Chapter 1
Video Streaming

M.-N. Garcia, S. Argyropoulos, N. Staelens, M. Naccari, M. Rios-Quintero and A. Raake

Abstract This chapter addresses QoE in the context of video streaming services. Both reliable and unreliable transport mechanisms are covered. An overview of video quality models is provided for each case, with a focus on standardized models. The degradations typically occurring in video streaming services, and which should be covered by the models, are also described. In addition, the chapter presents the results of various studies conducted to fill the gap between the existing video quality models and the estimation of QoE in the context of video streaming services. These studies include work on audiovisual quality modeling, field testing, and on the user impact. The chapter finishes with a discussion on the open issues related to QoE.

1.1 Introduction

With the multitude of video transmitted across the internet infrastructure, video QoE is of large interest for users, internet service or content providers, and component manufacturers alike. As a consequence, video-related services have received a lot of attention by research and development activities over the past years. As with other media applications such as audio entertainment or speech communication, three principle types of quality assessment can be distinguished:

(1) Explicit quality tests with users evaluating respective sequences in laboratory tests,

Marie-Neige Garcia, Assessment of IP-based Applications, Technische Universität Berlin, Germany, e-mail: marie-neige.garcia@telekom.de; Savvas Argyropoulos, Assessment of IP-based Applications, Technische Universität Berlin, Germany, e-mail: savvas.argyropoulos@telekom.de; Nicolas Staelens, Department of Information Technology, Internet Based Communication Networks and Services, Ghent University – iMinds, Belgium, e-mail: nicolas.staelens@intec.ugent.be; Matteo Naccari, BBC R&D, London, UK, e-mail: Matteo.Naccari@bbc.co.uk; Miguel Rios-Quintero, Assessment of IP-based Applications, Technische Universität Berlin, Germany, e-mail: Miguel.Rios-Quintero@telekom.de; Alexander Raake, Assessment of IP-based Applications, Technische Universität Berlin, Germany, e-mail: Alexander.Raake@telekom.de

(2) instrumental quality estimation algorithms, also called quality models, or
(3) possible additional consideration of context and user behavior in conjunction with the assessment of technical performance parameters, for example capturing service and user data during large-scale service use [17].

As for the case of other media, test methods involving human viewers can be distinguished according to the presentation method (method of constants vs. method of adjustment, and use of single versus multiple stimuli), and the scale being used for judgment. For an overview of subjective test methods, the interested reader is referred to [99], Chapter 4, and to the ITU recommendations [45, 46, 47, 30, 31, 19]. In addition, discussions and comparisons of methods can be found in [10, 73, 108, 27].

Quality models have been developed to complement or replace time-consuming and expensive viewing tests. For an overview, cf. e.g., [98, 9, 75].

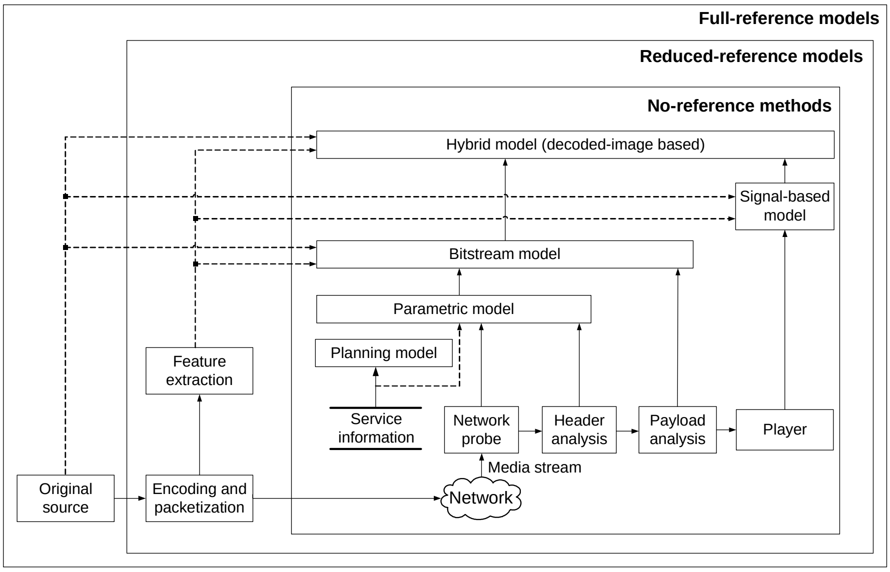


Fig. 1.1 Categorization of video quality assessment algorithms, adapted from [75].

The categorization of algorithms based on the type and amount of information employed for the quality assessment is depicted in Fig. 1.1, modified from [75]. From this figure, it can be seen that the quality models can be categorized in terms of:

- the amount of reference information they employ: No-Reference (NR), where the models do not have access to the original non-degraded signal, Reduced-Reference (RR), where the models have access to features extracted from the original signal (see box "Feature extraction" in Fig. 1.1), and Full-Reference

(FR), where the models have access to the original signal (“Original source” in Fig. 1.1),

- the type of information that is used for quality predictions: Signals (“signal-based model” in Fig. 1.1) and/or transmission-related parameters extracted from packet-header- (“Parametric model” in Fig. 1.1) or bitstream- information (“Bitstream model” in Fig. 1.1). Hybrid models take as input signal (pixel), bitstream, and/or packet-header information,
- the extent to which they include the explicit modeling of the human visual system.

This chapter mainly addresses quality models focusing on IP-based video streaming applications and the most widely used codec in this context, H.264 [38]. However, the scope of the presented models is broader: Signal-based models are usually not restricted to H.264, and they can be applied universally. Also, the structure of the other types of models make them generally adaptable to other codecs and network types.

Typically, video transmission over IP networks (e.g. for broadcast TV as in IP-based Television – IPTV) is performed using unreliable transport mechanisms, such as the Real Time Protocol (RTP) in conjunction with the User Datagram Protocol (UDP), to ensure limited delay and real-time operation. However, with the increase of available network bandwidth, multimedia content can now be delivered very efficiently using TCP and typically HTTP (e.g. in the case of Youtube or other so-called over-the-top services), which enables traffic reduction by the efficient usage of caches, for example in the vicinity of end-users. Different considerations for QoE in the context of UDP-based versus TCP-based transmission are discussed in Sections 1.2 and 1.3, respectively.

The chapter is structured as follows: Section 1.2 summarizes the video degradations that are encountered in case of streaming with unreliable transport, and respective approaches for instrumental assessment. These include the impact due to video coding as well as the packet-based transmission, possible packet loss and its concealment. Here, different types of models are outlined, including packet-header-based models for network planning and monitoring, bitstream-based models, pixel-based models and hybrid models. Section 1.3 discusses the differences between streaming over unreliable and over reliable channels, and how models initially developed for video streaming with unreliable transport can be used here, and which additional components are required to also handle adaptive streaming or re-buffering. In Section 1.4, the rather technical approach followed up to that point is re-considered, and current trends towards a more QoE-centric assessment of video streaming services are presented, including added modalities and audiovisual assessment, field rather than lab testing, and the impact due to the type of user. Finally, in Section 1.5, future work in the field of video streaming QoE assessment is discussed.

1.2 Video quality models for streaming with unreliable transport mechanisms

This section introduces the main degradations occurring due to compression or packet loss in the case of unreliable transport mechanisms. It also provides an overview of the different types of video quality models, with a focus on standardized models. Packet-, bitstream-, pixel-based, and hybrid models are addressed, as well as full-, reduced-, and no-reference models.

1.2.1 Video coding and packet error degradations

Blockiness, also referred to as *block distortion* or *tiling* [48], is a distortion of the image characterized by the appearance of an underlying block encoding structure. Block distortions are caused by coarse quantization. They comprise other identified degradations such as the *staircase effect*, *mosaic pattern effect* or the *DCT basis-image effect* [107]. In modern encoders, coarsely quantized block boundaries are usually filtered, reducing the visibility of the above artifacts but leading to *blurriness*, which is characterized by reduced sharpness of edges and spatial details [48].

In order to reduce the amount of video information that the system is required to transmit or process per unit of time, video frames may be skipped at the encoding stage. This may result in *jerkiness*, which is defined in [48] as a "motion which was originally smooth and continuous, but is now perceived as a series of distinct "snapshots" of the original scene". It is often observed in the case of high motion scene.

During the encoding process, video frames are assigned different types, which are called "I-frames", "P-frames" and "B-frames". The perceptual impact of packet loss depends on the type of the frame in which the loss occurs. Indeed, P- and B-frames are predicted from previous I- and P-frames, while I-frames are intra-coded and therefore do not depend on previous frames. As a consequence, if a loss occurs on an I- or a P-frame, the loss is typically propagated till the next I-frame. If a loss occurs on a reference B-frame, which is used in hierarchical coding, the loss propagates till the next P- or I-frame, i.e. it only affects the surrounding non-reference B-frames. There is no loss propagation if the loss occurs on a non-reference B-frame, since it is not referenced by other frames.

The perceptual impact of packet loss also depends on the packet loss concealment applied by the decoder. If *slicing* is applied as packet loss concealment, one packet loss results in the loss of the corresponding pixel-area as well as the pixel-area corresponding to the rest of the affected slice (see red rectangle in Fig. 1.2(a)). The decoder re-synchronizes its bitstream parsing process at the beginning of the next slice, using the slice header. Therefore, the spatial extent of the loss depends on the number of slices per frame. For instance, and as shown in Fig. 1.2(a), the video was encoded with one slice per frame, the loss of a packet yielded the loss of

(a) Loss occurred in the current frame (b) Loss propagated from previous frames

Fig. 1.2 Effect of packet loss when one slice per frame is used.

the rest of the frame, and the lossy area was replaced by the same area in the last reference frame.

In the case of slicing, packet loss may also yield *blockiness* effect, as shown in the red rectangle area of Fig. 1.2(b). Indeed, when loss occurs, content from the previous reference frame is usually copied to the lost portion of the hit frame. If there is motion in the sequence, this replaced content will not fit well the missing content. In the subsequent frames, the whole lost block of pixels will be displaced, resulting in a blocking artifact.

Another loss handling strategy is *freezing with skipping*. In this case, the frames affected by loss are completely discarded. According to the encoding process described above, the frames referencing the lossy frames are also discarded, and all discarded frames are replaced by the last unimpaired reference frame. The video is therefore perceived as frozen.

1.2.2 Packet-based models for network planning

In the case of network planning, parameters of the service to be deployed cannot be measured. Instead, planning assumptions are made. Typical model input parameters used for network planning are the average bitrate and the percentage of packet loss, as proposed by [92] for variable bitrate, random loss and motion-compensated transmission-error concealment. Since the impact of packet loss depends on the loss distribution [7, 21], burst-length-related parameters are commonly used, as presented in [90] for small video formats and in [21] for IPTV and High Definition (HD) video. In [105], the authors improved the ITU-T G.1070 model [36] by including a burstiness parameter computed from the burst density (fraction of lost or discarded packets in a burst period), the burst duration, the number of lost packets and the number of burst periods. Alternatively, the packet loss frequency may be used instead of the packet loss percentage and packet burst length [102]. Since the perceptual impact depends on the applied packet loss concealment (slicing or freezing with skipping) and on the number of slices per frame (in the case of slicing), both the packet loss concealment and the number of slices per frame should be considered by the model, as presented in [21] for the IPTV scenario.

1.2.3 Packet-based models for service monitoring

In the case of service monitoring, model parameters are extracted from the bitstream. The most applicable case of an encrypted bitstream is considered in this section. In this case, the model does not have access to the payload or to the pixel information.

The network planning models can be used in that case. However, additional information may be extracted from the encrypted bitstream such as the video frame boundaries and the video frame types (I-, P- or reference/non-reference B- frames). This allows a more accurate parametric description of the degradations and of the influence of the content.

Indeed, it is known that the impact of packet loss depends on the type of the frame in which the loss occurs [92]. In particular, a loss propagates till the next I-frame when it occurs in an I or a P frame (in contrast to a B-frame, see also Section 1.2.1). The number of impaired frames is therefore more appropriate than the percentage of packet loss for capturing the quality impact of packet loss, as proposed in [101] and [103] for slicing, and in the ITU-T P.1201.2 standard [43], for both slicing and freezing with skipping.

As previously mentioned, the spatial extent of the loss depends on the slice size in the case of slicing. This spatial extent is computed in both the ITU-T P.1201.2 and in [20]. In the ITU-T P.1201.2 standard, the spatial extent of the loss is combined with the loss duration in a single parameter describing the whole degradation for directly predicting the perceived quality.

It has often been observed that the spatio-temporal complexity of content influences the quality impact of coding artifacts and the visibility of the loss. For instance, slicing degradations are more visible in the case of panning or complex movements than in the case of almost static-content. In the case of an encrypted stream, this content complexity may be captured by the frame sizes and frame types, as proposed in [64, 43, 42, 103, 104, 101]. In particular, in the case of coding degradations, the I-frame sizes are used [43, 42], reflecting the observation that high I-frame sizes indicate low content complexity at low bitrates. The ratios between B, P and I frame sizes may also be used [104, 43]. In [43], these ratios capture the observation that similarly small B- and P- frames sizes, compared to I-frames sizes, indicate low temporal complexity.

1.2.4 Bitstream-based models

Bitstream-based objective quality models predict video quality using the encoded video as it is transmitted through the network (see Fig. 1.1). These models parse the video bitstream without reconstructing the pixel information and are particularly interesting for monitoring video quality at any point in the distribution network (network-based monitoring). However, parsing the bitstream requires coder specific implementations of each model.

By analyzing the encoded video bitstream beyond the packet headers as described in the previous section, additional information can be gathered at the frame level. For example, quantization and motion vector information can be obtained by parsing the video data [59, 82]. This provides a first indication of the video quality and video content characteristics [65].

In [66], quality is estimated in the compressed domain by considering the Quantization Parameter (QP), the motion and the bitrate allocation of inter macroblocks. Similar information has also been used in [59]. In [85], information on the location and spatial extent of the picture loss is extracted from the encoded bitstream to predict perceived video quality.

Besides predicting video quality, bitstream-based models have also been constructed to estimate the visibility of impairments due to, for example, packet loss in the network [77, 55, 56, 2, 88]. This approach can be used to verify if the delivery network is able to provide adequate QoE to the end-users [18, 37]. The bitstream features which are typically used to detect the visibility of loss are the duration and extent of the propagated error, the motion and the residual error of the degraded area in order to account for the spatio-temporal complexity of the video sequence and capture the perceptual effects caused by packet loss. Then, the prediction of visibility may be performed using classifications methods, such as generalized linear models [77, 55, 56], support vector regression [2], or binary trees [88].

Bitstream based quality assessment models are standardized in ITU-T Rec. 1202 [32], where P.1202.1 [33] refers to lower resolutions (from QCIF to HVGA) and P.1202.2 [34] refers to higher resolutions (from Standard Definition (SD) to HD). In P.1202.1, compression artifacts are computed based on the QP, the key frame rate, the frame rate, and the motion vector magnitude. Similarly, in P.1202.2, the compression degradations are computed based on two parameters: the average QP of the sequence and a parameter which denotes the content complexity (computed based on the bits per pixel and the QP).

The perceived distortion of slicing degradations depends on the effectiveness of the employed error concealment technique. Thus, the *level of visible artifacts* is computed based on the motion information, the residual energy of the erroneous area, and the error propagation extent to indicate how annoying is the slicing artefact. This reflects the principle that the parts of the sequence which can be easily predicted (e.g. low texture, low motion) can be efficiently concealed.

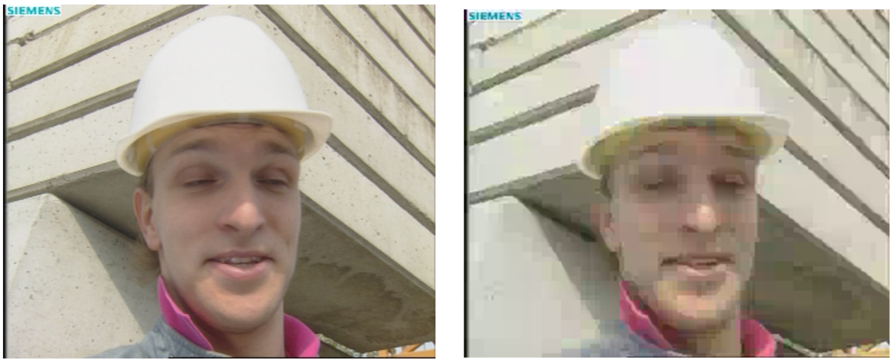
The distortion caused by each freezing event is computed based on the freezing duration and a motion term to reflect the fact that a freezing in the fast moving part of the video results in larger jerkiness and therefore causes larger perceptual annoyance. Finally, the overall freezing degradation is computed as the square root of the sum of the individual degradations of each freezing event.

1.2.5 Pixel-based models

Pixel-based models estimate the video or image quality using features or information associated with pixel data only (see Fig. 1.1). These models are therefore taking as input the decoded video. They may in addition use the reference (non-degraded) video signal. The most popular metric in this category is the Peak Signal to Noise Ratio (PSNR), which is computed based on the Mean Squared Error (MSE) between the two (degraded and non-degraded) video signals. PSNR is quite popular due to its inherent properties: it is simple to compute, parameter-free and memory-less, and as a result, it can be calculated locally without consideration of other source samples. The MSE is, however, a signal fidelity measure and not a perceived quality metric. In fact, it shows poor correlation with perceived quality mainly because it does not take into account the properties of the human visual system [28].

The Human Visual System (HVS) is characterized among other things of contrast sensitivity and masking which lead to a space and time varying sensitivity of the artifacts associated with lossy coding and packet errors. As an example, Fig. 1.3 shows the first frame of the *Foreman* sequence in CIF resolution. The sequence has been coarsely quantized and therefore coding artifacts, namely blockiness, are visible. However, due to the space varying sensitivity to distortion of the HVS, the blocking is more noticeable around the face, while it is less disturbing on the building wall. Therefore HVS-based models embed the HVS properties to weight more the artifacts present in image areas where the human eye is more sensitive and vice-versa. One of the first HVS-based model is the Visual Difference Predictor (VDP) [12], which uses the contrast sensitivity function of the HVS to weight differently the artifacts in image area. For a thoughtful review of HVS-based models, the interested reader is referred to [6].

Another well-known pixel-based model is the Structural SIMilarity (SSIM) index [94], which has been initially proposed for images and eventually extended for videos [95]. It is a full-reference model and for each pixel computes the distortion as the contribution of three terms: luminance, contrast and structure. The terms are then multiplied together and the final SSIM score is obtained as the average over all pixels. The work in [91] estimates the Mean Square Error (MSE) induced by channel errors by performing a maximum-a-posteriori estimation to detect the location of corrupted pixels. After these pixels are detected, the NORM algorithm ([67], see also Section 1.2.6) is run to estimate the final MSE. Considering also the temporal component of videos, the work in [74] proposes a general reduced-reference Video Quality Model (VQM). The VQM uses features computed over the pixels of the original and processed video and combines all of them according to a linear weighting. The features used by VQM are related to quality degradation introduced by lossy coding or channel errors. The VQM has been standardised as quality model



(a) *Foreman* original frame (b) Same frame coarsely quantized

Fig. 1.3 Space varying sensitivity of the HVS to coding artifacts. The blockiness is more visible in flat areas than in high contrast areas as the building wall.

in the set of models specified in the ITU-T J.144 [39]. Furthermore, VQuad-HD is a full-reference video quality model for high definition video signals, which was selected by ITU-T as Recommendation J.341 [40]. It is based on the computation of the following quality features: blockiness, slicing, blurring, and jerkiness, as defined in Section 1.2.1.

1.2.6 Hybrid models

Hybrid video quality assessment models employ a combination of packet information, bitstream information and the decoded reconstructed video sequence (see Fig. 1.1). In general, in a hybrid video quality assessment algorithm the features extracted or calculated from the bitstream (e.g., motion vectors, macroblock types, transform coefficients, quantization parameters, loss duration, etc.), and the information extracted by packet headers (e.g., bitrate, packet loss, delay, etc.) are combined with the features extracted from the decoded and reconstructed images in the pixel domain. Since the reconstructed image can be obtained from the decoding device, this type of model ensures that the error concealment method of the decoder is taken into consideration. Within the Video Quality Experts Group (VQEG), efforts are ongoing towards the joint construction and validation of novel objective hybrid video quality models [87].

The V-Factor [98] model inspects different parts of the encoded bitstream and extracts information from the packet headers and the encoded and decoded video to model the impact of packet loss during video streaming. In [63], coding parameters are extracted during the decoding process of the video to construct a hybrid bitstream-based quality model. The model is based on a linear combination of quantization parameters, bitrate and boundary strength parameters. The latter parameter influences the intensity of the de-blocking filter in order to minimize blockiness artifacts. In [57], the authors further extended an existing bitstream based objective video quality model [58] by including pixel-based features. These features are based on detecting bluriness, blockiness, motion continuity and other aspects describing video quality. Their results show that the hybrid model outperforms the bitstream-based model. Similarly, a hybrid no-reference model for H.264/AVC encoded sequences is proposed in [13]. It is based on the quantization parameter and a pixel difference contrast measure.

The work in [78] estimates the MSE induced by channel errors between the reconstructed signal at the encoder and decoder. The model targets MPEG-x and H.26x codecs and is designed in three different versions denoted as Full-Parse (FP), Quick-Parse (QP) and No-Parse (NP). The FP version estimates the square error on the luma component for each pixel and then provides the MSE at the required level of granularity (e.g. macroblock, slice, frame, etc.). The input data to the FP algorithm require entropy decoding and inverse quantisation only. The QP estimates the MSE at slice level using bitstream parameters such as packet headers, thus without requiring any decoding operation. Finally, the NP estimates the MSE at sequence level using a linear relationship between the packet loss rate and the MSE. As may be noted, the three different versions lead to a different trade-off between model complexity and accuracy, with the FP being the most complex and accurate version. Finally, the work in [67] describes a NO-Reference video quality Monitoring (NORM) model which estimates the MSE induced by channel losses for H.264/AVC coded videos at the macroblock level. The NORM algorithm models the channel distortion as the result of three contributions: lack of motion vectors, lack of prediction residuals and error propagation from previous frames due to motion compensation. The contribution to the MSE due to specific coding tools of the H.264/AVC standard (i.e. intra prediction and deblocking filter) is also addressed. The NORM estimate has also been used to devise a reduced-reference quality model based on SSIM [94].

1.3 Video quality models for streaming with reliable transport mechanisms

This section begins with a short description of the progressive download and adaptive streaming mechanisms and of the corresponding degradations. Subsequently, an overview of models developed for the quality assessment of progressive download and adaptive streaming services is presented.

1.3.1 Progressive download and adaptive streaming mechanism and degradations

There are two main types of video streaming over HTTP: (a) progressive download and (b) HTTP adaptive streaming (HAS). In progressive download, the client may begin the playback of the media before the whole media is downloaded. However, if the available throughput is lower than the bitrate of the media, the player will stall until enough data have been downloaded. This is perceived by the end users as *freezing without skipping*, which is typically called *rebuffering* or *stalling*. To avoid stalling during playback and enable smoother flow of video, HAS methods adapt to the available network conditions. In HAS applications, the video is encoded in multiple quality versions, called "representations", which are segmented in short intervals, typically between 2 and 10 seconds long. The adaptive client periodically requests segments of the video content from an HTTP server, and then decodes and displays this segment. The client may switch between different representations at each request depending (mainly) on the available bandwidth. The aim is to provide the best quality of experience for the user by avoiding stalling events. However, the perceived artefact in this case is the fluctuating quality of the video sequence and quality models should consider the impact of temporal quality adaptation. The interested reader is referred to [83] for more information on standardized HAS methods such as the Dynamic Adaptive Streaming over HTTP (DASH).

1.3.2 Progressive download models

An audiovisual quality model has been proposed in [26] for rebuffering degradations[1]. The model takes as inputs the number of stalling events and the average length of a single stalling event. It was developed based on the results of laboratory and crowdsourcing tests. Video sequences had different video resolutions, but no extremely small or high definition resolutions. The video durations were typical of Youtube videos (on average 5.54 *min* and up to 15 *min*).

Another model has been proposed in the ITU-T P.1202.1 Recommendation [33] for capturing the quality impact of rebuffering degradations, for low video resolutions (up to HVGA). The model is based on the ratio between the rebuffering duration of the sequence normalized to the total duration of the sequence (including the rebuffering duration). The ITU-T P.1202.1 model has been developed based on the results of subjective laboratory tests. The model was validated on video sequences of 16 s and up to 30 s. However, the rebuffering ratio parameter is normalized to the sequence duration and is therefore in principle applicable to longer sequences.

[1] Note that the focus was so far on visual stimuli and, therefore, video quality models. Due to its impact on the scientific work dedicated to rebuffering models, and although it is audiovisual, the model of [26] is presented in this section.

The rebuffering quality model of the ITU-T P.1201.1 Recommendation [42] has been developed using the same test databases as the P.1201.2 model. Similarly to [26], it uses as input parameters the number of stalling events and the average length of a single stalling event. In addition, the average distance between two rebuffering events is used as an input parameter to capture the distribution of the rebuffering events in the sequence.

Both the ITU-T P.1201.1 and P.1202.1 are addressing the lower video resolution application area. No rebuffering quality model has been so far standardized for higher resolution applications such as IPTV. However, the ITU-T is currently planning to develop a parametric model for progressive donwload ("P.NAMS-PD") valid for both lower and higher video resolution application areas. This model will also capture the quality impact of initial rebuffering. Note that the quality impact of initial rebuffering has also been studied in [25], with initial delays up to 30 s for 60 s video duration. This impact was found negligible compared to the quality impact of stalling events occurring during the playback of the video.

1.3.3 Adaptive streaming models

The main feature of video sequences transmitted using HAS is the changes in quality over time. Most of the existing quality assessment algorithms of Section 1.2 assume constant base (i.e. without packet loss) quality over the whole sequence and have been designed for short durations, typically between 10 and 20 seconds. With HAS services, the base quality is varying over time, and the quality adaptation typically lasts more than 20 seconds. In addition, these services facilitate the switching between devices and video resolutions (TV screen, tablet PC, smartphones) during playback. Quality models have therefore to be adapted to estimate quality for longer durations sequences (from a few seconds to several minutes), for fluctuating quality within the sequence, and even for switching between devices. Since quality models are preferably developed based on the results of subjective tests, a revision of the existing standardized subjective test methods [29, 45] is needed.

The impact of time-varying quality on human perception was investigated in [23], and it was shown that subjects react with different time constants to large sudden quality degradations or improvements. Moreover, it was shown that the location of a quality degradation or improvement influences subjects' overall judgements, revealing a recency effect. On the video domain, user perception of adapting quality was studied in [11, 69], revealing that the perception of spatial distortions over time can be largely modified by their temporal changes. Moreover, in [80], an hysteresis effect was observed in the subjective judgment of time-varying video quality. The video adaptation scheme of [100] for modeling the impact of bitrate and frame rate adaptation on perceived quality suggests that for video sequences with high temporal complexity, adaptation on frame rate will result in better quality than adapting QP; on the other hand, adaptation of QP is beneficial for video with low motion or fine texture details.

Quality assessment models for HAS can also employ existing models to assess the quality of short intervals (e.g., approx. 10 sec) and then combine these scores into a single quality estimate using temporal pooling techniques. There are several temporal pooling algorithms ranging from simple approaches which consider the maximum, minimum, or mean video quality to those which integrate the temporal properties of human perception, memory effects, and transient properties. For example, in [62], temporal pooling is based on motion, while in [70], a content adaptive spatial and temporal pooling strategy is proposed which takes into consideration the severity of the quality degradations. In [81], an evaluation of the most popular pooling techniques using PSNR and SSIM as quality predictors concluded that the plain average of individual quality scores can achieve comparable results with the most sophisticated pooling methods. Further extending this study, a temporal pooling scheme based on an auto-regressive moving-average (ARMA) model to simulate the adaptation of perceived quality over time was presented in [68]. It is based on the computation of standardized video quality models, such as J.144, J.341, and P.1201.2, on video chunks of short duration, typically from 5 to 15 seconds. Moreover, a penalty parameter was introduced into the model to take into consideration the abrupt quality degradation within the sequence and the frequency in representation switches. In conclusion, the aforementioned approaches indicate that objective estimates of short sub-sections of a video sequence can be efficiently pooled into a single score as long as the memory and recency effects are taken into account.

1.4 From video quality towards QoE

The models presented in the previous sections estimate the video quality of short sequences, typically 10 s, in the context of laboratory testing, where the viewing and listening environment as well as the task given to the user deviate from typical viewing and listening conditions. These models are therefore not estimating the QoE, and several aspects need to be taken into account to achieve a more QoE-centric assessment approach. These aspects include audio-visual quality, field testing, and user impact characterization. An overview of theses aspects is provided in this section.

1.4.1 Audiovisual quality models

Several studies on audiovisual perception, summarized in [60], have been conducted in the 80's. However, the first audiovisual quality models to be found in the literature appeared as late as in the 90's. At this time, these models addressed either analog degradations, such as audio and video noise [49, 35, 3], or compression artifacts, such as blockiness [8, 50, 52, 24]. For an overview of audiovisual quality models covering analog and compression degradations, see [106]. The interest in modeling audiovisual quality has risen again in the past ten years, reflected for instance by standardization activities such as the ITU-T Recommendations P.1201.1 and P.1201.2 [42, 43] or the Audiovisual High Definition Quality (AVHD) project of VQEG, which intends to evaluate audiovisual quality models for multimedia applications and HD resolution. Audiovisual quality models for mobile applications have been developed in 2005 and 2006 [79, 97], but the reported model versions do not 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. This influence of the degradation type has been studied in [22] for higher resolution application areas such as IPTV. In particular, the authors show that the impact of the audio quality on the overall perceived audiovisual quality is higher in the case of audio packet loss than in the case of audio coding degradations. In parallel to this finding, an audiovisual quality model has been proposed which captures the impact of the audio and video degradation types (coding vs. transmission-error degradations). Both compression and transmission-error degradations are covered as well in [4] for interactive scenarios and small video resolutions. Based on an extensive review of the literature, both [22] and [106] highlight that video quality generally dominates the perceived audiovisual quality, but that this dominance depends on the semantic audiovisual content. Finally, an overview of existing audiovisual quality models is provided in [71]. The authors show that a simple model based on the product of audio and video quality terms is valid for a wide range of scenarios and applications. When a small amount of data is available for a wide range of applications, it is indeed a safe choice to use a model as simple and with as few coefficients as possible to avoid overtraining.

1.4.2 Ecologically valid testing

As detailed in Chapter 10, subjective methodologies for assessing momentary QoE provide detailed guidelines describing how to conduct such experiments. These recommendations define different methods for presenting and rating video sequences. Typically, short duration (10s $\sim$ 15s) video sequences are presented to the test subjects and rated immediately after watching. In the case of longer video sequences (up to 30min), continuous quality evaluation is recommended where subjects rate quality while watching the video [29]. Specific instructions are provided to the test subjects on how to evaluate the video sequences at the beginning of the experiment. The assessment methodologies also specify requirements for the environment in which the subjective experiment is conducted. These requirements are formulated in terms of room illumination, subject seating position, screen calibration, etc. As such, subjective quality assessment experiments are usually conducted in controlled lab environments. These assessment methodologies are still actively used for measuring pure video or audiovisual quality.

The broad availability of high speed Internet access and growing number of multimedia-capable devices (such as smartphones and tables) enable watching video content anyplace, anytime. Thus, the environment in which video is actually consumed does not necessarily comply with the recommended controlled lab setting. Furthermore, according to its definition ([41, 61] and Chapter 2), Quality of Experience (QoE), is influenced by user expectations, context, and personal preferences. Therefore, capturing and understanding end-users' QoE goes beyond purely measuring video quality [76]. This calls for new subjective studies and methodologies [93, 15, 5, 72, 16] enabling episodic and multi-episodic subjective quality evaluation in more realistic and ecologically valid environments (cf. Section 10.4).

A first effort towards assessing QoE of IPTV services in real-life environments has been made in [86, 84]. This study involved conducting subjective experiments in subjects' own home environment under realistic viewing conditions. Comparing the results obtained with subjective tests conducted in a controlled lab highlighted the importance of the assessment environment, primary focus, and immersion on impairment visibility and quality perception. It is found that immersion has a major impact on impairment tolerance and overall QoE.

Field testing can provide new insights and findings which cannot necessarily be discovered in a controlled lab setting. In this respect, field testing should complement lab testing rather than replace it.

1.4.3 User impact

The viewing and listening environmental set-up is thoroughly controlled in standardized video quality test methods [30, 44]. These systematic methodologies reduce significantly the amount of noise in the results and enable the comparison of test results between labs. However, the outcome of these tests cannot be easily extrapolated to more realistic scenarios due to variable factors such as the context of use.

As presented in Chapter 4, the main limitation in the development of more realistic tests is the complex interaction between all determinant influential factors of the ecosystem. Hence, researchers from the field of social sciences and economy tried to analyse how factors such as user demography, overall service quality and context of use interact and influence the perception. New approaches were developed, such as the Theory of Acceptance Model (TAM) [14], and more recently the Unified Theory of Acceptance and Use of Technology (UTAUT) [1]. These theories intend to understand the development of intention of use of a service, based on strong behavioural elements such as external influential variables, perceived usefulness and perceived ease of use.

First uses of these theories can be seen in [96, 51], where modified versions of TAM are utilized to predict the adoption of IPTV services. The results are surprisingly aligned to studies in the context of mobile television [53], demonstrating that the content offered, the technological knowledge of the user, his/her attitude towards technology, and socio-economic factors are crucial for a positive experience of the service. The above-mentioned findings suggest that users with similar characteristics show similar behaviour, an observation that, in spite of being studied in other services for decades, has been little explored in the study of multimedia services. Among the most relevant factors, the degree of expertise of the users with the service greatly influences the way in which they perceive and evaluate quality. However, this type of classification only gives a general idea concerning the main factors influencing the users' evaluation process. It has therefore been necessary to combine cognitive and psycho-perceptual methods that provide the user with the freedom to assess the service attributes in their own words, as seen in the work of Strohmeier and Jumisko-Pyykkö [89, 54].

1.5 Discussion

The literature is rich in video quality models, with different model types according to the application needs. The first models were developed in order to address compression artifacts. Then, new models appeared for covering packet loss degradations yielding slicing or freezing with skipping degradations in the case of unreliable network such as RTP over UDP. Other models have recently been proposed for addressing the progressive download scenario.

These models all target short-term video quality predictions and ignore which user and in which context the user utilizes the video streaming service. Some studies have already been conducted for addressing longer term quality predictions. Other subjective tests have been conducted to address the context- and user-impact. However, this topic needs much more investigation, especially with the new emerging types of video streaming applications such as adaptive streaming and complex scenarios such as portable TV, where the user can watch TV in different locations and on different screens.

It is still open which subjective tests should be conducted to address these complex scenarios and to identify which factors, in addition to the perceived video quality, influence the overall QoE. Also open is the type of measurement tools to be targeted. Indeed, a "measurement-window" approach, with which quality scores are output for example every 10 s, is traditionally used in service monitoring. With the diversity of degradations, and in order to better capture the long-term QoE prediction, a "remembered-event" approach may become more appropriate, where an event represents any kind of degradations of any duration. In that case, efforts should be spent on the identification and characterisation of these "events", as well on the weighting of their contribution to the overall QoE.

References

1. Al-Qeisi, K.I.: Analyzing the use of utaut model in explaining an online behaviour: Internet banking adoption. Ph.D. thesis, Brunel University (2009)
2. Argyropoulos, S., Raake, A., Garcia, M.N., List, P.: No-reference video quality assessment for SD and HD H.264/AVC sequences based on continuous estimates of packet loss visibility. In: Int. Workshop on Quality of Multimedia Experience, pp. 31–36 (2011)
3. Beerends, J.G., Caluwe, F.: The influence of video quality on perceived audio quality and vice versa. In: J. Audio Eng. Soc. (1999)
4. Belmudez, B., Möller, S., Lewcio, B., Raake, A., Mehmood., A.: Audio and video channel impact on perceived audio-visual quality in different interactive contexts. In: IEEE Int. Workshop on Multimedia Signal Processing (2009)
5. Borowiak, A., Reiter, U., Svensson, U.P.: Quality evaluation of long duration audiovisual content. In: IEEE Consumer Communications and Networking Conference, pp. 337 –341 (2012)
6. Bosc, E., Le Callet, P., Morin, L., Pressigout, M.: Visual quality assessment of synthesized views in the context of 3D-TV. In: 3D-TV systems with depth-based-image-rendering architectures, techniques and challenges (2013)
7. Boulos, F., Parrein, B., Callet, P.L., Hands., D.: Perceptual effects of packet loss on H.264/AVC encoded videos. In: Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (2009)
8. Chateau, N.: Relations between audio, video and audio-visual quality. In: Contr. COM 12-61 to ITU-T Study Group 12 (1998)
9. Chikkerur, S., Sundaram, V., Reisslein, M., Karam, L.J.: Objective video quality assessment methods: A classification, review, and performance comparison. IEEE Trans. Broadcasting **57**(2), 165–182 (2011)
10. Corriveau, P., Gojmerac, C., Hughes, B., Stelmach, L.: All subjective scales are not created equal: The effects of context on different scales. Signal Processing **77**(1), 1–9 (1999)
11. Cranley, N., Perry, P., Murphy, L.: User perception of adapting video quality. Int. J. Hum.-Comput. Stud. **64**(8), 637–647 (2006)
12. Daly, S.: The visible difference predictor: an algorithm for the assessment of image fidelity. In: SPIE Human Vision, Visual Processing and Digital Display, vol. 1666 (1992)
13. Davis, A.G., Bayart, D., Hands, D.S.: Hybrid no-reference video quality prediction. In: Proc. of IEEE Int. Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–6 (2009)
14. Davis, F., Bagozzi, R., Warshaw, P.: User acceptance of computer technology: A comparison of two theoretical models. Management Science **35**, 982–1003 (1989)
15. De Moor, K., Ketyko, I., Joseph, W., Deryckere, T., De Marez, L., Martens, L., Verleye, G.: Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting. Mobile Networks and Applications **15**(3), 378–391 (2010)
16. De Pessemier, T., De Moor, K., Joseph, W., De Marez, L., Martens, L.: Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching. IEEE Transactions on Broadcasting **59**(1), 47–61 (2013)
17. Dobrian, F., Awan, A., Joseph, D., Ganjam, A., Zhan, J., Sekar, V., Stoica, I., Zhang, H.: Understanding the impact of video quality on user engagement. Comm. ACM **56**(3), 91–99 (2013)
18. DSL Forum Technical Report TR-126: Triple-play Services Quality of Experience (QoE) requirements. DSL Forum (2006)
19. EBU-SAMVIQ: SAMVIQ-Subjective assessment methodology for video quality. Report by the EBU Project Group B/VIM, European Broadcasting Union, CH–Geneva (2003)
20. Frossard, P., Verscheure, O.: Joint source/FEC rate selection for quality-optimal MPEG-2 video delivery. IEEE Transactions on Image Processing **10**(12), 1815–1825 (2001)
21. Garcia, M.N., Raake, A.: Parametric packet-layer video quality model for IPTV. In: Int. Conference on Information Science, Signal Processing and their Applications (2010)

22. Garcia, M.N., Schleicher, R., Raake, A.: Impairment-factor-based audiovisual quality model for IPTV: Influence of video resolution, degradation type, and content type. In: EURASIP Journal on Image and Video Processing, vol. 2011, pp. 1–14 (2011)
23. Gros, L., Chateau, N.: Instantaneous and overall judgements for time-varying speech quality: Assessments and relationships. Acta Acustica **87**(3), 367–377 (2001)
24. Hands., D.: A basic multimedia quality model. IEEE Transactions on Multimedia **6**(6), 806–816 (2004)
25. Hossfeld, T., Egger, S., Schatzand, R., Fiedler, M., Masuch, K., Lorentzen, C.: Initial delay vs. interruptions: Between the devil and the deep blue sea. In: Int. Workshop on Quality of Multimedia Experience (2012)
26. Hossfeld, T., Schatz, R., Biersack, E., Plissonneau, L.: Internet video delivery in youtube: From traffic measurements to quality of experience. In: Springer book on Data Traffic Monitoring and Analysis: From measurement, classification and anomaly detection to Quality of experience (2013)
27. Huynh-Thu, Q., Garcia, M.N., F., S., Corriveau, P.J., Raake, A.: Study of rating scales for subjective quality assessment of high-definition video. IEEE Trans. Broadcast. **57**(1), 1–14 (2011)
28. Huynh-Thu, Q., Ghanbari, M.: The accuracy of psnr in predicting video quality for different video scenes and frame rates. Telecommunication Systems **49**(1), 35–48 (2012)
29. ITU-R Rec. BT.500: Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union (2012)
30. ITU-R Rec. BT.500-12: Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, CH–Geneva (2009)
31. ITU-R Rec. BT.710-4: Subjective Assessment Methods For Image Quality in High-Definition Television. International Telecommunication Union, CH–Geneva (1998)
32. ITU-R Rec. P.1202: Parametric non-intrusive bitstream assessment of video media streaming quality. International Telecommunication Union (2012)
33. ITU-R Rec. P.1202.1: Parametric non-intrusive bitstream assessment of video media streaming quality - lower resolution application area. International Telecommunication Union (2012)
34. ITU-R Rec. P.1202.2: Parametric non-intrusive bitstream assessment of video media streaming quality - higher resolution application area. International Telecommunication Union (2012)
35. ITU-R SG12 COM 12-37: Extension of combined audio/video quality model. Bellcore, USA (1994)
36. ITU-T Rec. G . 1070: Opinion model for video-telephony applications. International Telecommunication Union (2007)
37. ITU-T Rec. G.1080: Quality of Experience requirements for IPTV services. International Telecommunication Union (2008)
38. ITU-T Rec. H.264: Advanced video coding for generic audiovisual services (2013)
39. ITU-T Rec. J.144: Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. International Telecommunication Union (2001)
40. ITU-T Rec. J.341: Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference. International Telecommunication Union (2011)
41. ITU-T Rec. P.10/G.100 Amd 2: Vocabulary for performance and quality of service (2008)
42. ITU-T Rec. P.1201.1: Parametric non-intrusive assessment of audiovisual media streaming quality - lower resolution application area. International Telecommunication Union (2012)
43. ITU-T Rec. P.1201.2: Parametric non-intrusive assessment of audiovisual media streaming quality - higher resolution application area. International Telecommunication Union (2012)
44. ITU-T Rec. P.910: Subjective video quality assessment methods for multimedia applications. International Telecommunication Union, CH–Geneva (1999)
45. ITU-T Rec. P.910: Subjective video quality assessment methods for multimedia applications. International Telecommunication Union (2008)

46. ITU-T Rec. P.911: Subjective audiovisual quality assessment methods for multimedia applications. International Telecommunication Union, CH–Geneva (1998)
47. ITU-T Rec. P.920: Interactive test methods for audiovisual communications. International Telecommunication Union, CH–Geneva (2000)
48. ITU-T Rec. P.930: Principles of a reference impairment system for video. International Telecommunication Union (1996)
49. ITU-T SG12 COM 12-20: Experimental combined audio/video subjective test method. Bellcore, USA (1993)
50. ITU-T SG12 D.038: Results of an audio-visual desktop teleconferencing subjective experiment. NTIA/ITS, USA (1998)
51. Jang, Y.H., Noh, J.M.: Customer acceptance of IPTV service quality. International Journal of Information Management **31**(6), 582–592 (2011)
52. Jones, C., Atkinson, D.: Development of opinion-based audio-visual quality models for desktop video-teleconferencing. In: IEEE Int. Workshop on Quality of Service (1998)
53. Jumisko-Pyykkö, S., Häkkinen, J.: Profiles of the evaluators: impact of psychographic variables on the consumer-oriented quality assessment of mobile television. In: International Journal of Information Management, pp.68210L-68210L-14 (2008)
54. Jumisko-Pyykkö, S., Strohmeier, D.: Cognitive styles and visual quality. In: IS&T/SPIE Electronic Imaging,vol. 4, no. 4, pp.86670K–86670K (2013)
55. Kanumuri, S., Cosman, P., Reibman, A., Vaishampayan, V.: Modeling packet-loss visibility in mpeg-2 video. IEEE Transactions on Multimedia **8**(2), 341–355 (2006)
56. Kanumuri, S., Subramanian, S., Cosman, P., Reibman, A.: Predicting h.264 packet loss visibility using a generalized linear model. In: IEEE Int. Conference on Image Processing, pp. 2245–2248 (2006)
57. Keimel, C., Habigt, J., Diepold, K.: Hybrid no-reference video quality metric based on multiway plsr. In: Proceedings of the 20th European Signal Processing Conference, pp. 1244–1248 (2012)
58. Keimel, C., Habigt, J., Klimpke, M., Diepold, K.: Design of no-reference video quality metrics with multiway partial least squares regression. In: Int. Workshop on Quality of Multimedia Experience, pp. 49–54 (2011)
59. Keimel, C., Klimpke, M., Habigt, J., Diepold, K.: No-reference video quality metric for HDTV based on H.264/AVC bitstream features. In: IEEE Int. Conference on Image Processing, pp. 3325–3328 (2011)
60. Kohlrausch, A., van de Par., S.: Audio-visual interaction in the context of multi-media applications. In: J.Blauert (Ed.), Communication Acoustics, pp. 109–138 (2005)
61. Le Callet, P., Möller, S., Perkis, A.: Qualinet white paper on definitions of quality of experience. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003) (2012)
62. Lee, K., Park, J., Lee, S., Bovik, A.: Temporal pooling of video quality estimates using perceptual motion models. In: IEEE Int. Conference on Image Processing, pp. 2493–2496 (2010)
63. Lee, S.O., Jung, K.S., Sim, D.G.: Real-time objective quality assessment based on coding parameters extracted from H.264/AVC bitstream. IEEE Transactions on Consumer Electronics **56**(2), 1071–1078 (2010)
64. Liao, N., Chen, Z.: A packet-layer video quality assessment model with spatiotemporal complexity estimation. EURASIP Journal on Image and Video Processing **5** (2011)
65. Lin, T.L., Kanumuri, S., Zhi, Y., Poole, D., Cosman, P., Reibman, A.: A versatile model for packet loss visibility and its application to packet prioritization. IEEE Transactions on Image Processing **19**(3), 722–735 (2010)
66. Lin, X., Ma, H., Luo, L., Chen, Y.: No-reference video quality assessment in the compressed domain. IEEE Transactions on Consumer Electronics **58**(2), 505–512 (2012)
67. Naccari, M., Tagliasacchi, M., Tubaro, S.: No-reference video quality monitoring for H.264/AVC coded video. IEEE Transactions on Multimedia **11**(5), 932–946 (2009)
68. Next Generation Mobile Networks: ”service quality definition and measurement”. White Paper (2013)

69. Ninassi, A., Le Meur, O., Le Callet, P., Barba, D.: Considering temporal variations of spatial visual distortions in video quality assessment. Selected Topics in Signal Processing, IEEE Journal of **3**(2), 253–265 (2009)
70. Park, J., Seshadrinathan, K., Lee, S., Bovik, A.: Video quality pooling adaptive to perceptual distortion severity. IEEE Transactions on Image Processing **22**(2), 610–620 (2013)
71. Pinson, M., Ingram, W., Webster, A.: Audiovisual quality components. In: IEEE Signal Processing, vol. 28, pp. 60–67 (2011)
72. Pinson, M., Janowski, L., Pepion, R., Huynh-Thu, Q., Schmidmer, C., Corriveau, P., Younkin, A., Le Callet, P., Barkowsky, M., Ingram, W.: The influence of subjects and environment on audiovisual subjective tests: An international study. IEEE Journal of Selected Topics in Signal Processing **6**(6), 640–651 (2012)
73. Pinson, M., Wolf, S.: Comparing subjective video quality testing methodologies. In: Proc. SPIE (2003)
74. Pinson, M.H., Wolf, S.: A new standardized method for objectively measuring video quality. IEEE Transactions on Broadcasting **50**(3), 312–322 (2004)
75. Raake, A., Gustafsson, J., Argyropoulos, S., Garcia, M.N., Lindegren, D., Heikkilä, G., Pettersson, M., Feiten, B., List, P.: [monitoring the quality of] ip-based mobile and fixed network audiovisual media services. IEEE Signal Processing Magazine **28**(6), 68–79 (2011)
76. Redi, J.: Visual quality beyond artifact visibility. In: SPIE Human Vision and Electronic Imaging XVIII (2013)
77. Reibman, A., Kanumuri, S., Vaishampayan, V., Cosman, P.: Visibility of individual packet losses in mpeg-2 video. In: IEEE Int. Conference on Image Processing, vol. 1, pp. 171–174 (2004)
78. Reibman, A.R., Vaishmpayan, V.A., Sermadevi, Y.: Quality monitoring of video over a packet network. IEEE Transactions on Multimedia **6**(2), 327–334 (2004)
79. Ries, M., Puglia, R., Tebaldi, T., Nemethova, O., Rupp, M.: Audiovisual quality estimation for mobile video services. In: 2nd IEEE Symposium on Wireless Communication Systems (2005)
80. Seshadrinathan, K., Bovik, A.: Temporal hysteresis model of time varying subjective video quality. In: IEEE Int. Conference on Acoustics, Speech and Signal Processing, pp. 1153–1156 (2011)
81. Seufert, M., Slanina, M., Egger, S., Kottkamp, M.: "to pool or not to pool": A comparison of temporal pooling methods for http adaptive video streaming. In: Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (2013)
82. Shi, Z., Chen, P., Feng, C., Huang, L., Xu, W.: Research on quality assessment metric based on H.264/AVC bitstream. In: Int. Conference on Anti-Counterfeiting, Security and Identification, pp. 1–5 (2012)
83. Sodagar, I.: The MPEG-DASH standard for multimedia streaming over the internet. MultiMedia, IEEE **18**(4), 62–67 (2011)
84. Staelens, N., Van den Broeck, W., Pitrey, Y., Vermeulen, B., Demeester, P.: Lessons Learned during Real-life QoE Assessment. In: EuroITV - 10th European Conference on Interactive TV (2012)
85. Staelens, N., Deschrijver, D., Vladislavleva, E., Vermeulen, B., Dhaene, T., Demeester, P.: Constructing a No-Reference H.264/AVC Bitstream-based Video Quality Metric using Genetic Programming-based Symbolic Regression. IEEE Transactions on Circuits and Systems for Video Technology (2013). To appear
86. Staelens, N., Moens, S., Van den Broeck, W., Mariën and, I., Vermeulen, B., Lambert, P., Van de Walle, R., Demeester, P.: Assessing quality of experience of IPTV and Video on Demand services in real-life environments. IEEE Transactions on Broadcasting **56**(4), 458–466 (2010)
87. Staelens, N., Sedano, I., Barkowsky, M., Janowski, L., Brunnström, K., Le Callet, P.: Standardized toolchain and model development for video quality assessment - the mission of the joint effort group in VQEG. In: Int. Workshop on Quality of Multimedia Experience (2011)

88. Staelens, N., Van Wallendael, G., Crombecq, K., Vercammen, N., De Cock, J., Vermeulen, B., Van de Walle, R., Dhaene, T., Demeester, P.: No-Reference Bitstream-based Visual Quality Impairment Detection for High Definition H.264/AVC Encoded Video Sequences. IEEE Transactions on Broadcasting **58**(2), 187–199 (2012)
89. Strohmeier, D.: Open profiling of quality: a mixed methods research approach for audiovisual quality evaluations. In: SIG Multimedia Rec.,vol. 4, no. 4, pp.5-6 (2012)
90. Tao, S., Apostolopoulos, J., Guerin, R.: Real-time monitoring of video quality in IP networks. IEEE/ACM Transactions on Networking **16**(5), 1052–1065 (2008)
91. Valenzise, G., Magni, S., Tagliasacchi, M., Tubaro, S.: No-reference pixel video quality monitoring of channel-induced distortion. IEEE Transactions on Circuits and Systems for Video Technology **22**(4), 605–618 (2012)
92. Verscheure, O., Frossard, P., Hamdi, M.: User-oriented qos analysis in mpeg-2 video delivery. Real-Time Imaging **5**(5), 305–314 (1999)
93. Waltl, M., Timmerer, C., Hellwagner, H.: A test-bed for quality of multimedia experience evaluation of sensory effects. In: Int. Workshop on Quality of Multimedia Experience, pp. 145 –150 (2009)
94. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
95. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measure. Signal Processing: Image Communication **19**(2), 121–132 (2004)
96. Weniger, S.: User adoption of IPTV: A research model. In: International Bled eConference (2010)
97. Winkler, S., Faller, C.: Perceived audiovisual quality of low-bitrate multimedia content. IEEE Transactions on Multimedia **8**(5), 973–980 (2006)
98. Winkler, S., Mohandas, P.: The evolution of video quality measurement: From PSNR to hybrid metrics. IEEE Transactions on Broadcasting **54**(3), 660–668 (2008)
99. Wu, H.R., Rao, K.R.: Digital Video Image Quality and Perceptual Coding. Taylor & Francis, USA–Boca Raton, FL (2006)
100. Xue, Y., Song, Y., Ou, Y.F., Wang, Y.: Video adaptation considering the impact of temporal variation on quantization stepsize and frame rate on perceptual quality. In: Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, pp. 70–74 (2013)
101. Yamada, T., Yachida, S., Senda, Y., Serizawa, M.: Accurate video-quality estimation without video decoding. In: Int. Conference on Acoustics, Speech, and Signal Processing (2010)
102. Yamagishi, K., Hayashi, T.: Parametric packet-layer model for monitoring video quality of IPTV services. In: IEEE Int. Conference on Communications (2008)
103. Yamagishi, K., Okamoto, J., Hayashi, T., Takahashi, A.: No reference video-quality-assessment model for monitoring video quality of IPTV services. In: IEICE Transactions on Communications (2012)
104. Yang, F., Song, J., Wan, S., Wu, H.: Content-adaptive packet-layer model for quality assessment of networked video services. IEEE Journal of selected topics in Signal Processing **6**(6), 672–683 (2012)
105. You, F., Zhang, W., Xiao, J.: Packet loss pattern and parametric video quality model for IPTV. In: Int. Conference on Computer and Information Science (2009)
106. You, J., Reiter, U., Hannuksela, M., Gabbouj, M., Perkis, A.: Perceptual-based quality assessment for audio-visual services: a survey. In: Journal of Signal Processing: Image Communication (2010)
107. Yuen, M., Wu, H.: A survey of hybrid MC/DPCM/DCT video coding distortions. In: EURASIP Journal Signal Processing, vol. 70 (1998)
108. Zieliński, S., Rumsey, F., Bech, S.: On some biases encountered in modern audio quality listening testsa review. J. Audio Eng. Soc. **56**(6), 427–451 (2008)