Techniques for measuring quality of experience

Kuipers, FA; Kooij, RE; de Vleeschauwer, D; Brunnstrom, K

DOI
10.1007/978-3-642-13315-2_18

Publication date
2010

Document Version
Accepted author manuscript

Published in
Proceedings of the 8th international conference on wired/wireless internet communications

Citation (APA)
Kuipers, FA., Kooij, RE., de Vleeschauwer, D., & Brunnstrom, K. (2010). Techniques for measuring quality of experience. In s.n. (Ed.), *Proceedings of the 8th international conference on wired/wireless internet communications* (pp. 216-227). Springer. https://doi.org/10.1007/978-3-642-13315-2_18

Techniques for Measuring Quality of Experience

Fernando Kuipers[1], Robert Kooij[2], Danny De Vleeschauwer[3], and Kjell Brunnström[4]

[1] Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands
F.A.Kuipers@tudelft.nl
[2] TNO, Brassersplein 2, 2612 CT, Delft, The Netherlands robert.kooij@tno.nl
[3] Alcatel-Lucent Bell NV, Copernicuslaan 50, 2018 Antwerp, Belgium
danny.de_vleeschauwer@alcatel-lucent.com
[4] Netlab, Acreo AB, Kista, Sweden Kjell.Brunnstrom@acreo.se

Abstract. Quality of Experience (QoE) relates to how users perceive the quality of an application. To capture such a subjective measure, either by subjective tests or via objective tools, is an art on its own. Given the importance of measuring users' satisfaction to service providers, research on QoE took flight in recent years. In this paper we present an overview of various techniques for measuring QoE, thereby mostly focusing on freely available tools and methodologies.

1 Introduction

In the 40 years of its existence, the Internet has gradually evolved from a small network where connectivity was key, to a large media-rich network in which the user is placed more and more central. Users do not only just consume content, but have also started actively producing content. Hand-in-hand with this evolution to a media-rich Internet, the user requirements have transcended requirements on connectivity and users now expect services to be delivered in par with their demands on quality. Research on how to measure user Quality of Experience (QoE) has consequently also blossomed in recent years. In the International Telecommunication Union (ITU) several standards related to QoE have been proposed or are under development. We refer to Takahashi et al. [18] for an overview of these standardization activities. According to the ITU-T Focus Group on IPTV (FG IPTV), *Quality of Experience (QoE) refers to the overall acceptability of an application or service, as perceived subjectively by the end-user*. QoE thereby includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.), where overall acceptability may be influenced by user expectations and context. This definition explicitly refers to QoE as a subjective measure and properly measuring QoE should therefore involve tests with actual users, which is a time-consuming and costly process. For service and network providers it is preferable to have tools that objectively reflect within reasonable accuracy the subjective mean opinion score of users.

In this paper we present a set of techniques to measure QoE. The paper is not an exhaustive survey of the state-of-the-art in QoE, but instead provides a comprehensive palette of freely available tools and methods for conducting various QoE experiments. The remainder of the paper is organized as follows. In Section 2 we discuss the various parameters that affect QoE. The following three sections discuss measuring QoE for audio (Section 3), gaming (Section 4), and video (Section 5). Section 6 delves into the use of QoE in designing applications, and we conclude in Section 7.

2 QoE Framework

What makes QoE a somewhat elusive concept is that the parameters that define QoE may differ per service. For instance, zapping times play a role in television services, while in gaming it is not an issue. However, it is clear that QoE comprises more than just audio and video quality. We can classify the parameters that affect QoE into three groups:

1. The quality of the video/audio content at the source.
2. Quality of Service (QoS), which refers to the delivery of content over the network.
3. Human perception, which includes expectations, ambiance, etc.

The quality of the content relates to the kind of codec used, for instance MPEG-2 or MPEG-4, bit-rate, etc. The QoS parameters that affect the performance of streaming services most are bandwidth, delay, jitter, and packet loss.

The first two QoE categories are fairly easily quantified, while the latter one is not. Human perception is usually captured by a Mean Opinion Score (MOS), which reflects the appraisal of some test panel. The MOS is expressed on a five-point scale (ITU-T P.800), where 5 = *excellent*, 4 = *good*, 3 = *fair*, 2 = *poor*, 1 = *bad*. The minimum threshold for acceptable quality corresponds to a MOS of 3.5.

In general, there are three possible methodologies for measuring QoE:

1. The no-reference model has no knowledge of the original stream or source file and tries to predict QoE by monitoring several QoS parameters in real-time.
2. The reduced-reference model has some limited knowledge of the original stream and tries to combine this with real-time measurements to reach a prediction on the QoE.
3. The full-reference model assumes full access to the reference video, possibly combined with the measurements conducted in a real-time environment.

The first model fits under the umbrella of the second, which on its turn can be brought under the third. The full-reference model therefore should be able to give the best accuracy, but it is a method that can only be applied if one has control over both end systems. A no-reference model can be more easily adopted, but might not always give accurate results. In the following section we will list both a no-reference and a full-reference model for capturing speech QoE.

3 QoE of Speech and Audio

The most prominent example of a no-reference model is the E-model (ITU-T Rec. G.107). It predicts the quality users experience during a voice conversation based on the end-device characteristics and the transport parameters. These characteristics and parameters are plugged into some functions internal to the E-model of which the coefficients were tuned based on subjective experiments. The E-model determines a rating R:

$$R = R_0 - I_s - I_d - I_e + A$$

where R_0 is the basic signal-to-noise ratio, I_s takes into account phenomena that occur simultaneously with the speech signal (like the loudness of the speech signal and the side-tone and quantization effects), I_d groups impairments associated with delay (such as, impairments due to echo and loss of interactivity), I_e accumulates the effects associated with special equipment (for example, the use of a low bit rate codec or packet loss), and A is an advantage factor (i.e., a decrease in R-rating a user is willing to tolerate because he or she has a certain advantage, e.g., being mobile).

While for traditional telephone conversations the factors R_0 and I_s were the most important, for packet-based networks (e.g., for VoIP applications), the terms I_d and I_e play the most prominent role (and A is considered to be just an additional budget in some mobile applications). Leaving out the advantage factor A and for end devices that are well tuned, $R_0 - I_s$ usually lies around 95. If echo is properly controlled (which we can safely assume in modern day applications), the term I_d remains practically constant for a one-way delay below 150 ms. If this delay (also termed interactivity bound) is exceeded, I_d increases (and R decreases) by about 1 point on the R-scale per 10 ms additional delay [10]. The term I_e includes the effect of using a low bit rate codec and the impact of packet loss. Each standard low bit rate codec has an associated I_e function that determines how much distortion the codec itself introduces and how much additional impairment is introduced by packet loss. It is not uncommon that a codec introduces 10 points of impairment on the R-scale and that per percent of packet loss an additional 5 points of impairment on the R-scale are introduced. The precise curves associated with various narrow-band codecs can be found in ITU-T Rec. G.107.

The rating R of the E-model can be translated to a MOS value as follows:

$$MOS = \begin{cases} 1 & R < 0 \\ 1 + 0.035R + (R-60)(100-R)\frac{7R}{10^6} & 0 \leq R \leq 100 \\ 4.5 & R > 100 \end{cases}$$

However, normally R is interpreted directly via the quality classes defined in ITU-T Rec. G.107. The current version of the E-model includes an extension to wideband codecs in an appendix, but this appendix is not normative yet. An alternative computational model is presented in Chen et al. [4], in which the user satisfaction is determined based solely on the parameters bit rate, jitter, and round-trip time.

An example of a full-reference model is the Perceptual Evaluation of Speech Quality (PESQ) model (ITU-T Rec. P.862). It assesses the listening-only quality of narrow-band speech encoded, packetized and sent over a (possibly packet-loss-prone) network. This model compares the original with the received signal and determines which differences result in annoying artifacts. For this purpose both signals need to be suitably aligned (which forms an integral part of the model). There is ongoing research on extending PESQ to wideband speech in Study Group 12 of the ITU-T.

While the listening quality for speech can be objectively measured by PESQ, the standard for objectively measuring audio QoE is called PEAQ (Perceptual Evaluation of Audio Quality). The PEAQ standard ITU-R BS.1387 was set in 1998 and was last updated in 2001. Like PESQ, PEAQ is a full-reference algorithm that analyzes the audio signal sample-by-sample after a temporal alignment of corresponding excerpts of reference and test signal. The algorithm takes perceptual properties of the human ear into account, integrating multiple model output variables into a single metric. PEAQ characterizes the perceived audio QoE, expressed as a MOS score, as subjects would do in a listening test according to ITU-R BS.1116. For educational use, there exists a free cross-platform program called Peaqb which accomplishes the same functions in a limited manner, as it has not been validated with the ITU data. PQevalAudio, another not validated implementation of the PEAQ basic model for educational use, is available from the TSP Lab of McGill University.

4 QoE of Interactive Gaming

In this section we consider the QoE of so-called First Person Shooter (FPS) games, because other types of online games, such as real-time strategy games and multiplayer role-playing games, pose less strict requirements with respect to network quality [7].

Many papers investigated the impact of delay, jitter, and packet loss on gameplay for FPS games, see for instance [1], [2], [22]. Most of these works focus on network performance metrics that game players can tolerate. Like the E-model for speech, Wattimena et al. [21] developed a G-model model for the prediction of the QoE of FPS games. Subjective tests were performed with FPS games such as Quake IV and later Quake III, Unreal Tournament, Counter Strike, and Halo. The finding was that the G-model produced a well-founded lower limit of gaming QoE.

In order to apply the G-model one first determines the round-trip time (RTT) of 10000 UDP packets that are send over the network. Define $PING$ as the average of all RTTs and $JITTER$ as the 99.9%-quantile of all RTT's minus the minimum RTT (ITU-T Rec. Y.1541).

We formulate the G-model as

$$MOS_{Gaming} = \max\left(4.33 - 3.08 \cdot 10^{-9}X^3 + 1.18 \cdot 10^{-5}X^2 - 1.15 \cdot 10^{-2}X, 1\right)$$

where the network impairment $X = \min\left(PING + 0.686 * JITTER, 650\right)$. This formulation is slightly different from [21], since we now use a standardized way of measuring $JITTER$.

According to [21] the correlation between the original G-model and the subjective data (of 33 test subjects) is very high (correlation coefficient $\rho = 0.98$). Note that, according to the G-model, FPS games are insensitive to packet loss.

5 QoE of Video

In this section, we discuss how to measure the QoE of streaming video services at the end user. Visually the most important factors for the video quality are: viewing distance, display size, resolution, brightness, contrast, sharpness, colorfulness, and naturalness. A distinction should be made between fidelity and quality, where fidelity stands for the closeness of the processed video to the original. For instance, for a low-quality original, a high-fidelity reproduction will still have low quality.

Visible distortions will most often lower the perceived quality. These may be introduced by lossy compression. The most widely used codecs use block-based Discrete-Cosine Transform (DCT) with motion compensation followed by quantization of the coefficients as a compression scheme, e.g. MPEG-2 and ITU-T H.264. The main distortion introduced by such codecs are: blockiness, blurring, color bleeding, ringing, DCT basis image effect, staircase effects on slanted lines, false edges, jagged motion, chrominance mismatch, mosquito noise, flickering, and aliasing.

For video containing audio the synchronization between video and audio is also an important perceptual factor.

Another source of errors occurs during transmission. The three most important transmission artefacts are packet loss, delay, and delay variations (jitter). The visual effect of lost information is highly dependent on the codec and on the type of information that is lost. Some errors might be concealed using error concealment strategies, whereas others have severe effects.

Finally, also channel zapping time can be identified as an important factor.

The perceived quality factors that are listed for streaming video also hold for video conferencing. In addition, due to its interactive nature, the same strict delay requirements as for VoIP hold. We shall address several of the above-mentioned components in the following subsections.

5.1 Visual quality

Commonly used video quality metrics are:

- Peak-Signal-to-Noise-Ratio (PSNR) gives the ratio (in dB) between the signal power of the original signal versus the power of a reconstructed compressed signal. PSNR is usually derived via the mean squared error (MSE) between the two signals in relation to the maximum possible value of the luminance of the images. Although PSNR may not accurately reflect the QoE, as demonstrated in [9], it continues to be a popular method to evaluate the quality difference among videos.
- Video Quality Metric [17] (VQM) is a software tool developed by the Institute for Telecommunication Science (ITS) to objectively measure perceived video quality. It measures the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block distortion, and color distortion, and combines them into a single metric, by using a linear combination of these parameters. The Video Quality Experts Group[5] (VQEG) Phase II validation tests show that VQM has a high correlation with subjective video quality scores and as a result it has been adopted by ANSI (ANSI T1.801.03-2003), and by ITU-T (ITU-T J.144, and ITU-R BT.1683) along with three other metrics as standard for measuring video quality.
- Structural Similarity Index [19] (SSIM) uses a structural distortion based measurement approach. Structure and similarity in this context refer to samples of the signals having strong dependencies between each other, especially when they are close in space [20]. The rationale is that the human vision system is highly specialized in extracting structural information from the viewing field and it is not specialized in extracting the errors.

An important step in full- and reduced-reference video quality metrics to reach an accurate quality score is that the sampled videos need to be calibrated. The calibration consists of estimating and correcting the spatial and temporal shift of the processed video sequence with respect to the original video sequence. This is usually an integral part of the more sophisticated metrics, but it needs to be added separately to, for instance, PSNR to give sensible results.

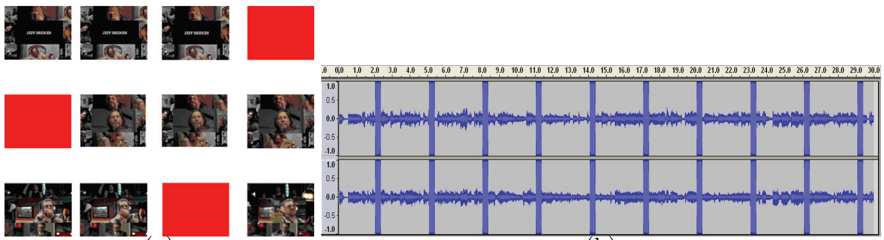
The standard for video quality measurements by the ITU covers full-reference (ITU-T Rec. J.144 and ITU-R BT.1683) and reduced-reference (ITU-T Rec. J.249) methods for standard definition TV, as well as a full-reference (ITU-T Rec. J.247) and reduced-reference (ITU-T Rec. J.146) for multimedia. All these standards contain various metrics that have performed statistically equally well in evaluations performed by VQEG. However, there are currently no standards covering no-reference methods. A trend in recent years has been to propose so-called hybrid models that incorporate information from the bit stream as well as quality evaluation on the video itself. Although a good correlation with subjective quality can be reached, the disadvantage is that the models are tailored to a specific codec [5].

[5] http://www.its.bldrdoc.gov/vqeg/

5.2 Audio-Video Synchronization

Audio-video synchronization refers to the relative timing of sound and image portions of a television program, or movie.

The ITU-R BT.1359-1 recommendation states that the tolerance from the point of capture to the viewer/listener shall be no more than 90 ms audio leading video to 185 ms audio lagging behind video.



(a) (b)

Fig. 1. (a) Video with red markers, and (b) accompanying audio with corresponding beep markers.

Analyzing A/V synchronization can be done with an "artificially generated" video test sample (e.g., see Fig. 1). The test sample includes a video component and an audio component. The video component and the audio component both contain markers. The video marker displays between a first video state and a second video state, a red full screen image. Similarly, the audio waveform alternates between a first audio state and a second audio state, an audio "beep." The video and audio waveforms are temporally synchronized to transition from one state to another at the same time.

By comparing the audio and video tracks at the receiving end, any dis-synchronization can be noticed. This approach was applied in [16] to test the A/V synchronization of four video conferencing applications. Also the audio delays among participants was measured by injecting in the video an artificial DTMF (dual-tone multi-frequency) tone. The audio was sent and recorded at one client. Other participants kept their speakers and microphones on, but did not produce extra audio. Based on the recorded audio tracks, the difference between the time the audio marker was sent from the client and the time the same audio marker was heard again at the client was extracted. The time difference is approximately twice the one-way audio delay plus the processing delay at a client.

5.3 Network QoS

As we have argued, QoE is determined by more than the QoS provided by the network. However, network or service providers only have control over their own equipment/network, and therefore it is important for them to know the relation between QoS and QoE. Accurately measuring QoS parameters like bandwidth and delay, is a research topic on its own, but fortunately QoE is an end-to-end measure that sees the network as a black box. This means that we can omit the details of the network and correlate the QoE of certain applications to (artificially introduced) artefacts like delay and packet losses. EvalVid [12] is a tool that facilitates just that. It can be used with the network simulator NS-2 to simulate the network environment and trace the video packets in different controlled communication conditions [11]. A video source file (usually in raw YUV format) is encoded into a compressed video file (either MPEG-4 or H.264). Then the compressed video file is time-stamped and transmitted via UDP packets over a real or simulated network to the receiver. By comparing the time-stamps of packets (and their type) at sender and receiver, frame loss and frame jitter can be computed. Also the PSNR can be returned by EvalVid. By precisely knowing and controlling the QoS parameters of the simulated environment, a correlation between QoE and QoS parameters could be established.

5.4 User Synchronization

While watching a football match it could be disturbing to hear the neighbors scream "GOAL" while you still are watching the pre-goal action. Such phenomena are referred to as user or peer lags. While watching the same channel, users' content might not be synchronized. Measuring the different "lag delays" can be done by using another artificial video displaying a timer. Each second (or at some other precision) a sequential number is shown. By using PlanetLab, or other similar platforms that give control over different geographically dispersed computers, one can set up several clients that are under full control. This approach has been applied in [15]. Unsynchronized users can especially affect the QoE when they are experienced in multi-party communications.

5.5 Start-up and zapping time

One of the key elements of QoE of IPTV is how quickly users can change between TV channels, which is called channel zapping. Kooij et al. [13] conducted a number of subjective tests in order to get insight in the relation between QoE and zapping time. For the tests described in [13], during channel zapping, a black screen was visible which contained the number of the target channel. The test subjects (21 in total) could select one of the following five opinion scores: 5 = *excellent zapping quality*, 4 = *good zapping quality*, 3 = *fair zapping quality*, 2 = *poor zapping quality*, 1 = *bad zapping quality*.

The following model was proposed for the relation between zapping time (in seconds) and the QoE (expressed in MOS) of channel zapping:

$$MOS_{Zapping} = \max\left(\min\left(-1.02\ln(ZappingTime) + 2.65, 5\right), 1\right)$$

From this relation it was deduced that in order to guarantee a MOS of at least 3.5, we need to ascertain that zapping time < 430 ms.

The model presented in [13] was based on "lean forward" zapping, where test subjects were switching channels by clicking buttons with a mouse on a computer screen, and on fixed zapping times. Experiments reported in [14] include "lean backward" experiments, where the test subjects were sitting on a sofa with a remote control. For "lean backward" zapping the requirement for the zapping time could be relaxed to 670 ms. In [14] subjective experiments were also conducted with varying zapping times. The zapping time was uniformly distributed, with a variance bounded by the mean zapping time. In this case, to obtain a MOS rating of at least 3.5, the maximum allowed variance, and thus also the maximum allowed mean zapping delay, was 460 ms.

In [8] the QoE of channel zapping was assessed when, during zapping, advertisements are displayed, instead of a black screen. Based on Figure 2, we can conclude that:

- Users prefer advertisements only when the zapping time is sufficiently large.
- Short zapping times lead to insufficient time to view the advertisement. On the other hand, a long zapping time is also still not appreciated.

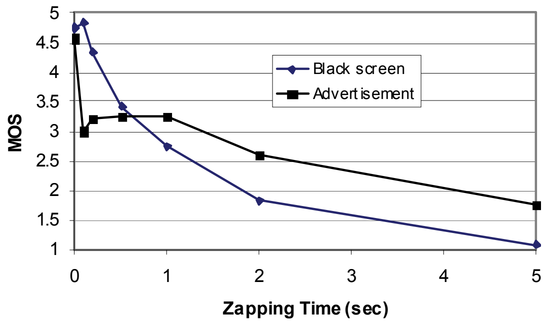


Fig. 2. MOS for "black screen" and "advertisement."

Analogous to the G-model for black screens, explained above, in [8] the following QoE model for advertisements was suggested:

$$MOS_{Ads} = \max\left(y_1, \min\left(y_2, y_3\right)\right)$$

where $y_1 = -15.8(ZappingTime) + 4.58$, $y_2 = 0.10 \ln(ZappingTime) + 3.27$, and $y_3 = -0.93 \ln(ZappingTime) + 3.27$.

In [15], with Wireshark and a small script that zaps from channel to channel, the zapping times of the P2PTV application SopCast have been measured. The used script simply starts a counter when a channel is clicked and it stops when enough data to be displayed has been fetched. The script was zapping among 20 popular and less popular channels. With an average zapping time of 50 seconds, SopCast faces an unacceptable delay. This delay was predominantly caused by the end systems (a SopCast buffer and a media player buffer needed to be filled) and by the process of finding peers to download content from. The actual network delay (QoS) was of minor influence.

6 Using QoE for design

In this last section, we show how QoE models can be used when designing a system. Distributing television over a packet-based network will be our case study. First the quality with which the television signal will be transported needs to be determined (which can be assessed with models of Section 5.1). This desired quality impacts the required bit rate, which normally has to be chosen as low as possible. Given the desired quality, the required bit rate mainly depends on the interval between Intra frames (I-frames). I-frames can be decoded without making reference to other frames in the sequence, while Predicted frames (P-frames) and Bidirectionally predicted frames (B-frames) are encoded making reference to other frames. I-frames typically require 2 to 4 times more bits than P-frames for a difficult and easy scene respectively, while P-frames in turn usually require 3 times more information than B-frames. Spacing the I-frames further apart decreases the bit rate. Figure 3 shows the bit rate penalty to be paid by choosing a smaller interval between I-frames.

I-frames are anchor points in the sequence: the effect of a lost packet stops propagating at I-frames and a user can only tune in at these anchor points. The longest time a user has to wait to tune into a channel is equal to the interval between I-frames. Section 5.5 argued that in order to reach a MOS of 3.5 the zapping delay can be at most 460 ms. Figure 3 shows that for such a small interval between I-frames the bit rate penalty is high.

In Degrande et al. [6] a system is discussed where the interval between I-frames is much larger than this 460 ms such that the bit rate penalty is small, while still maintaining a small zapping delay. For that purpose the most recent video information is kept in a circular buffer in the network. As soon as a user zaps to a channel, a recent I-frame is retrieved from that circular buffer and this allows the user to immediately tune in. However, since the user tunes in to an I-frame of the recent past (say an I-frame sent on the channel T seconds ago) he or she lags T s behind with respect to the current information that is sent over the channel. In order to catch up, the T s of video information needs to be burst to the user. This temporarily increases the bit rate for that user after a zap. The bursting rate has to be tuned such that this is not overwhelmingly large. However, the smaller the bursting rate is chosen, the longer the catch-up time is. Which of the two situations, 1) choosing a small interval between I-frames paying a high bit rate penalty, or 2) choosing a large interval between I-frames with a circular buffer to assist fast zapping and bursting to catch up, is best in terms of the required bit rate, depends on the frequency with which users are likely to change from one channel to the next.

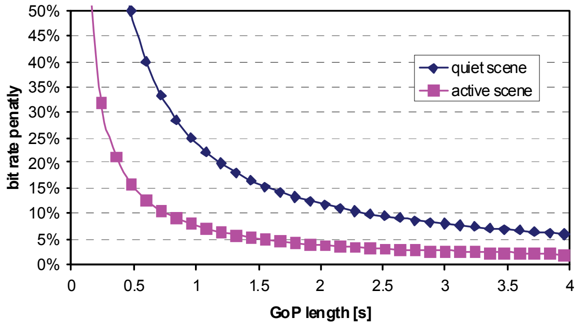


Fig. 3. Bit rate as a function of the interval between I-frames.

7 Conclusions

In this paper we have provided an overview of some of existing techniques and tools for conducting Quality of Experience (QoE) measurements. QoE is a subjective measure, which makes quantifying it difficult. Consequently, one could find drawbacks to each of the presented techniques in this paper. On the other hand, these techniques are some of the most accessible ones and they do provide us with some insight into QoE, which could for instance already be used in the design of an application or the tuning of a network.

References

1. G. Armitage, L. Stewart, "Limitations of using Real-World, Public Servers to Estimate Jitter Tolerance Of First Person Shooter Games," Proc. of ACM SIGCHI ACE2004 Conference, Singapore, June 2004.
2. T. Beigbeder, R. Coughlan, C. Lusher, J. Plunkett, E. Agu, M. Claypool, "The Effect of Loss and Latency on User Performance in Unreal Tournament 2003®", Proc. of NetGames'04, Oregon, USA, 2004.
3. C.J. Van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatiotemporal model of the human visual system," Proc. of SPIE, vol. 2668, pp. 450-461, 1996.
4. K.-T. Chen, C.-Y. Huang, P. Huang, and C.-L. Lei, "Quantifying Skype User Satisfaction," Proc. of SIGCOMM'06, Pisa, Italy, September 11-15, 2006.
5. A.G. Davis, D. Bayart, and D.S. Hands, "Hybrid No-Reference Video Quality Prediction," Proc. of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Bilbao, Spain, May 13-15, 2009.

6. N. Degrande, K. Laevens, D. De Vleeschauwer, and R. Sharpe, “Increasing the user perceived quality for IPTV services,” IEEE Communications Magazine, Vol. 46, Issue 2, pp. 94-100, Feb. 2008.
7. M. Dick, O. Wellnitz, and L. Wolf, “Analysis of factors affecting Players’ Performance and Perception in Multiplayer Games,” Proc. of NetGames’05, Hawthorne (NY), USA, 2005
8. B.E. Godana, R.E. Kooij, and O.K. Ahmed, “Impact of Advertisements during Channel Zapping on Quality of Experience,” Proc. of The Fifth International Conference on Networking and Services, ICNS 2009, Valencia, Spain, April 20-25, 2009.
9. Q. Huynh-Thu and M. Ghanbari, “Scope of validity of PSNR in image/video quality assessment,” Electronics letters, vol. 44, no. 13, June 19, 2008.
10. J. Janssen, D. De Vleeschauwer, M.J.C. Büchli, and G.H. Petit, “Assessing Voice Quality in Packet-Based Telephony,” IEEE Internet Computing, Vol. 6, No. 3, pp. 48-56, May/June 2002.
11. C.-H. Ke, C.-K. Shieh, W.-S. Hwang, and A. Ziviani, “An Evaluation Framework for More Realistic Simulations of MPEG Video Transmission,” Journal of Information Science and Engineering, vol. 24, pp. 425-440, 2008.
12. J. Klaue, B. Rathke, and A. Wolisz, “EvalVid - A framework for video transmission and quality evaluation,” Proc. of the International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, pp. 255-272, 2003.
13. R.E. Kooij, O.K. Ahmed, and K. Brunnström, “Perceived Quality of Channel Zapping,” Proc. of the fifth IASTED International Conference on Comm. Systems and Networks, Palma de Mallorca, Spain, pp. 155-158, Aug. 28-30, 2006.
14. R.E. Kooij, F. Nicolai, O.K. Ahmed, and K. Brunnström, “Model validation of channel zapping quality,” Proc. of Human Vision and Electronic Imaging Conf., Jan. 19-22, 2009.
15. Y. Lu, B. Fallica, F.A. Kuipers, R.E. Kooij, and P. Van Mieghem, “Assessing the Quality of Experience of SopCast,” Int. J. Internet Protocol Technology, vol. 4, no. 1, pp.11–23, 2009.
16. Y. Lu, Y. Zhao, F.A. Kuipers, and P. Van Mieghem, “Measurement Study of Multiparty Video Conferencing,” Proc. of IFIP Networking 2010, Chennai, India, May 10-14, 2010.
17. M. Pinson and S. Wolf, “A New Standardized Method for Objectively Measuring Video Quality,” IEEE Transactions on Broadcasting, vol. 50, no. 3, September 2004.
18. A. Takahasi, D. Hands, and V. Barriac, “Standardization Activities in the ITU for a QoE Assesment of IPTV,” IEEE Communications Magazine, pp. 78 - 84, February 2008.
19. Z. Wang and Q. Li, “Video quality assessment using a statistical model of human visual speed perception,” Journal of the Optical Society of America A, vol. 24, no. 12, pp. B61-B69, December 2007.
20. Z. Wang and A.C. Bovik, “Mean squared error: love it or leave it? - A new look at signal fidelity measures,” IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009.
21. F. Wattimena, R.E. Kooij, J.M. van Vugt, and O.K. Ahmed, “Predicting the perceived quality of a First Person Shooter: the Quake IV G-model,” Proc. of NetGames’06, Singapore, October 30-31, 2006.
22. S. Zander and G. Armitage, “Empirically Measuring the QoS Sensitivity of Interactive Online Game Players,” Proc. of Australian Telecommunications Networks & Applications Conference 2004, Sydney, Australia, December 2004.