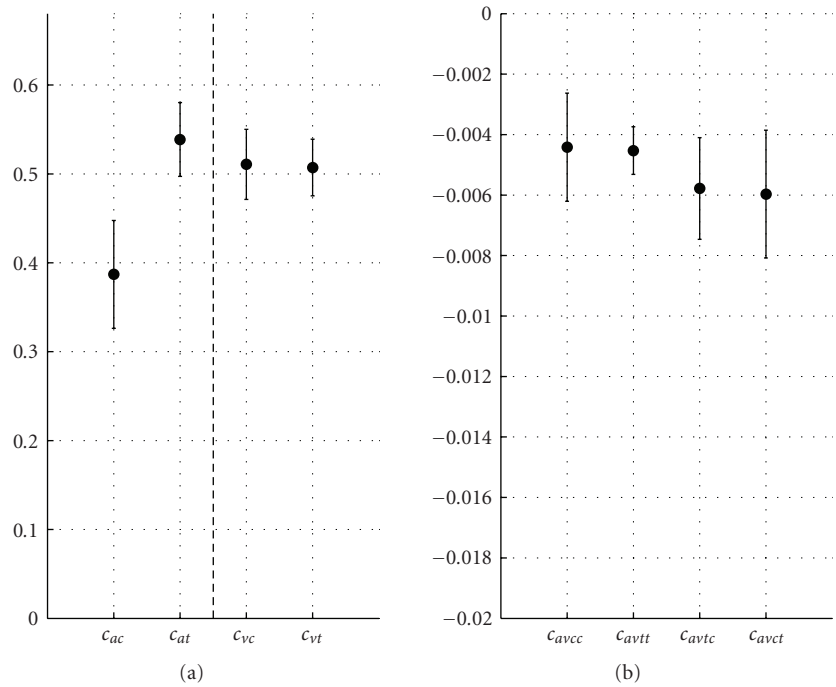

Figure 4: Regression coefficients and 95% confidence intervals. SD IF-based model.

of coefficients c_{ac} and c_{at}, which are linked to the quality impact of audio only. As apparent from Figures 3 and 4, and Table 6, these coefficients are statistically different both for SD and HD. This means that (a) audio quality alone shows a significant impact on audiovisual quality, when the coding- and packet-loss-related contributions to audio quality are separated, and (b) the impairment due to audio packet loss impacts audiovisual quality differently from that due to audio coding. Hence, when a coding-only audio impairment and a transmission-related audio impairment of equal value $I\text{cod}_A = I\text{tra}_A$ are presented to users in an audiovisual context, the packet-loss-related impairment plays a larger

Table 6: Significance of regression coefficients and of differences between regression coefficients, for HD and SD, from (4).

	Nonsign. coeff.	Sign. diff. coeff.-pairs
HD all	None	$\{c_{ac}, c_{at}\}$
HD A	N.A.	N.a.
HD B	None	$\{c_{ac}, c_{at}\}$
HD C	c_{ac}, c_{avcc}, c_{avct}	$\{c_{ac}, c_{at}\}$
		$\{c_{avcc}, c_{avtt}\}$, $\{c_{avcc}, c_{avtc}\}$
		$\{c_{avct}, c_{avtt}\}$, $\{c_{avct}, c_{avtc}\}$
HD D	None	$\{c_{ac}, c_{at}\}$
HD E	None	None
SD all	None	$\{c_{ac}, c_{at}\}$
SD A	c_{avcc}, c_{avtc}	$\{c_{avcc}, c_{avtt}\}$, $\{c_{avcc}, c_{avct}\}$
		$\{c_{avtc}, c_{avtt}\}$, $\{c_{avtc}, c_{avct}\}$
SD B	None	None
SD C	c_{ac}, c_{avcc}, c_{avct}	$\{c_{ac}, c_{at}\}$, $\{c_{vc}, c_{vt}\}$, $\{c_{avtc}, c_{avtt}\}$
		$\{c_{avcc}, c_{avtt}\}$, $\{c_{avcc}, c_{avtc}\}$
		$\{c_{avct}, c_{avtt}\}$, $\{c_{avct}, c_{avtc}\}$
SD D	None	$\{c_{ac}, c_{at}\}$
SD E	c_{avcc}, c_{avct}	$\{c_{ac}, c_{at}\}$
		$\{c_{avcc}, c_{avtt}\}$, $\{c_{avcc}, c_{avtc}\}$
		$\{c_{avct}, c_{avtt}\}$, $\{c_{avct}, c_{avtc}\}$

role for audiovisual quality than the coding-related one. Both of these effects cannot be captured by the quality-based model, where the audio-only quality was not found to have a significant impact on the overall audiovisual quality. Since c_{ac} is significantly different from c_{at}, assumption "A" can be rejected, confirming that the degradation type does have an influence on how the audio component impacts audiovisual quality. This supports the idea of impairment-factor-based modeling approach.

No significant difference was found between the regression coefficients of $I\text{cod}_V$ and $I\text{tra}_V$, and between the regression coefficients of all the interaction terms. This may mean that the impact of video on audiovisual quality is independent of the video degradation type, and that the impact of the interaction between audio and video qualities on the audiovisual quality is independent of the audio and video degradation types. This may also mean that the influence of the audio and video degradation types has been compensated by the influence of the content type during the modeling process. Indeed, if the analysis is done per content, $I\text{cod}_V$ is shown to have a higher impact on the audiovisual quality than $I\text{tra}_V$, but for another content, the opposite is observed; on average, $I\text{cod}_V$ and $I\text{tra}_V$ will have the same impact on audiovisual quality, and will thus not have significantly different regression coefficients.

All these results provide us with interesting insights into the subjects' attention in the context of audiovisual quality assessment. Indeed, in an audiovisual test the subjects seem to focus more on video, as in a video-only test, while the audio is only subconsciously attended to. With their main attention on the video, the subjects pay similar attention to stationary degradations such as compression artifacts as to more time-varying degradations such as transmission errors, just as in a video-only test. The users' attention is attracted more to the audio only in case of transient audio degradations such as audio frame loss. This may explain why—across contents—the coefficients of $I\text{cod}_V$ and $I\text{tra}_V$ are not significantly different while the coefficient of $I\text{tra}_A$ is significantly bigger than the one of $I\text{cod}_A$.

For investigating the impact of the content on audiovisual quality, we rerun the regression analysis on ratings averaged per content over all subjects. The obtained regression coefficients are shown in Table 5, rows "A" to "E". Coefficients not significantly different from zero are shown in column "Nonsign. coeff." of Table 6 for each resolution (referred to by "HD" and "SD"), and separately for each content (rows "A" to "E"). Moreover, we want to verify if assumption "A" still can be rejected when modeling the audiovisual quality per content. For this purpose, in column "sign. diff. coeff.-pairs", we indicate for each resolution and content, if $c_{ac} \neq c_{at}$, or $c_{vc} \neq c_{vt}$, or if one of the coefficients of the multiplicative terms of (4) is significantly different from any other.

It can be observed that, for some of the contents, some regression coefficients are nonsignificant (e.g., coefficient $c_{ac}, c_{avcc}, c_{avct}$ of content C for HD) but, for other contents, they are (e.g., contents B, D, and E for HD). This implies that different model patterns for different contents may increase the overall performance of the model. Moreover, for several contents, c_{ac} is significantly different from c_{at}, confirming that the audio-only quality does have an impact on the perceived audiovisual quality, and that this impact depends on the audio degradation type. This is especially true for content C, for both SD and HD, for which regression coefficients for the terms containing $I\text{cod}_A$ ($c_{ac}, c_{avcc}, c_{avct}$) are all nonsignificant, contrary to the regression coefficients of the terms containing $I\text{tra}_A$ (c_{at}, c_{avtt}, c_{avtc}).

In addition, c_{vc} is significantly different from c_{vt} for the content C (soccer) of the SD model, highlighting the importance of the video degradation type for this content on the overall audiovisual quality. Note that we already observed in the quality-based model that for SD and content C, $\gamma \neq 0$ in (2). The video-only quality and degradation type seem to play a bigger role for content C than for the other contents. Regarding the coefficients of the multiplicative terms ($c_{avcc}, c_{avtt}, c_{avtc}, c_{avct}$), they are significantly different for several contents (content C for HD, contents A, C, and E for SD). This confirms that assumption "A" needs to be rejected also when modeling the audiovisual quality per content. All those results are in favor of developing an impairment-factor based model, which, in addition, takes into account the audiovisual content type.

4.5. Estimation of Impairment Factors. In a real instrumental assessment situation, the impairment factors are computed from measurements done on the audio and video streams and not from subjective tests. The model input information can either be the decoded audio and video (i.e., input to a signal-based model) or information extracted from

the bitstream, or, in a more light-weight fashion, from transport-header information, requiring much lower processing resources. As input, our model takes parameters extracted from transport-header information, such as audio and video bitrates or packet-loss rates. A more detailed list of the employed parameters is given in the column "Parameters" of Table 2.

In a leastsquare curve fitting procedure using separately the subjective audio and video test results described in Section 2 as target values, we have identified the following relations for the different impairment factors $I\text{cod}_X$ and $I\text{tra}_X$:

$$I\text{cod}_X = a1 \cdot \exp(a2 \cdot \text{bitrate}) + a3, \tag{5}$$

$$I\text{tra}_A = (b0 - I\text{cod}_A) \cdot \frac{Pfl}{(b1 + Pfl)}, \tag{6}$$

$$I\text{tra}_{Vf} = (c0 - I\text{cod}_V) \cdot \frac{Ppl}{(I\text{cod}_V \cdot (c1 \cdot \mu + c2) + Ppl)} \tag{7}$$

$$I\text{tra}_{Vs} = d0 \cdot \log\left(d1 \cdot Ppl \cdot \text{bitrate}^{d2} \cdot \mu^{d3} + 1\right). \tag{8}$$

Here, ai, $i \in \{1,2,3\}$, bj, $j \in \{0,1\}$, ck, $k \in \{0,1,2\}$, dl, $l \in \{0,1,2,3\}$ are the curve-fitting coefficients. The coefficients ai depend on the used codec and on the video resolution. The coefficients bj depend on the audio codec and packet loss concealment. The coefficients ck and dl depend on the codec and on the video resolution. $I\text{tra}_{Vf}$ and $I\text{tra}_{Vs}$ are the quality impact due to video packet loss with, respectively freezing and slicing as packet loss concealment. Pfl is the audio frame-loss rate in percentage, and Ppl is the packet-loss-rate in percentage. μ is the number of video packets lost in a row. In the audiovisual tests we conducted, we used uniform loss.

Thus, for predicting the audiovisual quality of IPTV services using the impairment-factor-based model, we first extract parameters from the audio and video packet trace, insert these into (5), (6), (7), and (8), and finally insert the predicted impairment factors into (4).

It should be noted that changing the configuration of the video encoder, for example in terms of the group of picture (GOP) structure properties or the number of slices per frame, will affect the perceived quality. However, these changes do not introduce new types of degradations. As a consequence, they can be captured by simply modifying the video-quality model (Equations (5), (7), and (8)). For instance, additional parameters such as the GOP length could implicitly or explicitly be included in this model. As long as the changed settings do not introduce new types of degradations, there is no need to modify either of the two variants of the audiovisual quality model.

4.6. Model Evaluation. The impairment-factor- and quality-based models have been evaluated against the audiovisual subjective test dataset used for developing the model, the "training" dataset, as well as a subjective test dataset not used for training the model, the "evaluation" dataset. The latter contains sources B, C, and E listed in Section 2 as well as the processed versions of those videos, using the same conditions as listed in Table 2, except for the freezing conditions (due to processing issues, freezing packet loss concealment was present only in the anchor conditions, making the test database still balanced in terms of quality range and perceptual dimensions, but the ratings for freezing conditions could not be used for evaluating the model. Further, note that loss processing was done independently for the training and evaluation datasets, yielding different loss instances in the decoded audio and video). The same test procedure and set-up as the ones described in Section 2 were followed. 18 naïve subjects participated in the evaluation test.

Four model variants are compared for each resolution: the content-blind (Q) and -aware (Qc) quality-based models, and the content-blind (IF) and -aware (IFc) impairment-factor-based models. The content-blind models use the same set of coefficients for all contents (see rows "HD all" and "SD all" in Tables 4 and 5). The content-aware models use one set of coefficients per content (see rows "H" D B to E and "SD" A to E in Tables 4 and 5).

4.6.1. Performance Indicators. The performance of the models is evaluated by computing the Pearson correlation coefficient (R) and the so-called modified root mean square error (rmse*) between the predicted and the subjective quality values. These quality values have been previously converted from the 100-point model scale back to the 11-point scale used in the subjective tests by applying reverse transforms of the conversions described at the beginning of Section 3. rmse* has been used to evaluate the model candidates in the development of the new ITU-T standard for full-reference speech quality assessment P.OLQA (objective listening quality assessment, future ITU-T Recommendation P.863). It explicitly takes the degree of uncertainty of subjects' judgments into account and is defined as follows:

$$\text{rmse}^* = \sqrt{\frac{1}{N-d} \sum \text{Perror}(i)^2}, \tag{9}$$

with

$$\text{Perror} = \max(0, |Q(i) - Qp(i)| - ci_{95}(i)). \tag{10}$$

Here, N is the number of audiovisual sequences, i is the index of the audiovisual sequence, ci_{95} is the 95% confidence interval of the sequence i, Q is the subjective audiovisual quality, and Qp is the predicted audiovisual quality.

Since the rmse* is not commonly used in previous research work, the root mean square error rmse is also given. This may ease the comparison with the performance of other models in the literature.

The significance of the difference of the correlation and rmse* (but not rmse for clarity purposes) is further tested following the VQEG HDTV evaluation procedure described in [29].

Performance results are summarized in Tables 7 and 8 for, respectively, HD and SD, for the content-blind (Q) and, -aware (Qc) quality-based models according to (2), and

Table 7: Performance for HD, for the training ($_t$) and evaluation ($_e$) data; audio and video quality and impairment factors are derived from subjective tests (Subj.) or predicted from audio and video quality models (Pred.); *Q*: content-blind Q-based model, *Qc*: content-aware Q-based model, *IF*: content-blind IF-based model, *IFc*: content-aware IF-based model; in italic: the respective model performs significantly better than the corresponding basic model *Q*; in bold: significantly better performing model between *IF* and *IFc*.

Subj.	R_t	rmse_t	rmse_t^*	R_e	rmse_e	rmse_e^*
Q	0.94	0.68	0.30	0.94	0.71	0.28
Qc	0.94	0.60	*0.21*	0.94	0.64	*0.22*
IF	0.96	0.53	*0.15*	0.95	0.57	**0.14**
IFc	**0.98**	0.39	**0.07**	0.95	0.58	*0.24*
Pred.	R_t	rmse_t	rmse_t^*	R_e	rmse_e	rmse_e^*
Q	0.91	0.67	0.26	0.91	0.71	0.26
IF	0.93	0.63	*0.20*	0.92	0.70	0.25

Table 8: Performance for SD. See Table 7 caption for more details.

Subj	R_t	rmse_t	rmse_t^*	R_e	rmse_e	rmse_e^*
Q	0.94	0.56	0.20	0.92	0.65	0.26
Qc	0.95	0.52	*0.18*	0.92	0.64	0.26
IF	0.94	0.55	0.18	0.93	0.58	*0.20*
IFc	0.95	0.48	**0.14**	0.92	0.64	*0.22*
Pred.	R_t	rmse_t	rmse_t^*	R_e	rmse_e	rmse_e^*
Q	0.91	0.63	0.23	0.86	0.72	0.26
IF	0.91	0.65	0.28	0.87	0.73	0.25

for the content-blind (*IF*) and, -aware (*IFc*) impairment-factor-based models according to (4). Table 7 (resp., 8) shows the performance of the HD (resp., SD) audiovisual quality models on the training (R_t, rmse_t, rmse_t^*) and evaluation (R_e, rmse_e, rmse_e^*) dataset, when the impairment factors and audio and video qualities are either derived from subjective tests (section "Subj."), or predicted by the audio and video quality models defined in (5), (6), (7), (8) and (1) (section "Pred."). If a model performs significantly better than the content-blind quality-based model *Q*, the corresponding performance indicator (R_y, rmse_y^*, $y \in \{t, e\}$) is marked in italic; if one of the two impairment-factor-based models is performing better than the other, the respective performance values are printed in bold. Since the audio- and video-quality models are not content-dependent, the second part of Tables 7 and 8 only shows the performance numbers for the content-blind models *Q* and *IF*. Indeed, having one set of coefficients per content can be a benefit only if the predicted impairment factors and audio and video qualities are content dependent.

The subjective results ("Subj." in Tables 7 and 8) are used for validating the impairment-factor versus quality-based approach and the content-based approach versus content-blind approach while the data referred to as "Pred." are used for checking how robust our models are against the prediction error introduced by the audio- and video-quality models. Note that for both "Subj." and "Pred." parts, the audiovisual quality models have been trained on the audio and video qualities and impairment factors derived from the subjective tests, not quality values predicted by the audio and video quality models.

Figures 5 and 6 show the performance of the content-blind impairment-factor-based model on the evaluation dataset for HD and SD, when the impairment factors are derived from the subjective tests. This corresponds to the most valid way of evaluating the audiovisual impairment-factor-based model, since the evaluation data are unknown to the model, and for audiovisual quality prediction the model directly uses the subjective results from the audio- and video-only tests, instead of the audio- and video-only quality models with their possible prediction errors.

We will start the model performance comparison with general observations for all results, then continue by evaluating the benefit of taking the degradation type into account. In a third stage, we will analyze the advantage of considering the content type in the model. At last, we will analyze the robustness of the models against the prediction errors introduced by the audio- and video-quality models.

Performance indicators used in the following analysis are the Pearson correlation *R* and the rmse* values reported in Tables 7 and 8.

We can first observe that all model variants obtain good performance results, especially for HD, where the models always obtain correlations above 0.91, up to 0.98, rmse* is between 0.07 and 0.30 (on the 11-point scale used in the tests). The SD model variants obtain slightly lower performance, with correlation values ranging from 0.86 to 0.95, and rmse* is between 0.14 and 0.28. As expected,

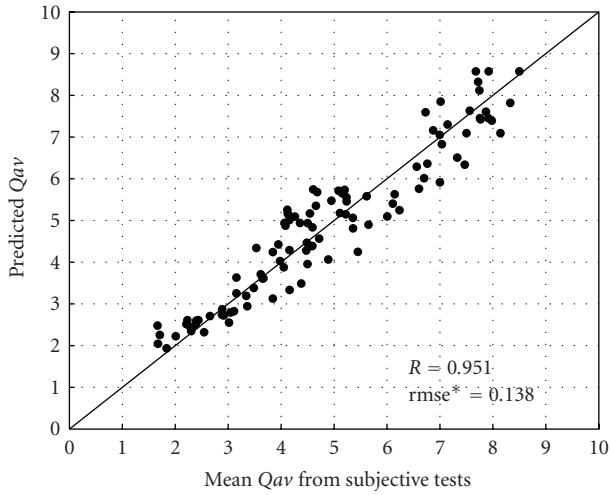

Figure 5: Performance of the content-blind HD impairment-factor-based model on unknown subjective data. Impairment factors are derived from the subjective tests.

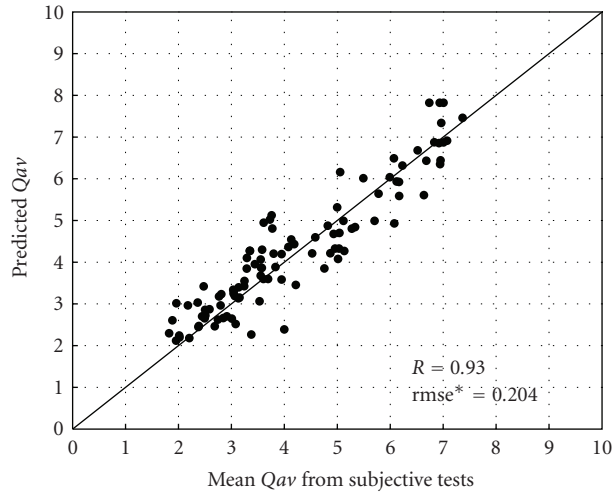

Figure 6: Performance of the content-blind SD impairment-factor-based model on unknown subjective data. Impairment factors are derived from the subjective tests.

the rmse* is the most discriminative performance measure between models.

4.6.2. Model Evaluation for HD. For HD, the content-blind impairment-factor-based model (*IF*) always performs better then the content-blind quality-based model (*Q*). This best performance is always significant, except for unknown data when the impairment-factors are predicted from the audio and video quality models. A possible explanation for this exception is the slightly lower performance of the audio and video quality models on the evaluation data compared to the training data. Since the rmse* takes into account the confidence interval of each sequence (see (10)), the slightly higher confidence interval values of the evaluation data compared to the training data ease obtaining good performance for all models and thus increase the difficulty of achieving significant difference between the rmse* of different models. A promising result is that the impairment-factor-based model variants *IF* and *IFc* in all cases perform better than the quality-based model variants *Q* and *Qc*. Considering the content in the modeling further improves the performance of the models in all cases except for the evaluation data with the impairment-factor-based model, this may be due to an overtraining of the model. Indeed, even though contents used in the evaluation dataset are identical to some of the contents of the training dataset, different conditions were used between the two sets. Moreover, the processing chains were different, yielding different perceptual impacts for similar conditions. As a consequence, the evaluation set can be considered to represent a case where we use different contents between the datasets. Since the coefficients are content specific, the prediction for one content can even be worse than when using the coefficient set of "row all", which were obtained using ratings from several contents.

4.6.3. Model Evaluation for SD. The advantage of using the impairment-factor-based approach is not as clear for SD as for HD. However, we can notice that the impairment-factor-based model *IF* performs better than the quality-based model *Q* in all cases, except for the training data, where the audio and video quality are predicted from the audio and video quality models. A small more detailed diagnosis shows that this exception may be explained by a lower performance of the video quality model and respective impairment factors on the training data ($R = 0.90$) than on the evaluation data ($R = 0.94$). This lower performance affects the impairment-factor-based audiovisual quality model *IF* more because, contrary to the quality-based model *Q* (see Table 4, row "SD all" and (2)), it contains video-only terms ($I\text{cod}_V$ and $I\text{tra}_V$, see Table 5, row "SD all" and (4)). In other words, the video quality prediction error may propagate more in case of the impairment-factor-based model than in case of the quality-based model. As in the case of HD, considering that the influence of the content further improves the performance of the models in all cases, except for the evaluation data with the impairment-factor-based model.

4.6.4. Comparison with Models Described in the Literature. We wish to compare the performance of our models to the performance of models described in the literature. Note that, in almost all studies, the models are quality based, their performance is computed using the training dataset, and the audio and video quality terms *Qa* and *Qv* of the models are fed with the subjective test values. Since validating the models on unknown data is considered to be more suitable, we prefer to show the performance of our best-performing models, that is, the content-blind impairment-factor-based models for both SD and HD, on the evaluation dataset. As a comparison point, we also depict the performance results for our content-blind quality model variants. This leads to the

Table 9: Models performance comparison, "Cod." compression artifacts, "trans.": transmission errors. In italic: model performance evaluation on training data.

Model	Degradation type	α	β	γ	ζ	Correlation[a]	Scale[b]
Bellcore93 [2]	Analog	1.33	0	0	0.11	*0.99*	9
Bellcore94 [3]	Analog	1.07	0	0	0.11	*0.99*	9
NTIAITU98 [6]	Cod.	1.54	0	0	0.12	*0.93*	9
NTIAJones98 [7]	Cod.	−0.677	0.217	0.888	0	*0.98*	5
Beerends99 v1 [4]	Analog	1.12	0.007	0.24	0.09	*0.98*	9
Beerends99 v2 [4]	Analog	1.45	0	0	0.11	*0.97*	9
FT98 [5]	Compression	1.76	0	0	0.10	*0.96*	9
BT04 (head and shoulder) [8]	Cod.	1.15	0	0	0.17	*0.85*	0–100
BT04 (high motion) [8]	Cod.	0.95	0	0.25	0.15	*0.82*	0–100
Ries07 (fast movement) [10]	Cod.	−0.922	0.569	0.506	0.170	*0.91*[c]	5
Ries07 (video call) [10]	Cod.	−0.631	0.214	0.012	0.118	*0.90*[c]	5
Winkler06 v1 [11]	Cod., frame-rate	−1.51	0.456	0.770	0	*0.94*	11
Winkler06 v2 [11]	Cod., frame-rate	1.98	0	0	0.103	*<0.94*	11
NTT05 [17]	Compr., frame-rate, delay	N.A.	N.A.	N.A.	N.A.	*0.94*	5
Impairment-based T-V-M[d] HD	Cod., trans.	N.A.	N.A.	N.A.	N.A.	0.95	11
Quality-based T-V-M[d] HD	Cod., trans.	28.49	0	0.13	0.006	0.94	11
Impairment-based T-V-M[d] SD	Cod., trans.	N.A.	N.A.	N.A.	N.A.	0.92	11
Quality-based T-V-M[d] SD	Cod., trans.	30.99	0	0	0.006	0.91	11

[a]Correlation coefficients are assumed to be Pearson correlation coefficients.
[b]Number of categories.
[c]Audio and video quality predicted from models.
[d]Proposed model.

correlations listed in Table 9. Degradation types addressed by each model are also shown, indicating that all other data has been obtained without considering transmission errors.

The content-blind impairment-factor-based model obtains high correlation values, similar to most of the other models. This is even more valuable since our model can be applied to both coding and transmission errors, that is a wider range of degradation types. However, since the models from the literature have been derived for different video formats and applications, comparing correlation coefficients does not allow any conclusions to be drawn on which model performs the best, but rather gives us an indication of relative performance of our model. Moreover, we did not have access to the rmse* (neither the rmse) for most of the models found in the literature. This measure would have been more appropriate for comparing the models, since, as previously mentioned, it is more discriminative, also in the light of the underlying test data.

A brief summary of the results discussed here in detail is given at the beginning of the conclusion.

5. Conclusions and Outlooks

Based on the results of five quality perception tests, we have presented different audiovisual quality models for IPTV services for each SD and HD video resolution: a content-blind and a content-aware quality-based model, and a content-blind and a content-aware impairment-factor-based model. By definition, the content-blind models use the same set of coefficients for all contents while the content-aware models have one set of coefficients per content (see Tables 4 and 5). All models have been developed using the same subjective test data. Based on a correlation analysis of the test results (see Table 3) and the comparison of regression coefficients for different model variants, we have shown that both the audiovisual content type and the degradation type have an influence on the perceived audiovisual quality, with different effects between SD and HD.

As shown by regression analysis of our data in terms of the quality-based model, the audiovisual quality interaction plays the main role for audiovisual quality, both in case of SD and HD. However, a clear difference can be observed for the role of the video-only quality: while a nonzero coefficient was found for one content only in case of SD, all but one content lead to nonzero coefficients for video-only quality in the case of HD. Obviously, the video part has more importance in this case.

The advantage of an impairment-factor-based rather than a quality-based approach could be substantiated by our regression analysis for both SD and HD, mainly due to the more fine-grained inclusion of audio: while audiovisual quality was not found to be dependent on the degradation type for video, it was shown to be more affected by audio frame loss than by audio coding, in spite of the equal role of the two degradation types for audio-only quality.

This difference is assumed to be due to the video-only-like perception mode in an audiovisual context, where the users' attention is explicitly drawn to the audio quality only when transient events such as loss events occur.

These findings are directly linked with the performance of the respective models: the SD and HD content-blind impairment-factor-based models perform better than the other models on unknown data with, for HD, a Pearson correlation of 0.95 and an rmse of 0.57 on the 11-point scale used in the subjective tests. Both impairment-factor-based variants perform better than the quality-based variants, and they provide a more fine-grained diagnosis of the audiovisual quality.

However, the proposed models have some limitations: when the audio and video qualities and impairment factors are predicted from audio and video quality models, the impairment-factor-based variants are less robust to audio and video quality prediction errors than the quality-based variants. More studies are necessary for identifying the thresholds of audio and video quality prediction errors at which the impairment-factor-based variants start to perform worse than the quality-based variants.

The main limitation may, however, be the fact that content-dependent models also require content-specific datasets from which they are derived. Of course the question arises at what point of specificity to stop to avoid overfitting of the models and to cease the otherwise neverending task of subjective tests. We tried to overcome this limitation by focussing on the content types that so far appear to be the most popular ones broadcasted via IPTV: movies, sports, music videos, and so forth. With regard to video quality degradation type, we would like to differentiate slicing and freezing for the interaction between video and audio qualities. At last, more analyses are necessary for extending the audio- and video-model components to more diverse degradations such as other loss distributions, video and audio encoder and decoder settings, and the audiovisual model to audiovisual synchronization artifacts. Since the impairment-factor-based approach was developed for a range of coding and loss settings, however, it is expected that it will be applicable to many of these cases as well.

References

[1] A. Kohlrausch and S. van de Par, "Audio-visual interaction in the context of multi-media applications," in *Communication Acoustics*, J. Blauert, Ed., pp. 109–138, Springer, Berlin, Germany, 2005.

[2] ITU-T SG12 COM 12-20 (Bellcore, USA), "Experimental Combined Audio/Video Subjective Test Method," December 1993.

[3] ITU-R SG12 COM 12-37 (Bellcore, USA), "Extension of Combined Audio/Video Quality Model," September 1994.

[4] J. G. Beerends and F. E. De Caluwe, "Influence of video quality on perceived audio quality and vice versa," *Journal of the Audio Engineering Society*, vol. 47, no. 5, pp. 355–362, 1999.

[5] N. Chateau, "Relations between audio, video and audiovisual quality," Contr COM 12-61 to ITU-T Study Group 12, 1998.

[6] ITU-T SG12 D.038. (NTIA/ITS, USA), "Results of an audio-visual desktop teleconferencing subjective experiment," February 1998.

[7] C. Jones and D. J. Atkinson, "Development of opinion-based audio-visual quality models for desktop video-teleconferencing," in *Proceedings of the 6th IEEE International Workshop on Quality of Service*, May 1998.

[8] D. S. Hands, "A basic multimedia quality model," *IEEE Transactions on Multimedia*, vol. 6, no. 6, pp. 806–816, 2004.

[9] J. You, U. Reiter, M. H. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio-visual services: a survey," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482–501, 2010.

[10] M. Ries, R. Puglia, T. Tebaldi, O. Nemethova, and M. Rupp, "Audiovisual quality estimation for mobile streaming services," in *Proceedings of the 2nd International Symposium on Wireless Communications Systems (ISWCS '05)*, pp. 173–177, September 2005.

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

[12] B. Belmudez, S. Moeller, B. Lewcio, A. Raake, and A. Mehmood, "Audio and video channel impact on perceived audio-visual quality in different interactive contexts," in *Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP '09)*, October 2009.

[13] J. Allnatt, *Transmitted-Picture Assessment*, John Wiley & Sons, Chichester, UK, 1983.

[14] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning," 2005.

[15] A. Raake, M. N. Garcia, S. Möller et al., "T-V-model: parameter-based prediction of IPTV quality," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 1149–1152, April 2008.

[16] M.-N. Garcia and A. Raake, "Parametric packet-layer video quality model for IPTV," in *Proceedings of the 10th International Conference on Information Sciences Signal Processing and Their Applications (ISSPA '10)*, 2010.

[17] K. Yamagishi and T. Hayashi, "Analysis of psychological factors for quality assessment of interactive multimodal service," in *Human Vision and Electronic Imaging X*, vol. 5666 of *Proceedings of SPIE*, pp. 130–138, 2005.

[18] M. N. Garcia and A. Raake, "Impairment-factor-based audio-visual quality model for IPTV," in *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEx '09)*, pp. 1–6, July 2009.

[19] B. Feiten, A. Raake, M.-N. Garcia, U. Wüstenhagen, and J. Kroll, "Subjective quality evaluation of audio streaming applications on absolute and paired rating scales," in *Proceedings of the 126th AES Convention*, 2009.

[20] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," 1999.

[21] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality," 1996.

[22] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," 2002.

[23] A. Raake, M. N. Garcia, B. Feiten, and S. Möller, "Parametric quality prediction for IP-based audio," in *Proceedings of the 155th Meeting of Acoustical Society of America (Acoustics '08)*, 2008.

[24] S. Pechard, D. Barba, and P. Le Callet, "Video quality model based on a spatio-temporal features extraction for H.264-coded HDTV sequences," in *Proceedings of the Picture Coding Symposium (PCS '07)*, 2007.

[25] Y. X. Liu, R. Kurceren, and U. Budhia, "Video classification for video quality prediction," *Journal of Zhejiang University: Science*, vol. 7, no. 5, pp. 919–926, 2006.

[26] M. Ries, C. Crespi, O. Nemethova, and M. Rupp, "Content based video quality estimation for H.264/AVC video streaming," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 2670–2675, March 2007.

[27] A. Khan, L. Sun, and E. Ifeachor, "Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, June 2009.

[28] M.-N. Garcia, R. Schleicher, and A. Raake, "Towards a content-based parametric video quality model for IPTV," in *Proceedings of the 3rd International Workshop on Perceptual Quality of Systems (PQS '10)*, 2010.

[29] VQEG, "Report on the validation of video quality models for high definition video content," Tech. Rep., VQEG, 2010.