Alexander Raake, Jörgen Gustafsson, Savvas Argyropoulos, Marie-Neige Garcia, David Lindegren, Gunnar Heikkilä, Martin Pettersson, Peter List, and Bernhard Feiten

IP-Based Mobile and Fixed Network Audiovisual Media Services

Current approaches for quality monitoring

This article provides a tutorial overview of current approaches for monitoring the quality perceived by users of IP-based audiovisual media services. The article addresses both mobile and fixed network services such as mobile TV or Internet Protocol TV (IPTV). It reviews the different quality models that exploit packet-header-, bit stream-, or signal-information for providing audio, video, and audiovisual quality estimates, respectively. It describes how these models can be applied for real-life monitoring, and how they can be adapted to reflect the information available at the given measurement point. An outlook gives insight into emerging trends for near- and mid-term future requirements and solutions.

Digital Object Identifier 10.1109/MSP.2011.942472
Date of publication: 1 November 2011

INTRODUCTION

IP-based video streaming or broadcast services such as video on demand (VoD), Web video (e.g., Youtube), IPTV, and mobile TV gain increasing popularity. To ensure that a service is of the desired high quality, methods are required for planning, optimizing, monitoring, and maintaining the service performance. Performance is often assessed in terms of quality of service (QoS), i.e., technical performance indicators at the protocol layer such as lost, dropped, or resent packet information and delay statistics, or at the lower layers information from streaming server logs, digital subscriber line access multiplexer

 1053-5888/11/$26.00©2011IEEE

(DSLAM) information (such as link errors), radio error information (such as transmission block errors), or information from the client site based on set-top box (STB) or mobile device logs.

TO ENSURE THAT A SERVICE IS OF THE DESIRED HIGH QUALITY, METHODS ARE REQUIRED FOR PLANNING, OPTIMIZING, MONITORING, AND MAINTAINING THE SERVICE PERFORMANCE.

However, ensuring the correct functioning of the technical system is no longer sufficient, especially when the QoS performance indicators, as they are, for example, defined in the service level specifications, tolerate certain quality degradations for efficiency reasons. As a consequence, it is indispensable to assess performance in terms of user-perceived quality, often referred to as quality of experience (QoE). In that case, monitoring the QoE from the end user's perspective allows a final judgement to be made on how well the service fulfills the promises made towards the user.

Quality can be defined as the "result of [the] judgment of the perceived composition of an entity with respect to its desired composition" [1], and is internal to the user. In an ultimately valid way, it can be measured only in perception tests with human subjects. Such tests are time- and resource-consuming, and cannot be carried out to continuously monitor the quality of a running service with its numerous users. In this case, a better alternative is to use quality models that map QoS-related performance indicators, or representations of the transmitted multimedia signals, to user-perceived quality as obtained from subjects in laboratory tests.

The assessment of media-signals transmitted over telecommunication links gained broader attention when the telephone had long transformed into a large-scale service. In the 1930s and 1940s, audio quality was addressed mainly in terms of how well speech could be understood, resulting in one of the first audio-quality related models, the Articulation Index [2]. A more holistic view on speech quality was taken in the 1960s, when aspects of naturalness were addressed [3], and interactive voice communication under delay was assessed [4]. For an overview of audio and speech quality assessment see [5] and [6]. The assessment of television system quality dates back at least until the 1950s [7], with increasing research activity between the 1960s and 1980s (see [8] for an overview). Image and video quality models were first proposed in the 1970s and 1980s (see, e.g., [9]). First integration functions for combining audio and video quality estimates to an estimate of audiovisual quality were proposed in the early 1990s (see [10] for an overview).

The early assessment methods consider the degradations typical of analog systems, such as signal attenuation and noise. With the advent of digital and packet-based transmission, speech, audio, and video coding as well as error-prone transmission channels have led to new types of quality degradations, such as coding or packet loss artifacts, which require novel assessment methods. Several models have been developed in the meantime, which address the degradation of today's audiovisual media transmission systems.

Based on such models, monitoring solutions have been proposed that enable information from different locations of the end-to-end delivery chain to be converted into estimates of QoE. Such methods can be used in service probes for active or passive monitoring. These probes often form a part of a global service monitoring framework, which allows a link to be established with diagnostic QoS information. Moreover, this framework provides a service-wide picture of QoE and the sources of respective problems, for example, in terms of QoS fault isolation or performance diagnosis [11], [12].

PERCEPTUAL QUALITY TESTS

Audio, speech, video, and audiovisual quality tests should be conducted using methods recommended by the International Telecommunication Union (ITU) or European Broadcasting Union (EBU). The use of standardized methods ensures that tests are reproducible and comparable between different laboratories. Commonly used and established methods are described in P.800 [13] for speech; BS.1116 and BS.1534 [14], [15] for audio; BT.500, BT.710, P.910, and EBU-SAMVIQ for video [16]–[19]; and P.911 and P.920 for audiovisual quality tests [20], [21].

These recommendations describe different dedicated test methods, specifying aspects such as the employed rating scales and the task for which they are most appropriate (see, e.g., Table 2 in [16]). Furthermore, the standards provide guidance on the test material to be used in the test (type, duration, etc.), on the test environment and set-up, on the number of subjects to be recruited and possible methods of subject screening.

Single-stimulus methods such as the Absolute Category Rating (ACR) best reflect the everyday usage situation of watching or listening to multimedia sequences. Such methods allow a high number of conditions to be rated in a short time but are less sensitive to small quality differences. Paired or multiple comparison methods such as the Degradation Category Rating (DCR), BS.1116, Multiple Stimuli with Hidden Reference and Anchor (MUSHRA), or Subjective Assessment Methodology for Video Quality (SAMVIQ) enable more discriminative test results but can be more time-consuming than single-stimulus methods, and they do not reflect the everyday multimedia service usage either. Comparisons of methods and scales commonly used in video quality tests are reported in [22]–[24].

Hidden or direct reference stimuli are often used, typically provided by the nondegraded source sequence. The test file duration is commonly five to eight seconds for speech and ten to 16 seconds for video and audiovisual sequences, but may be several minutes when continuous evaluation methods are applied, e.g., Single-Stimulus Continuous Quality Evaluation (SSCQE) [16].

Perceptual quality tests may suffer from bias. Zielinski et al. [25] provide an overview of possible types of bias for audio listening quality tests. One type is the "response-mapping bias" that reflects how subjects map their opinion to the quality rating scale, for example, due to the quality range and degradation types the test stimuli cover. Another source of bias can be the test interface. When conducting perceptual tests, it is vital to know these problems and to take measures to avoid them (see [25] and [26]). To the authors' knowledge, no comparable study is available for video. However, due to the fundamental perspective of [25], it can be assumed that it also applies to other modalities. An entry point into the literature on comparing video quality test methods is, for example, [24].

MODEL DEVELOPMENT USING QUALITY TEST RESULTS

Quality model development can be split into the following two parts: 1) definition of the model input and identification of respective features and 2) pooling and mapping of these features to an integral quality index. For a better separation and selection of features in terms of perceptual dimensions, it can be useful to run tests for multidimensional analysis as in [27], or qualitative tests as in [28].

In the service monitoring case, the selection of the input features depends on the available data: the signal/media, the bit stream, or both signal and bit stream (see the section "Instrumental Quality Models"). It is obvious that the best suitable input features for multimedia quality models are those that can directly be related with user perception. In case of a large number of candidate features, the finally used set may need to be identified using multivariate analysis techniques. Typical features represent (temporally and/or spatially) local artifacts or degradation due to coding or packet loss. Typical examples of pooling are: spatial pooling for image or video quality in terms of artifacts across individual frames (see e.g., [29] for image quality); temporal pooling of features or of short-term quality to an integral or longer-term quality estimate, see [6], [30], and [31] for examples for speech and video, respectively. Note that when several features associated with perceptually different types of degradations coexist, the model developer has to decide how they should best be combined. In [8] and [32], it is assumed that features that are related with a certain type of quality impairment can be mapped onto an appropriate (perceptual) quality rating scale, and that the resulting *impairment factors* I_i are additive on this scale

$$Q = \max\left(Q_{\min}\left[Q_0 - \sum_{i=1}^{n} I_i\right]\right), \qquad (1)$$

with Q as the integral speech, audio, or video quality; Q_0 as the best possible base-quality in a given context; $Q_{\min}$ the lowest quality level possible on the respective model scale; and $I_i = g_i(f_i)$ the impairment factors calculated using a set of functions $g_i(\cdot)$ of the feature vectors f_i, with $I_i \in [Q_{\min}, Q_0]$.

Instead of an additive formation of integral quality, other models use a multiplicative combination of the terms associated with individual impairment types, for example, in the form

$$Q_{\text{int}} = Q_0 \cdot \left(1 - \prod_i I_i\right), \qquad (2)$$

where, by design, the features f_i are mapped to impairment values I_i in $[0,1]$ using a different set of functions $h_i(f_i)$.

PERFORMANCE EVALUATION OF QUALITY MODELS

Quality models aim at predictions that match perceived quality ratings as closely as possible. To ensure a guaranteed performance of a given model, two different approaches exist: Standardization of the model, as it is typically done by ITU-T or the Video Quality Experts Group (VQEG), or standardization of a procedure and testplan for model validation, as recently proposed by the IPTV Interoperability Forum (IIF) of the Alliance for Telecommunications Industry Solutions (ATIS) [33]. Model standardization within ITU or VQEG is typically conducted as a competition, where a number of proponents compete to determine the winning and then standardized model (as in case of the POLQA or PESQ development, [34], [35]), or as a partial or full collaboration. See the section "Instrumental Quality Models" for more examples of respective ITU standards. In the case of the ATIS IIF approach, not the model, but the validation procedure and criteria are standardized. Then, an independent lab conducts the respective quality tests, and evaluates the model performance. In turn, in case that the model is standardized, the test data is typically provided by the proponents or an independent lab group.

In both cases, the performance of a quality model is evaluated using statistical metrics. Common metrics are the Pearson correlation coefficient (R) and the root mean square error (rmse). R is the linear correlation between the predicted and the subjective quality values on a scale from -1 to 1. The rmse measures the difference between the two data sets in terms of per-condition or per-stimulus mean square error and can be expressed on the model output or quality test scale.

In the development of ITU-T's latest full reference (FR) speech quality model POLQA (see the section "Signal-Based Models"), the modified rmse (rmse*) was used instead of rmse (see [34] for details). The performance criteria rmse and rmse* will be employed for evaluating the quality monitoring models P.NAMS and P.NBAMS currently developed by ITU-T Study Group 12. The rmse* explicitly includes the variation in the subjects' ratings

$$\text{rmse}^* = \sqrt{\frac{1}{N-d} \cdot \sum_{i=1}^{N} P_{\text{error}}(i)^2} \qquad (3)$$

$$P_{\text{error}}(i) = \max\left(0, |\,\text{MOS}(i) - \widehat{\text{MOS}(i)}| - ci_{95}(i)\right). \qquad (4)$$

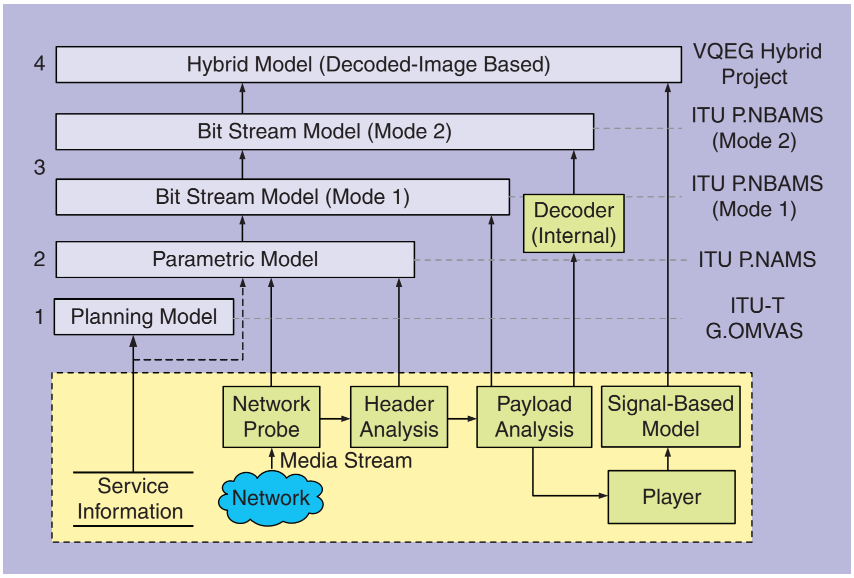
Here, N is the number of samples; d a correction term for the case that a dth-order mapping was used to map the subjective

[PREDICTING THE INTEGRAL QUALITY OF MULTIMEDIA SERVICES INVOLVES ESTIMATING AUDIO, VIDEO AND AUDIOVISUAL QUALITY.]

test data and model predictions; i is the index of the media sample; and ci_{95} the 95% confidence interval for that sample. MOS(i) and MOS(i) are the subjective and predicted quality, respectively.

INSTRUMENTAL QUALITY MODELS

Multimedia quality assessment models can be described in terms of the level at which input information is extracted. For example, information may stem from protocol headers or payload-related bit stream data. As depicted in Figure 1, models can be categorized into planning, parametric, signal-based, bit stream-based, or hybrid models, as detailed in the following sections. In addition, the models can be classified according to the amount of information they need from the original signal into no reference (NR), reduced reference (RR), and FR models. FR models have access to the original source sequence, which is compared with the processed sequence. RR models use the processed sequence together with a set of parameters extracted from the source sequence. NR models calculate quality predictions only based on the processed sequence. An overview of audiovisual quality models is given in Table 1 and in the following text. For more details on speech quality models, see [36].



[FIG1] Overview of quality assessment models that exploit different levels of information extracted from the media stream. Note: Planning and parametric models require additional service information, which they cannot extract from stream information (e.g., codec, bitrate, and loss characteristics for planning models; PLC or slicing information for parametric models; see text for details).

SIGNAL-BASED MODELS

Signal-based models exploit the decoded signal as input to calculate a quality score. Many of these models include aspects of human perception. Several ITU recommendations for signal-based models exist. In the speech quality domain, the FR-model P.862 (PESQ) [35] has been widely used over several years. The recently developed P.863 (POLQA) [34] is expected to supersede it. A respective NR speech quality model is described in [37]. For FR audio quality assessment, BS.1387 (PEAQ) [39] has been standardized.

In the video quality domain, J.144 [40] specifies models for FR SDTV quality assessment, J.341 [42] for FR HDTV

[TABLE 1] OVERVIEW QUALITY MODELS, SEE TEXT FOR DETAILS.

MODALITY	FR	RR	NR	P	BS(1)	BS(2)	IMPAIRMENT	OUTPUT	EXAMPLES
S	X						C, L	QUAL.	P.862/PESQ [35], P.863/POLQA [34]
S			X				C, L	QUAL.	P.563 [37]
S			X	X			C, L	QUAL.	G.107/E-MODEL [32], P.564 [38]
A			X	X			C, L	QUAL.	G.OMVAS, P.NAMS
A	X						C	QUAL.	BS.1387/PEAQ [39]
V	X						C	QUAL.	J.144 [40]
V	X						C, L	QUAL.	J.247 [41], J.341 [42]
V		X					C, L	QUAL.	J.246 [43]
V			X	X			C, L	QUAL.	P.NAMS; [44]–[49]
V			X	X			C, L, R	QUAL.	P.NAMS; [50]
V			X		X	X	C, L, R	QUAL.	P.NBAMS (MODE 1, MODE 2); R FOR MOBILE CASE
V			X		X		C	QUAL.	[51]–[53]
V			X			X	C	QUAL.	[54]
V	X[1]	X[1]	X[1]		X		L	VIS.	[55], [56]
V			X			X	L	MSE	[57]
A/V			X	X			C, L	QUAL.	G.1070 [58]; P.NAMS; [10]; G.OMVAS

S: Speech; A: Audio; A/V: Audiovisual; V: Video; P: Parametric/planning; BS(1): Bit stream mode 1, ITU-T P.NBAMS activity; BS(2): P.NBAMS Bit stream mode 2. Impairment: [C: Coding; L: Loss; R: Rebuffering]; Output: [Qual.: Quality; Vis.: Visibility; MSE: Mean Squared Error]. X1:Respective models available in NR, RR, and FR versions.

quality assessment, and J.246 and J.247 [41], [43] for RR and FR video quality assessment, respectively. See [9] for an overview of additional signal-based models and measures such as peak signal to noise ratio (PSNR). So far no audiovisual signal-based model has been standardized. Although the accuracy of the quality estimation for the video models is high, their high processing demands make them less well suited for larger-scale live network monitoring. Instead, they are better suited for off-line assessment or active monitoring in specific end-point test equipment. RR and FR models, which have the best accuracy of the signal-based models, must also deal with the nontrivial problem of synchronizing the reference signal to the decoded signal (see, e.g., [59] for a comparison of some video-related methods).

THE VISIBILITY AND THUS VIDEO QUALITY IMPACT OF PACKET LOSS HIGHLY DEPENDS ON DECODER PLC, AND THE SPATIOTEMPORAL COMPLEXITY OF THE CONTENT.

PARAMETRIC AND PLANNING MODELS

Parametric quality models predict the impact of coding and IP network impairments on multimedia QoE. Models of this type do not access the packet payload; instead, they use information extracted from packet headers, i.e., transport stream (MPEG-2 TS), Real-Time Transport Protocol (RTP), or packetized elementary stream (PES) headers, depending on the level of encryption. Parametric models do not explicitly address aspects such as the source quality or encoder or player implementations. To include these, service information (Figure 1) can be used to chose from different sets of curve-fitting coefficients. Parametric audio, video, and audiovisual quality models are currently developed and standardized in ITU-T Study Group 12 under the provisional name P.NAMS. Two different P.NAMS models are developed; one for the low bit-rate application area including mobile TV, and one for the high bit-rate application area that includes services such as IPTV.

Planning models can be considered as a variant of parametric models, where the input information is not acquired from an existing service, but is estimated based on service information available during the planning phase. The ITU-standardized E-model [32] is a prominent example for speech services. Another example is ITU-T Rec. G.1070 [58], a planning model for interactive video-telephony services. A corresponding model for streaming media is currently developed by ITU-T under the provisional name G. OMVAS.

Parametric models that are applied for service monitoring must provide media sequence-related quality estimates. In both mobile and fixed services, the quality degradations captured during monitoring relate to compression, packet loss, and delay, with delay leading to dropped packets or client rebuffering. How these degradations are handled by parametric models is discussed next.

COMPRESSION

Degradation of audio or video quality due to compression is mainly determined by the number of bits allocated to different audio or video frames, respectively. For video, frame rate and resolution are additional factors that impact quality. In case of video compression, the bit allocation mainly depends on

- the video resolution and frame rate
- the employed codec type and profile
- the codec implementation
- the targeted bit-rate
- the group of picture (GOP) structure
- the spatiotemporal complexity of the transmitted video content.

In case of audio, the bit allocation depends on the spectral bandwidth and the dynamics of the signal. Speech codecs such as Adaptive Multirate (AMR) employ a speech source model as the basis for quantization, whereas audio codecs such as Advanced Audio Coding (AAC) and MPEG-1 Layer 3 (MP3) exploit auditory masking.

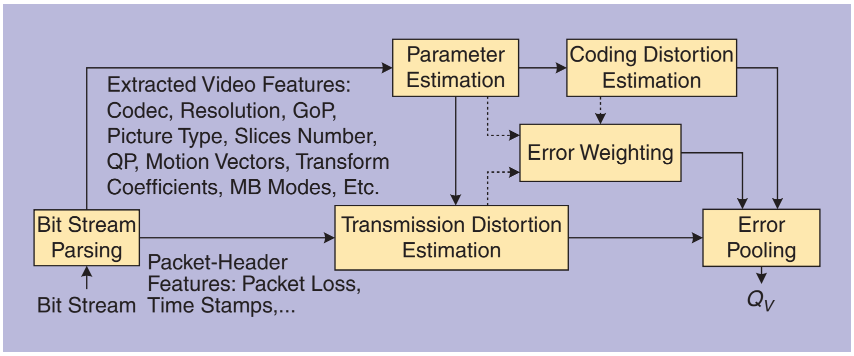
The media resolution, sample rate, and employed codec type and profile are usually known in terms of service information and do not need to be extracted from packet headers (see Figure 1). At transport level, the spatiotemporal complexity of the video content is unknown. However, since high spatial complexity results in larger I-frames, while high temporal complexity usually yields larger P- and B-frames, a prediction of video complexity from frame size statistics was proposed [60].

Most parametric video quality models estimate the impact of video compression using only, for a given resolution, the bit-rate, codec type, and profile, and possibly the GOP structure [44]–[46]. In the case of mobile services, the video frame-rate is used in addition [58] (wireline services such as IPTV have a fixed frame-rate). Audio quality models typically use the codec type and bitrate as input. The quality is then basically modeled as an exponential, logarithmic, or power function of a combination of the respective input parameters.

PACKET LOSS

The packet loss rate is commonly used in the literature as model input for describing the amount of lost packets [44], [46]. Bursty losses are described with the average number of packets lost in a row, the burst density (the fraction of packets lost in a burst) and the burst duration [47], [48], or the packet loss frequency [45]. For audio, consecutive losses can affect the efficiency of the packet loss concealment (PLC) method, which can typically handle one or two subsequently lost audio frames.

The quality impact of packet loss highly depends on the PLC method employed by the decoder. Video PLC can be classified as slicing or freezing. A slice typically corresponds to a certain area of the picture that is encoded independently of other areas in that picture. If affected by loss, the decoder conceals the affected slice area with data from adjacent regions or frames, or motion compensated content. In case of freezing concealment, the picture freezes until the next intact reference frame. Simple frame repetition may be used for audio PLC. More advanced methods make predictions based on surrounding audio frames, and typically introduce less frame-boundary-related artifacts.



[FIG2] Architecture of a bit stream-based video quality assessment model.

In general, increasing the number of slices increases the robustness of the video sequence to packet losses (since it reduces the spatial extent of the loss), but reduces the coding efficiency, due to added slice headers and interrupted predictive coding at slice boundaries. Information such as the PLC, or the number of slices per frame in case of transport streams cannot be extracted from packet headers, and thus has to be available to the parametric model as side information. In case of slicing, the spatiotemporal extent of the loss depends on the frame type, slice size and number of packets per frame. The loss-related degradation propagates until the next intact reference frame that does not reference other erroneous frames. If the frame type can be identified, the loss-impact can be estimated using parameters such as the number of distorted (invalid) frames, see, e.g., [49].

REBUFFERING

Rebuffering for video and audio is an important degradation caused by packet loss or a too-narrow streaming bandwidth. This parameter is common in YouTube-type streamed audiovisual media and Web radio or Spotify-like audio. The decoder decides whether to use rebuffering or freezing/slicing when a late loss occurs. YouTube will rebuffer until the data is received to continue playback. Hence, if one packet is lost or simply late, the client will ask for a retransmission until a sufficient portion of the sequence is received to continue playback. A parametric video quality model comprising rebuffering is described in [50]. If a precise measurement is desired, rebuffering information must be measured in the client and reported back to the monitoring model location. If this is not achievable, it is possible to estimate buffer behavior depending on the acquired packets and the encoded media bitrate, providing a rough estimate of the rebuffering degradation.

VIDEO BIT STREAM MODELS

In addition to transport layer information, bit stream quality models exploit information from the elementary stream. Video bit stream quality models for IPTV and mobile applications are currently being standardized by ITU-T in the Study Group 12 workitemP. NBAMS. Twodifferentmodesareunderconsideration:in Mode 1, the model does not fully decode the payload, but only parses the bit stream and extracts features; in Mode 2, the model can fully decode the bit stream (including the inverse transformation), and use the additional information from the reconstructed pixel values.

The architecture of a bit stream quality assessment model is depicted in Figure 2. Initially, the bit stream headers are parsed to extract transport-related information such as TS and/or RTP time stamps and sequence numbers for packet loss detection. Additionally, the payload of the video bit stream is parsed, and different features are extracted. The most important features are the picture type, the number of slices, the quantization parameter (QP), the motion vector and type of each macroblock (MB) and its partitions, and the transform coefficients of the prediction residual. Subsequently, these features are further processed to compute the inputs to the coding distortion and transmission distortion estimation modules, and to determine the parameters employed in the perceptual error weighting module, which can include aspects of the human visual system, such as visual attention. Typically, a bit stream model separately evaluates the degradation due to compression or preprocessing of the sequences and the distortions due to network impairments. These distortions can be locally evaluated (e.g., at the frame or GOP level) and are pooled to provide a unique estimation of the quality of the video sequence (Q_V in Figure 2).

In the following, a number of example model algorithms are summarized; some general considerations on quality monitoring including content information can be found in [61]. A coding error estimation method based on the distribution of transform coefficients was proposed in [52]. In this approach, the perceptual model is based on the spatiotemporal contrast sensitivity function (CSF) applied to the transform domain for local error weighting. Furthermore, the spatial information (SI) and temporal information (TI) parameters defined in P.910 [18] are explicitly employed for the coding distortion assessment in [54]. Another method based on the extraction of spatial and temporal features from the bit stream and the prediction of video quality based on blocking and flickering artifacts was presented in [51]. Content-dependent parameters were selected from a set of spatiotemporal features derived from the video bit stream, such as the AC transform coefficients, QP, and the motion vectors in the method presented in [53].

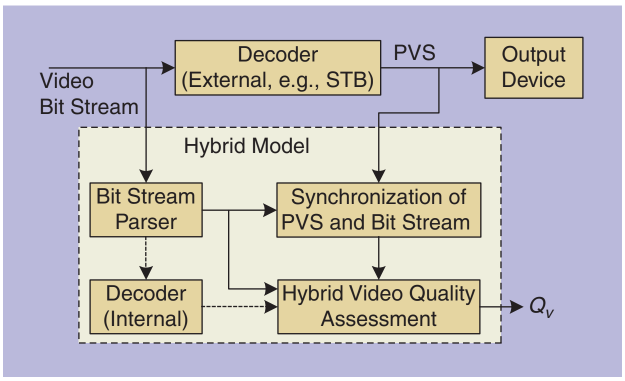
For the evaluation of quality degradations due to network impairments, the models should take into account that packet losses are not equally important and produce a different distortion depending on the pattern of loss events, their position within the sequence, the video content, and neighboring information that may mask the induced error. The visibility of packet loss in H.264/AVC streams was investigated in [55]. Based on different features extracted from the video bit stream, a general linear model (GLM) was proposed to predict the visibility of packet loss degradation.

The concept of mean time between failures was introduced in [62] to evaluate video quality based on failure statistics. The automatic video quality (AVQ) method evaluates compression and transmission artifacts, using the quantization step size and activity in scenes, and the perceptibility of video artifacts [56]. Finally, a no-reference method for quality monitoring based on the estimation of the mean squared error distortion at the MB level was presented in [57].

In [63], a model that directly incorporates motion information in an information theoretic framework was presented. Furthermore, visual attention is modeled in [64] from motion and contrast information for error-weighting based on saliency maps and pooling of global and local quality.

HYBRID VIDEO QUALITY ASSESSMENT MODELS

The block diagram of a hybrid video quality assessment model is depicted in Figure 3. Similar to the bit stream models, which may perform full decoding of the received bit stream, the hybrid video quality assessment models exploit information from the packet headers, the elementary stream, and the reconstructed pictures. The critical difference is that the information for the reconstructed pictures is obtained from the processed video sequence (PVS) generated by an external decoder (e.g., an STB in case of IPTV) and not from an internal decoder within the model. Thus, the reconstructed pictures used in the model are identical to the ones observed by the viewers, and there is no mismatch between the assumed and the actual packet loss handling and concealment method. Nevertheless, the model may also employ an internal decoder (like in Mode 2 of video bit stream models), since certain useful decoder-level information is not available from closed systems such as most STBs. If the bit stream features are not time-aligned with the picture-related features, the quality predictions become inaccurate. Thus, a crucial step in the hybrid video quality model is the time-alignment of the video bit stream with the decoded reconstructed images. Hybrid quality assessment models are currently examined in the Hybrid Perceptual/Bit stream project of VQEG [65]. In addition, the Joint Effort Group (JEG) is a new activity of VQEG that proposes an alternative, collaborative action for NR hybrid video models [66].



[FIG3] Block diagram of a hybrid video quality assessment model.

3-D VIDEO QUALITY MODELS

Models of three-dimensional (3-D) video quality must—in addition to the quality attributes for two-dimensional (2-D) video—consider 3-D-related quality aspects including depth perception, the naturalness of a scene, and the visual discomfort. Additional artifacts may also be present such as geometric distortions due to camera limitations [67], color differences, and spatial and temporal offsets between the left and the right view, for example, due to the employed (2-D-based) loss concealment. Different transmission techniques are currently being discussed for 3DTV applications, such as (the stereoscopic) frame-packing, which typically means side-by-side inclusion of two views in one frame and subsequent 2-D encoding, video plus depth encoding, or multiview coding (MVC) [68]. In case of stereoscopic coding approaches, instrumental assessment methods as for 2-D can be adopted [69]; the quality of each view can be calculated with a 2-D video model, and an additional modeling approach is used for the depth-related coding and transmission distortions. In case of video plus depth encoding, the specific reconstruction distortions have a very strong impact on quality and must be addressed with completely new types of models.

AUDIOVISUAL QUALITY MODELS

Audiovisual quality depends on the audio quality, the video quality, their interaction, and audiovisual impairments such as lip-sync problems. The simplest solution for estimating audiovisual quality is to use a function of audio and video quality estimates, and calculate the audiovisual quality score regardless of what type of degradation affects the audio and video scores. Linear regression with or without interaction between audio quality Q_A and video quality Q_V were proposed as quality integration functions (e.g., [10] and [70]), with respectively chosen coefficients α_i to calculate audiovisual quality Q_{AV}

$$Q_{AV} = \alpha_1 + \alpha_2 \cdot Q_A + \alpha_3 \cdot Q_V + \alpha_4 \cdot Q_A \cdot Q_V. \tag{5}$$

More detailed predictions can be achieved by using intermediate audio and video features that underlie the overall audio and video scores, and map these to an audiovisual quality score. The increased accuracy, however, comes with an increased complexity of the model. For a comparison of approaches, see [10].

MONITORING WIRELINE AND MOBILE SERVICES

The following description focuses on the application of quality models for monitoring QoE of IP-based audiovisual services. Respective tools consider the impact of IP network impairments on the quality of mobile audiovisual streaming and IPTV applications over transport formats such as User Datagram Protocol (UDP), RTP, Transmission Control Protocol (TCP), and Hypertext Transfer Protocol (HTTP), employing further protocols and data formats such as MPEG-2 TS, and various Internet Engineering Task Force (IETF) payload formats.

Three main criteria can serve for classifying and selecting the best suitable approach in a given context:

- *Target application area*
  - The service to be assessed: on demand streaming, broadcast, interactive, real-time communication etc.
  - The implementation of the service in terms of the protocols, audio and video resolutions, codecs, the manufacturer(s), and respective (proprietary and nonproprietary) implementations
- *Implementation of the assessment approach*
  - The locations of the quality model and of the probe for input data acquisition along the distribution chain
  - The monitoring may be done during service operation in a nonintrusive, i.e., passive manner, or off-line, as active (intrusive) monitoring, with dedicated equipment that inserts test traffic
- *Employed quality model*
  - The modality the model applies to: audio, video, or audiovisual
  - The type of information available to the model, for example, packet headers, encoded bit stream information, decoded signals, and the amount of reference information used (NR, RR, and FR models).

The quality model may be deployed directly in the measurement probe that acquires the model input information, or independently in a different location. Possible locations are the end-user devices or other midnetwork monitoring points. The locations of the model and measurement probe determine the mode of operation. Four such modes of operation are considered by ITU-T Study Group 12 in the P.NAMS development. They are referred to by a two-letter code, each one respectively for the probe and the model location, with three possible location designation letter choices: N for network (head-end, server, or another point along the delivery chain), C for client, and B for both. The four modes of operation are:

NN Probe and model are located in the network. The model can use known service information (see Figure 1) and streaming data captured at this point.

BN Probe located in the network and client—model located in the network. In this case, the information about the end point needs to be collected through measurement reporting protocols such as [71] and [72].

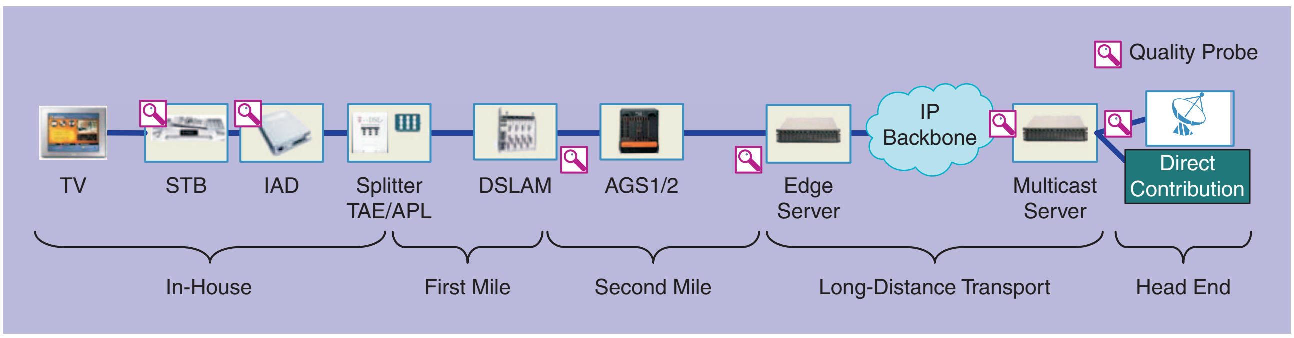
CN Probe is located in the client—model is located in the network. All model input information other than service information known in prior is measured at the client and transmitted to the model through measurement reporting protocols.

CC Probe and model are located in the client.

Another important issue for service deployment is how to increase error resilience. Common methods are forward error correction (FEC) with redundant information for error correction added to the stream, and automatic repeat request (ARQ), where missing information is requested and resent from a server. Depending on where in the delivery chain the error correction method is deployed, and where the monitoring information is captured, error correction has to explicitly be taken into account for QoE predictions. Two principle approaches are conceivable to capture this case:

- The packet stream is converted into a corrected stream that reflects an assumed behavior of FEC or ARQ.
- The input parameters/features to the model are converted into corrected values that simulate the effect of prior application of FEC or ARQ.

For different services and network structures, an adequate selection of measurement points is required (see Figure 4). The available information may vary, and not all model types can be applied in all situations. For example, if selective decryption is not efficiently feasible, the quality of encrypted media streams can only be assessed with parametric models using header information.



[FIG4] IPTV end-to-end delivery chain. Potential monitoring points are indicated by the "quality probe" symbols.

FIXED-NETWORK MEDIA SERVICE MONITORING

A typical wireline IPTV end-to-end service architecture for linear TV is depicted in Figure 4. It consists of a head end; a central play-out service (here based on a multicast server); an IP backbone for long-distance transport; some additional distributed servers at the network edges; the second mile aggregation network and switches (AGS); the first mile access network with the DSLAM; the Internet access devices (IADs); and the in-house network with the user's end device.

AN ESSENTIAL ENTITY OF THE QUALITY MONITORING SYSTEM IS A CENTRAL SERVICE MANAGEMENT THAT STEERS THE DISTRIBUTED PROBES AND COLLECTS AND DIGESTS THE MEASUREMENT RESULTS.

A complete monitoring system typically provides various types of information on system performance. The following description concentrates on monitoring the IPTV streams in terms of audiovisual quality. A quality monitoring probe can be applied at different stages of the delivery chain; see Figure 4. In the head end, an FR signal-based or hybrid model can provide an optimal basis for comparing measurement results with that of following measurement points. The next potential measurement points are located after the play-out server, at the different edge servers, at the AGS, and at the DSLAM. Here, parametric or bit stream quality models could be applied. Probes can also be placed in the homes, inside the IAD, in the STBs, or as an additional box linked between the IAD and the STB, monitoring all TV streams requested from the STB. The best placement of the probes depends on the implementation details of the whole system.

An essential entity in the architecture of the quality monitoring system is a central service management that steers the distributed probes and collects and digests the measurement results. The IPTV-service example is based on multicast transmission of RTP packets carrying MPEG-2 transport streams. For optimizing the channel switching time, parts of the stream are sent as unicast traffic coming from the edge server. Lost packets are retransmitted via unicast as well (ARQ), and need to be considered during QoE monitoring. A further challenge is that the payload of the MPEG-2 transport stream is encrypted for many of the streams. The quality monitor probe has to support the dissection of the whole IP traffic and extract the IPTV-related packets from it. The different IPTV streams are then analyzed in depth. Depending on the mode of operation, parametric, bit stream, or hybrid models according to the section "Instrumental Quality Models" are employed at one or several locations.

The quality model needs to be tailored to the processing power of the device it is implemented in. A parametric model can run on the IAD, while on the STB, a bit stream or even hybrid model could be implemented (CC mode). Alternatively, a lightweight probe on the client can send information to a more complex model located in the network (BN, CN modes). For a broad deployment, a solution based on standards is required. The upcoming P.NAMS and P.NBAMS standards together with appropriate standards for the communication between the end-user devices and the service monitoring systems, such as TR-069 and TR-135 [71], [73], optimally support the integration of QoE monitoring solutions in the IAD and STB. In this way, the STB or the IAD can measure what the end user actually experiences, and these measurements can be collected by the service management system in a cost-efficient manner. As an alternative to a probe or model on the STB or IAD, an additional box can be used in between these devices that sends quality reports to the service management system. Because of the extra costs, such end-user probes are only placed in exemplary users' homes, or in cases where a service engineer has to solve specific problems.

MOBILE MEDIA SERVICE MONITORING

For mobile media, the same type of monitoring as in the fixed case is principally possible, with quality probes located at different nodes in the distribution network. However, most of the quality impact is usually related to problems in the last wireless hop, which makes network-internal measurements less useful for assessing user-perceived quality. Hence, the only way to get a complete picture of QoE is to make the measurements directly in the mobile client, and report them back to the system. For 3GPP devices, QoE reporting was standardized in Release 6 for RTP-based streaming, and extended in Release 8 and 9 to HTTP streaming, progressive download and multimedia telephony (MMtel) [74], [75].

During a multimedia session, the mobile client continuously collects quality-related measurements, and periodically reports these back to a QoE reporting server that is run by the operator or the service provider. By using the reported QoE metrics as input to a parametric multimedia model, a good estimate of the end-user quality can be obtained. The data collected by the QoE reporting server is typically combined with other network or node measurements so that the cause of any quality degradation can be identified. Note that each mobile client reports QoE metrics regarding the received media, so to get the full picture, the QoE server can combine the reports from different clients.

Practical implementations typically address RTP-based streaming, HTTP streaming, and MMtel. The mobile client can be preconfigured via Open Mobile Alliance Device Management (OMA-DM) [76] with a default QoE reporting configuration. This default configuration is used by the mobile client if no session-related QoE configuration is specified. Session-related QoE reporting could, for instance, be based on a QoE configuration in the session initiation sequence provided by the RTP streaming server. The default configuration can specify that QoE reports shall be sent for selected sessions, all sessions, or for a random subset (for example, 5%). It is possible to specify which metrics are reported, and how often these are measured. In this way, the QoE reporting can be fine-tuned to achieve a good understanding of service QoE, without consuming unnecessary uplink radio capacity.

FUTURE TRENDS AND CHALLENGES

The standardization of parametric and bit stream-based as well as hybrid models will allow current monitoring requirements to be complemented by standardized models. Respective ITU-T standards are expected to be available by mid-2012.

ALTERNATIVE METHODS ARE ESPECIALLY REQUIRED FOR 3-D MEDIA QUALITY ASSESSMENT.

In the meantime, the media landscape is changing and expanding very rapidly, with more complex and rich service offerings as well as converged mobile-fixed solutions brought to the general public. People will expect the same or even better media quality as with the legacy solutions, and efficient and accurate methods for assessment of end-user service quality is one key element in this scenario.

MODEL DEVELOPMENT

One dimension of evolution is seen in the model development as such. The first aspect along these lines considers the ecological validity of the subjective test data that underlie all current models: to guarantee controlled laboratory settings, ecologically valid scenarios are not taken into account when asking subjects for quality ratings. However, the design and context of most tests and thus the respective model predictions differ considerably from how multimedia services are used in practice. Short sequences ranging from ten to 30 s duration are typically used instead of more TV- or video-typical viewing durations; the lab-test environment is physically treated to yield controlled optical and acoustical conditions, and thus lacks the naturalness of home or outdoor media consumption; the modeling data consists of conscious quality ratings rather than more hedonic user reactions, and so on. These drawbacks are deliberately chosen for the benefit of highly controlled settings and thus reproducible results. However, extrapolation or mapping to the actual usage scenario will help service providers and manufacturers to have a clearer picture of the relevance and value of certain model predictions. In [77], a first attempt was made along these lines.

Alternative methods are especially required for 3-D media quality assessment. While current test methods and models mainly provide fidelity-type data, more hedonic features or aspects of the "fidelity of the experience" such as media immersion need to be addressed. For 3-D, this is of cardinal relevance: for example, for video at a fixed bitrate, a viewer in a lab-test is likely to prefer a 2-D version over a 3-D version in terms of sheer image quality, but at home may still chose the 3-D video due to its higher hedonic appeal.

Accordingly, also for quality modeling, 3-D audio and video will be subjects of further research and development: although first considerations exist on how current 2-D video quality models may be adapted to 3-D, real 3-D models that are at par with their 2-D counterparts in terms of prediction accuracy are not yet available. This is partly due to the specific implications of 3-D video and related problems such as visual fatigue and diplopia [78], but also due to the employed coding and transmission strategies; see the section "3-D Video Quality Models." The situation for 3-D audio quality models is a very similar one. Subjective tests have indicated that timbral features are more important than spatial ones [79]. First respective models have recently been developed, but are signal based and cannot be used in a monitoring context [80]. For the ultimate goal of 3-D audiovisual quality prediction, no models are currently available.

IMPLEMENTATION OF SERVICE QUALITY MEASUREMENTS

Ideally, all new services or new types of devices should include built-in support for quality assessment and respective reporting. This will enable network operators and service providers to use such quality measurements as one important input to the optimization and trouble-shooting of their services and content-delivery networks. This requires standardization of not only media quality models, but also of concepts and protocols, which together can be used to efficiently implement quality assessment and client quality feedback reporting.

Existing approaches for integrating QoE and QoS data from different sources need to be extended [11], and the advantages of data warehouse-based solutions in combination with semiautomatic or automatic optimization of networks need to be exploited in more depth. This way, existing methods for fault localization, control, and network tomography will be complemented by respective QoE data, enabling a more comprehensive and QoE-based service monitoring, possibly considering aspects of user behavior and preferences. Such a multiservice perspective that includes user-related information will enable service-centric optimization and control of networks, more dedicated resource and service management, and to efficiently plan and implement comparable new services.

AUTHORS

Alexander Raake (alexander.raake@telekom.de) received his doctoral degree in electrical engineering and information technology from Ruhr-University Bochum, Germany, in 2005, and his electrical engineering diploma from RWTH Aachen, Germany, in 1997. From 1998 to 1999 he was a researcher at EPFL, Switzerland. Between 2004 and 2009 he held postdoc and senior scientist positions at LIMSI-CNRS, France, and Deutsche Telekom Laboratories, Germany, respectively. Since 2009, he has been an assistant professor at Deutsche Telekom Laboratories, Technical University Berlin. His research interests are in multimedia technology and QoE. He has been active in ITU-T since 1999 and is currently corapporteur of Q.14/12 on audiovisual quality.

Jörgen Gustafsson (jorgen.gustafsson@ericsson.com) obtained his M.Sc. degree from Linköping University, Sweden, in 1993. He then joined Ericsson Research, Luleå, Sweden, where he currently works as a master researcher in the field of network management. His research interests are in network management, performance monitoring, QoE, and end-user quality perception. Since 2001, he has been involved in ITU-T standardization, where he currently acts as corapporteur for question Q.14/12 on monitoring models for audiovisual services.

Savvas Argyropoulos (savvas.argyropoulos@telekom.de) received the diploma and Ph.D. degrees in electrical and

computer engineering from the Aristotle University of Thessaloniki, Greece, in 2004 and 2008, respectively. He is currently a research scientist at Deutsche Telekom Labs, Technical University of Berlin, Germany. He was also a research associate at the Informatics and Telematics Institute, Centre for Research and Technology Hellas. His research interests include video quality assessment, video coding and transmission, and distributed source coding. He received the 2008 European Biometrics Research Award. He is a Member of the IEEE.

Marie-Neige Garcia (marie-neige.garcia@telekom.de) holds a master's degree in engineering from the Electronics Superior Institute of Paris (France). She is currently pursuing her Ph.D. degree on audio-visual quality modeling of IPTV services at Deutsche Telekom Labs, Technical University of Berlin, Germany. From 2004 to 2006, she was in charge of the evaluation of text-to-speech systems and video technologies in both French and European projects at the Evaluation and Language Resources Distribution Agency, France. Her research topics are video and audiovisual quality assessment and quality modeling. She contributes to the ITU-T SG12 Q13 and Q14 activities and to VQEG.

David Lindegren (david.lindegren@ericsson.com) is an experienced researcher in the department of Wireless Access Networks at Ericsson Research in Luleå, Sweden. He joined Ericsson immediately after receiving his M.Sc. degree from the Luleå University of Technology in 2007. His main research interests are end-user subjective quality assessment for multimedia services together with radio measurements. He has been involved in related ITU-T standardization since 2008.

Gunnar Heikkilä (gunnar.heikkila@ericsson.com) is a senior specialist in the Department of Wireless Access Networks at Ericsson Research, Luleå, Sweden. He joined Ericsson in 1987 after obtaining an M.Sc. degree at Luleå Technical University. His current research areas are service quality measurements and end-user quality perception. He is also active in related standardization, mainly within 3GPP and ETSI.

Martin Pettersson (martin.m.pettersson@ericsson.com) is working as a senior researcher in the multimedia department at Ericsson Research in Stockholm, Sweden. He started at Ericsson after obtaining his M.Sc. degree in information and communications technology at Lund University, Sweden, in 2005. His main research interests are in subjective and objective video quality assessment and 3-D video. He is also interested in research of image and video compression. He has been involved in ITU-T standardization and VQEG activities since 2008.

Peter List (peter.list@telekom.de) graduated in applied physics in 1985 and received the Ph.D. degree in 1989 from the University of Frankfurt/Main, Germany. Since 1990, he has been with Deutsche Telekom, currently as a senior expert for video coding and quality at Deutsche Telekom Laboratories. He has actively contributed to international standardization of video compression technologies in MPEG, ITU, and several European projects.

Bernhard Feiten (bernhard.feiten@telekom.de) received the diploma and Dr.-Ing. degrees in electronic engineering from the Technische Universität Berlin in the field of speech synthesis, psychoacoustics, and audio bitrate reduction. He worked as an assistant professor at Technische Universität Berlin in the field of communication science and digital signal processing. Since 1996, he has been with Deutsche Telekom, currently as senior expert and project leader at Deutsche Telekom Laboratories. His research fields comprise audio and video coding algorithms, broadcasting applications, high-quality Internet media distribution and streaming, and QoE monitoring. He is actively working in standardization, mainly 3GPP SA4.

REFERENCES

[1] U. Jekosch, *Voice and Speech Quality Perception—Assessment and Evaluation*. Berlin: Springer-Verlag, 2005.

[2] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Amer.*, vol. 19, no. 1, pp. 90–119, 1947.

[3] R. Bücklein, "Hörbarkeit von Unregelmäßigkeiten in Frequenzgängen bei akustischer Übertragung," *Frequenz*, vol. 16/1962, pp/ 103–108, 1962.

[4] R. M. Krauss and P. D. Bricker, "Effects of transmission delay and access delay on the efficiency of verbal communication," *J. Acoust. Soc. Amer.*, vol. 41, no. 2, pp. 286–292, 1966.

[5] S. Bech and N. Zacharov, *Perceptual Audio Evaluation*. Hoboken, NJ: Wiley, 2006.

[6] A. Raake, *Speech Quality of VoIP—Assessment and Prediction*. Hoboken, NJ: Wiley, 2006.

[7] P. Mertz, A. Folwer, and H. Christopher, "Quality rating of television images," *Proc. Inst. Radio Eng.*, vol. 38, pp. 1269–1283, 1950.

[8] J. Allnatt, *Transmitted-Picture Assessment*. New York: Wiley, 1983.

[9] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660–668, 2008.

[10] M.-N. Garcia, R. Schleicher, and A. Raake, "Impairment-factor-based audio-visual quality model for IPTV: Influence of video resolution, degradation type, and content type," *EURASIP J. Image Video Processing*, vol. 11, no. 629284, pp/ 1–14, 2011.

[11] A. Mahimkar, Z. Ge, A. Shaikh, J. Wang, J. Yates, Y. Zhang, and Q. Zhao, "Towards automated performance diagnosis in a large IPTV network," in *Proc. ACM SIGCOMM*, 2009, pp. 1–12.

[12] A. Begen, C. Perkins, and J. Ott, "On the use of RTP for monitoring and fault isolation in IPTV," *IEEE Network Mag.*, vol. 24, no. 2, pp. 14–19, 2010.

[13] ITU, "Methods for subjective determination of transmission quality," ITU-T Rec. P.800, Int. Telecomm. Union, Geneva, Switzerland, 1996.

[14] ITU, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," ITU-R Rec. BS.1116-1, Int. Telecomm. Union, Geneva, Switzerland, 1997.

[15] ITU, "Method for the subjective assessment of intermediate quality levels of coding systems (MUSHRA)," ITU-R Rec. BS.1534-1, Int. Telecomm. Union, Geneva, Switzerland, 2003.

[16] ITU, "Methodology for the subjective assessment of the quality of television pictures," ITU-R Rec. BT.500-12, Int. Telecomm. Union, Geneva, Switzerland, 2009.

[17] ITU, "Subjective assessment methods for image quality in high-definition television," ITU-R Rec. BT.710-4, Int. Telecomm. Union, Geneva, Switzerland, 1998.

[18] ITU, "Subjective video quality assessment methods for multimedia applications," ITU-T Rec. P.910, Int. Telecomm. Union, Geneva, Switzerland, 1999.

[19] ITU, "SAMVIQ-Subjective assessment methodology for video quality," EBU-SAMVIQ, Int. Telecomm. Union, Geneva, Switzerland, 2003.

[20] ITU, "Subjective audiovisual quality assessment methods for multimedia applications," ITU-T Rec. P.911, Int. Telecomm. Union, Geneva, Switzerland, 1998.

[21] ITU, "Interactive test methods for audiovisual communications," ITU-T Rec. P.920, Int. Telecomm. Union, Geneva, Switzerland, 2000.

[22] P. Corriveau, C. Gojmerac, B. Hughes, and L. Stelmach, "All subjective scales are not created equal: The effects of context on different scales," *Signal Process.*, vol. 77, no. 1, pp. 1–9, 1999.

[23] M. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," in *Proc. SPIE*, vol. 5150, p. 573, 2003.

[24] Q. Huynh-Thu, M.-N. Garcia, F. Speranza, P. J. Corriveau, and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 1–14, 2011.

[25] S. Zieliński, F. Rumsey, and S. Bech, "On some biases encountered in modern audio quality listening tests—A review," *J. Audio Eng. Soc.*, vol. 56, no. 6, pp. 427–451, 2008.

[26] S. Möller, *Assessment and Prediction of Speech Quality in Telecommunications*. Norwell, MA: Kluwer, 2000.

[27] M. Waltermann, A. Raake, and S. Möller, "Quality dimensions of narrowband and wideband speech transmission," in *Acta Acustica utd with Acustica*, vol. 96, no. 6, pp. 1090–1103, 2010.

[28] S. Jumisko-Pyykko, J. Hakkinen, and G. Nyman, "Experienced quality factors—Qualitative evaluation approach to audiovisual quality," in *Proc. SPIE*, vol. 6507, 2007, p. 6507M.

[29] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. IEEE ICIP*, 2006, pp. 2945–2948.

[30] B. Weiss, S. Möller, A. Raake, J. Berger, and R. Ullmann, "Modeling call quality for time-varying transmission characteristics using simulated conversational structures," *Acta Acustic utd with Acustica,* vol. 95, no. 6, pp. 1140–1151, 2009.

[31] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *IEEE J. Select. Topics Signal Proc.*, vol. 3, no. 2, pp. 253-265, 2009.

[32] ITU, "The E-model, a computational model for use in transmission planning," *ITU–T Rec. G.107, Int. Telecomm. Union*, Geneva, Switzerland, 2009.

[33] ATIS-IFF, "Technical report on a validation process for IPTV perceptual quality measurements," *ATIS-0800035, Alliance for Telecommunications Industry Solutions, IPTV Interoperability Forum,* 2010.

[34] ITU, "Perceptual objective listening quality assessment (POLQA)," *ITU–T Rec. P.863, Int. Telecomm. Union*, Geneva, Switzerland, 2010.

[35] ITU, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," *ITU–T Rec. P.862, Int. Telecomm. Union*, Geneva, Switzerland, 2001.

[36] S. Möller, W.-Y. Chan, N. Côté, T. Falk, A. Raake, and M. Wältermann, "Speech quality estimation: Models and trends," *IEEE Signal Processing Mag.*, vol. 28, no. 6, pp. 18–28, 2011.

[37] ITU, "Single-ended method for objective speech quality assessment in narrow-band telephony applications," *ITU–T Rec. P.563, Int. Telecomm. Union*, Geneva, Switzerland, 2004.

[38] ITU, "Conformance testing for narrowband voice over IP transmission quality assessment models," *Int. Telecommun. Union, CH-Geneva, ITU–T Rec. P.564,* 2007.

[39] ITU, "Perceptual evaluation of audio quality (PEAQ)," *ITU–R Rec. BS.1387, Int. Telecomm. Union*, Geneva, Switzerland, 2001.

[40] ITU, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," *ITU–T Rec. J.144, Int. Telecomm. Union*, Geneva, Switzerland, 2004.

[41] ITU, "Objective perceptual multimedia video quality measurement in the presence of a full reference," *ITU–T Rec. J.247, Int. Telecomm. Union*, Geneva, Switzerland, 2008.

[42] ITU, "Objective perceptual multimedia measurement of HDTV for digital cable television in the presence of a full reference," *ITU-T J.341, Int. Telecomm. Union,* Geneva, Switzerland, 2010.

[43] ITU, "Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference," *ITU–T Rec. J.246, Int. Telecomm. Union*, Geneva, Switzerland, 2008.

[44] O. Verscheure, P. Frossard, and M. Hamdi, "User-oriented QoS analysis in MPEG-2 video delivery," *Real-Time Imag.*, vol. 5, p. 305–314, 1999.

[45] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," in *Proc. IEEE ICC*, 2008, pp. 110–114.

[46] H. Koumaras, C.-H. Lin, C.-K. Shieh, and A. Kourtis, "A framework for end-to-end video quality prediction of MPEG video," *J. Vis. Commun. Image Rep.*, 2009.

[47] F. You, W. Zhang, and J. Xiao, "Packet loss pattern and parametric video quality model for IPTV," in *Proc. IEEE/ACIS ICIS*, 2009, pp. 824–828.

[48] M.-N. Garcia and A. Raake, "Parametric packet-layer video quality model for IPTV," in *Proc. ISSPA*, 2010, pp. 349–352.

[49] T. Yamada, S. Yachida, Y. Senda, and M. Serizawa, "Accurate video-quality estimation without video decoding," in *Proc. IEEE ICASSP*, 2010, pp. 2426–2429.

[50] J. Gustafsson, G. Heikkilä, and M. Petterson, "Measuring multimedia quality in mobile networks with an objective parametric model," in *Proc. IEEE ICIP*, 2008, pp. 405–408.

[51] O. Sugimoto, S. Naito, S. Sakazawa, and A. Koike, "Objective perceptual video quality measurement method based on hybrid no reference framework," in *Proc. IEEE ICIP*, 2009, pp. 2237–2240.

[52] T. Brandao and M. P. Queluz, "No-reference quality assessment of H.264/AVC encoded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1437–1447, 2010.

[53] M.-N. Garcia, R. Schleicher, and A. Raake, "Towards a content-based parametric video qualiy model for IPTV," in *Proc. VPQM*, 2010, pp. 20–25.

[54] K. Yamagishi, T. Kawano, and T. Hayashi, "Hybrid video-quality-estimation model for IPTV services," in *Proc. IEEE Globecom*, 2009, pp. 1–5.

[55] S. Kanumuri, S. Subramanian, P. Cosman, and A. Reibman, "Predicting H.264 packet loss visibility using a generalized linear model," in *Proc. IEEE ICIP*, 2006, pp. 2017–2020.

[56] N. Suresh and N. Jayant, "AVQ: A zero-reference metric for automatic measurement of the quality of visual communications," in *Proc. VPQM*, 2007, pp. 1–4.

[57] M. Naccari, M. Tagliasacchi, and S. Tubaro, "No-reference video quality monitoring for H.264/AVC coded video," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 932–946, 2009.

[58] ITU, "Opinion model for video-telephony applications," *ITU-T Rec. G.1070, Int. Telecomm. Union*, Geneva, Switzerland, 2007.

[59] M. Barkowsky, R. Bitto, J. Bialkowski, and A. Kaup, "Comparison of matching strategies for temporal frame registration in the perceptual evaluation of video," in *Proc. VPQM*, 2006, pp. 1–4.

[60] A. Clark, "Method and system for viewer quality estimation of packet video streams," U.S. Patent 2009/0 041 114 A1, 2009.

[61] K. Watanabe, K. Yamagishi, J. Okamoto, and A. Takahashi , "Proposal of new QoE assessment approach for quality management of IPTV services," in *Proc. IEEE ICIP*, 2008, pp. 2060–2063.

[62] N. Suresh and N. Jayant, "'Mean time between failures': A subjectively meaningful video quality metric," in *Proc. IEEE ICASSP*, vol. 2, 2006, pp. 1–4.

[63] Q. Li and Z. Wang, "Video quality assessment by incorporating a motion perception model," in *Proc. IEEE ICIP*, 2007, pp. 173–176.

[64] J. You, J. Korhonen, and A. Perkis, "Attention modeling for video quality assessment: Balancing global quality and local quality," in *Proc. IEEE ICME*, 2010, pp. 914–919.

[65] Video Quality Experts Group (VQEG). [Online]. Available: www.vqeg.org

[66] Joint Effort Group (JEG). [Online]. Available: http://wiki.vqeg-jeg.org/

[67] W. Chen, J. Fournier, M. Barkowski, and P. L. Callet, "New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone," in *Proc. SPIE*, vol. 7863, 2011, p. 78631O.

[68] G. Akar, M. Tekalp, C. Fehn, and R. Civanlar, "Transport methods in 3DTV—A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1622–1630, 2007.

[69] P. Campisi, P. L. Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proc. EUSIPCO,* 2007, pp. 2110–2114.

[70] D. Hands, "A basic multimedia quality model," *IEEE Trans. Multimedia*, vol. 6, no. 6, pp. 806–816, 2004.

[71] BBF, "Data model for a TR-069 enabled STB," *Broadband Forum, TR-135*, 2007.

[72] IETF, "RTP control protocol extended reports (RTCP XR)," *Internet Eng. Task Force, IETF RFC 3611*, 2003.

[73] BBF, "CPE WAN management protocol (CWMP)," *TR-069, Broadband Forum*, Nov. 2010.

[74] 3GPP, "Transparent end-to-end packet-switched streaming service (PSS); Protocols and codecs," *3GPP TS 26.234*, June 2011, v10.1.0.

[75] 3GPP, "IP multimedia subsystem (IMS); Multimedia telephony; Media handling and interaction," *3GPP TS 26.114,* June 2011, v11.0.0.

[76] OMA, "Enabler release definition for OMA device management," *OMA-ERELD-DM-V1_2-20070209-A, Open Mobile Alliance,* 2007, appr. Vers. 1.2.

[77] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "Assessing quality of experience of iptv and video on demand services in real-life environments," *IEEE Trans. Broadcast.*, vol. 56, no. 4, pp. 458–466, 2010.

[78] M. Lambooij, I. Wijnand, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, no. 4, pp. 030 201–030 201-14, 2009.

[79] F. Rumsey, S. Zielinski, R. Kassier, and S. Bech, "On the relative importance of spatial and timbral fidelities in judgements of degraded multichannel audio quality," *J. Acoust. Soc. Amer.*, vol. 118, no. 2, pp. 968–976, 2005.

[80] F. Rumsey, S. Zielinski, P. Jackson, M. Dewhirst, R. Conetta, S. George, S. Bech, and D. Meares "QESTRAL (Part 1): Quality evaluation of spatial transmission and reproduction using an artificial listener," in *Proc. 125th AES Conv.*, 2008.

SP