

QoE for Mobile Streaming

Vlado Menkovski and Antonio Liotta
Eindhoven University of Technology
The Netherlands

1. Introduction

After its debut in personal computers and home entertainment devices, streaming video has found its place in the mobile environment as well. However, the characteristics of the mobile devices present a different set of challenges for delivering satisfactory video services. They are typically smaller in size designed to be handheld or carried on person in some manner. They have reduced screen size and somewhat limited computational power due to restrictions in size and power consumption. These technical challenges necessitate the adaptation of the content to the particular device so that successful reproduction is possible and desired quality is met. When the mobile device is exposed to unadjusted content in spatial resolution and frame-rate, unnecessary penalties in computational load are incurred mostly due to processing of the downscaling. This in turn leads to higher power consumption and can lower the quality if the device fails to execute the task on time. In general the issue of over and under-provisioning is present in any video streaming service. However, in mobile streaming it is particularly evident due to specific limitations of the mobile devices and the larger variety in the design of the user interface and the mode of interaction.

Another key feature of the mobile devices is that they rely on wireless access to connect to the outside world to gather information and use services. The types of wireless networks used are varied, consisting of a multitude of technologies that have different characteristics. This variability results in different bandwidth capability, round trip time and reliability of the link. Wireless connections are also more prone to errors because of exposure to various sources of noise and diminished reliability due to limitations in coverage. All these factors affect the networked services, but particularly multimedia services such as video streaming that rely on high throughput and high reliability of the link for successful playback (Wu et al., 2001).

Furthermore, due to the specific mechanics of consumption of the content in handheld devices the perception of quality is different than in desktop or large screen devices. This creates the need for separate exploration of the process of optimizing the tradeoff between the used resources and delivered quality from traditional video delivery systems.

To satisfy the ambition to deliver high quality of the service to its customers, providers need foremost an accurate estimation of the perceived quality. In order to be accurate in this estimation, they have to account for the unique characteristics of the mobile devices. These characteristics range from the way that the device is held, and used, to its screen size and computational power as well as the user expectations.

Quality of Experience (QoE) is a metric that quantifies this multifaceted, multidimensional factors that account for the perceived quality. QoE appears in the literature by different definitions, but generally it is agreed that QoE measures the quality experienced while using a service (Jain, 2004). Accurate estimation of QoE for video streaming on mobile services allows for control over the delivered quality and an implementation of user centric management of these services. User centric management focuses on managing the services according to the user or customer needs. In other words it offers constant quality to the user rather than constant amount of spent resources from the provider, which increases the value of the service to the customers (V. Menkovski et al., 2010).

This chapter begins with a discussion on the factors that affect the QoE in video streaming with particular interest on streaming for mobile devices. Next the chapter continues on to the objective and subjective methods for QoE estimation. This is followed by an elaboration on the impact that different video streaming technologies have on QoE. That section covers technologies for multimedia encoding such as video compression, transcoding, scalable video coding and multi description coding. Then a section on technologies for network transmission and reproduction of video and multimedia content follows. Finally we conclude with a summary of the current state of QoE and its possible future directions.

2. Factors that affect QoE in multimedia streaming

Video quality is affected by a multitude of factors in every stage of its existence from the creation of the content to transmission and finally reproduction. There is a vast amount of research on quality of the recording equipment, lightning conditions and other factors that contribute to quality at the content creation. However, these topics are out of scope of this chapter. Our interest here is to discuss how the quality is affected by the process of encoding, transport and presentation of the multimedia. Multimedia QoE is a complex metric that depends on many parameters. In the literature however QoE is modeled restricted to more specific aspects of a service. The authors of (Tran & Mellouk, 2010) implement a QoE model for network services by defining key performance indicators of service availability, accessibility, continuity, responsiveness, and delay. Other efforts focus on freezes in the playback (Huynh-Thu & Ghanbari, 2010) and their relationship with the spatial quality.

Adaptive services can vary the quality of the content as availability of resources changes. However, this variation of quality has negative effects on the QoE. The authors of (Zink et al., 2003) conclude that the rate and the amplitude of the variation are proportional to the degradation of the QoE. Audio and video synchronization also has a significant effect on the QoE. The investigation of media synchronization in (R. Steinmetz, 1996) concludes that the effect on quality depends on the type of content. In another investigation of audio and video correlation and lip synchronization (Mued, et al., 2003) concludes that the effects are different if the multimedia is of passive or active communication.

However, arguably the most important QoE factor in multimedia streaming is the video quality (S. Winkler & Mohandas, 2008). Significant amount of degradation of video quality comes from the encoding or the compression process. The main goal of the encoding process is to reduce the size of the original video. Uncompressed video requires very large storage space and is not suitable for transmission over restricted network channels. Lossy encoding

is necessary to accommodate the restrictions in data throughput in the transmission channel, storage and processing capabilities of the terminal devices. This loss of fidelity can come from the encoder itself or as a part of a pre-encoding process to downgrade the spatial and temporal resolution of the video. In this manner the loss of quality from this initial step can come from downscaling of spatial resolution, frame-rate or bit-rate. Similar degradation is present in the audio, when sampling rate, sampling frequency and bit-rate are downscaled.

The spatial resolution of the digital video is one of the key factors for the size of the video after encoding. The recording equipment usually has much higher resolution than what can be practically used in video streaming applications. Particularly for mobile devices the resolution needs to be adjusted to the limitations of the screen resolution, computational power or network throughput. Lowering the spatial resolution creates the effect of blockiness. Blockiness is a visual impairment effect, particularly visible around the edges. It makes the edges jagged and pixelized.

Another way to restrict the size is to decrease the color depth, which makes the amount of data per pixel smaller. However, the images with smaller color palette seem less natural. With the advent of screen technology and the increase in user expectation this technique is rarely used nowadays.

On the other hand, a more commonly used technique, in addition to the spatial resolution decrease, is temporal resolution decrease or decrease in frame-rate. Frame-rate is usually kept to less than 30 frames per second due to the characteristics of the human visual systems. However framerate acceptability depends on the type of content (Apteker et al., 1995). Certain types of content that have low mobility and small spatial resolution, frame-rates as low as 10 frames per second can be acceptable. This is particularly useful for very low bit-rate channels in mobile environments where lower frame-rates help achieve the required low throughput.

The digital video encoding process produces a video stream with a specific bit-rate. The bit-rate is directly linked to the quality of the video stream and most encoders accept bit-rate setting as input. It can be either set as a soft (indication) or as a hard (constraint) limit on the encoder in constant bit-rate encoding. In variable bit-rate the indicator is usually a quality setting and the resulting stream is with constant quality rather than having a constant bit-rate. Based on this setting and the complexity of the video, the encoder compresses the video with certain average bit-rate. The bit-rate is also a very important factor for the content delivery system because the costs incurred for data the transport and storage are directly associated with it. Furthermore, specific parts of the transmission have limitations on the throughput that they can handle. Similar limitations exist in the terminal devices.

Therefore the amount of bits required to encode the video, or bits per second of video (i.e. bit-rate), depends on the type of encoding algorithm, the complexity of the video and the desired quality. Typical MPEG-like algorithms will introduce increasingly larger amounts of blockiness and blurriness as the bit-rate is reduced. In other words the video data will be more coarsely quantized in the frequency domain, which will lead to blockiness in the decoded video. The encoder attempts to limit the blockiness effect on low spatial frequencies of the video, which are less perceptible to the viewers. However, very constrained compressions result in highly visible artifacts. Another type of artifact resulting from the encoding is blurriness. This one arises from inadequate temporal fidelity of the encoded video.

The audio is of course as important as the video in the overall QoE of multimedia content, and many cases even more important. Even so, due to the fact that it is significantly easier to compress audio (as it needs less resources), most delivery systems use higher-quality compression level on the audio and focus on optimizing the video compression.

Finally, significant amount of artifacts and impairments come from the transmission errors, delays and jitter in the data reception (Kanumuri, et al., 2006)(Reibman & V. Vaishampayan, 2003). Typical symptoms are visual artifacts due to errors, freezes in reproduction and startup delays. These have significant effect on the QoE and depend on the undying transmission technology, as discussed in the following sections.

3. Estimating the video quality

Since video quality has such a significant effect on the overall QoE of multimedia streaming services, this section is dedicated to a review of a range of methods for video quality assessment. Video quality estimation has been of significant importance since the early days of digital video, so many methods have been devised. These are generally grouped into 'objective' and 'subjective' methods. The objective methods focus on objectively measurable impairments present in the video. This is done by comparing the impaired video to the original unimpaired video, in a full reference approach. Furthermore, the reduced reference (RR) and no-reference methods (NR) use partial information or no information on the original video to estimate the quality. These methods are more flexible, because they do not need to have copies of the original material, but suffer from reduced accuracy. On the other hand, a subjective methods aim at quantifying quality based on direct response by the person. These are more accurate, as they better take into account the human perception factor. However, they are more costly and cumbersome.

3.1 Objective video quality methods

Video quality estimation benefits extensively from the exhaustive work done in the image quality domain. Many image quality assessment (IQA) methods are used for estimation of video quality by evaluating the quality of individual pictures and, then, averaging over time.

The method that cannot be omitted when an overview of objective video quality methods is made is the peak signal to noise ratio (PSNR) (equation 1). PSNR estimates the difference between the original image and the distorted one by calculating the mean squared error (MSE) between the two signals and giving the ratio between the maximum of the signal and the MSE. Regardless of its significant drawbacks (mainly its low accuracy) (Avcibaş, et al., 2002), PSNR is still very present in video quality analysis. It is easy to compute and provides a first impression on the quality achieved.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{orig}(i, j) - I_{imp}(i, j)]^2$$

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right)^2 \quad (1)$$

A more advanced IQA approach that tries to circumvent the drawbacks of PSRN is the structure similarity index method (SSIM) (Z. Wang, et al., 2004). SSIM focuses on the structural distortions in the image, not purely on the bit-errors, to calculate a more accurate quality value.

These approaches however do not consider the specific temporal effects of video. The video quality metric (VQM) model of Pinson and Wolf (Pinson & Wolf, 2004) is a linear combination of loss of spatial information, orientation shift, color impairments and moving edge impairments to calculate an overall video quality index. Because this model weighs in motion components, in addition to structural components, it covers a larger portion of the QoE factors in video.

A conceptually similar approach is DVQ (digital video quality), developed by Watson, et al. (Watson et al., 2001). In addition to the structural temporal distortions, DVQ incorporates factors such as light adaptation, luminance and chrominance distortions and contrast masking. These factors are motivated by the human visual system (HVS) models.

Another HVS-inspired method is PVQM (perceptual video quality metric) (Hekstra et al., 2002). This method computes the video quality index as a linear combination of three indicators: edginess of the luminance, normalized color error and temporal de-correlation.

To address the importance of the motion degradation for quality estimation, Seshadrinathan and Bovik developed the MOTion-based Video Integrity Evaluation (MOVIE) VQA index (Seshadrinathan & Bovik, 2009). This index calculates the impairments not only in space and time, but also in space-time. This algorithm as well as the other full-reference VQA makes a tradeoff between accuracy and computational resources. Even though MOVIE is an objective method that correlates well with subjective feedback from viewers, its high cost in computational power and memory limits its implementation in real-time systems.

A more suitable alternative for real-time quality estimation in content delivery systems are the RR and NR methods. The methods or models are usually much more specific than the FR methods and only deal with specific types of impairments and are not very accurate for general use.

Gunawan and Ghanbari have developed a RR method (Gunawan & Ghanbari, 2008) that uses local harmonic strength features from the original video to calculate the amount of impairments or quality of the affected video. Harmonic gains and loss correlate well with two very common types of impairment present in MPEG encoded video, i.e. blockiness and blurriness. This RR method has low overhead of the harmonic data of 160 to 400 bits per second, which is a negligible amount compared to the size of the video.

NR methods are even more flexible than RR because they are applicable to any video environment, even ones that do not have any information on the original source of the video. Naturally their accuracy and generality is highly constrained. NR models are frequently used to calculate the impact of transport errors on the delivered video. In (Reibman & V. Vaishampayan, 2003) authors present a model for estimating MSE caused by packet loss, by examining the video bit-stream. A single packet loss does not affect only the pixels of the video frame that lost information in that packet but, due to the temporal compression mechanisms of MPEG videos, these errors are propagated by the motion vectors in the subsequent frames. The location of the packet is of significant importance,

because different types of frames carry information of different nature. A similar approach for MPEG2 video is presented in (Kanumuri et al., 2006). This method uses different machine learning (ML) algorithms to predict the visibility of the lost packet on the presented video. The additional complexity that these NR methods face is that they are not aware of the decoder's approach to conceal the error. A typical concealment approach is zero-motion concealment, in which a lost macroblock is concealed by a macroblock in the same location from a previous frame. However, the visibility of this concealment depends on the content at this position, size of the screen and many other factors that are generally related to the overall QoE.

Another group of methods exist that also do not use any information about the original video, but falls in the group of Hybrid Subjective and Objective methods. These methods first start with feature extraction from the specific system that they are evaluating. These features are objectively measurable in the particular system. Then this feature space is explored and labeled with a subjective study that provides subjective feedback on the human perception of quality. This step can be done in a full reference manner by presenting to the participants both the unimpaired and impaired video or, in a non-reference method, by only showing them the impaired video.

Finally, with statistical methods or ML techniques a model is built, which maps the objective features and the subjective labels that is later used in an online NR objective fashion for assessment of quality. An example of this approach is given in (Venkataraman et al., 2009) where the authors create a k-dimensional feature space for a specific application. In this work an example is given for 3 features of a video streaming system, bit-rate, delay and loss, which form a 3-dimensional data-space. Based on the subjective data, this space is divided into regions and the regions are assigned QoE values. When an objective measurement is later made, the datapoint falls into one of the regions and is classified with the specific QoE.

The authors of this chapter also developed methods of this type using ML algorithms. In (V. Menkovski et al., 2010) and (V. Menkovski et al., 2009) decision tree models are built from existing subjective datasets that are further used to estimate the QoE, based on the measurements on the objective features.

We have also explored online learning methods (V. Menkovski et al., 2010) for dynamic environments where the QoE accuracy of the models is short lived.

3.2 Subjective video quality methods

Subjective methods are directly estimating the multimedia quality from the targeted audience, the human viewer. They are the most accurate methodology and are therefore used as a reference for the rest of the methods (S. Winkler & Mohandas, 2008). There are different ways to carry out subjective studies. The most commonly used and standardized approach is done by rating (Recommendation ITU-R BT. 500-11, 2002). The recommended setup for a subjective study is that the viewing conditions are strictly controlled, 15 or more non-expert participants are selected and prepared. The grading scale is defined as Mean Opinion Score (MOS) and it has five values: 1 Bad, 2 Poor, 3 Fair, 4 Good, 5 Excellent. Similarly for impairments: Very annoying, Annoying, Slightly annoying, Perceptible and Imperceptible. There is also a comparisons scale for Differential MOS (DMOS) going from "Much worse" to "Much better" with an additional value in the middle "The same".

The actual rating can be done with Absolute category rating for the MOS or by a variety of double stimulus and single stimulus methods. Example of a subjective study is (De Simone et al., 2009.), which deals with impairments from H.264 compression and transmission over noisy channels. A typical evaluation of many objective methods with a subjective study was done in (Seshadrinathan et al., 2010), where the authors used DMOS.

However, these methods all rely on rating that is inheritably biased due to the variance in the internal representation of the rating scale by the subjects (Krantz, 1968)(Krantz, 1972)(Krantz et al., 1971)(Shepard, 1978)(Shepard, 1981). This bias and variance is propagated to the test output and results in the inefficiency of this type of subjective studies. This is not surprising, in fact psychophysicists have argued for a long time that the human perceptual system is more accurate at grasping 'differences' rather than giving absolute rating values (Watson, 2000).

In a later analysis (Winkler, 2009) of the properties of subjective rating, Winkler discusses the MOS variability and standard deviation in a set of subjective databases. The analysis shows that the standard deviation of MOS for the midrange is between 15-20% of the scale and decreases at the edges.

The further analysis of the subjective scales (Brooks & Hestnes, 2010) presents some interesting perspectives. A striking one is that the MOS scale is an ordinal qualitative scale and should not be used as a quantitative scale. Therefore analysis of MOS in decimal values would be invalid as well as its variability calculated to less than one fifth of the scale. In other words, numbers associated with the 5 labels do not represent actual distance between them. To avoid these issues with the scales, the authors propose a label-free scale or labels only at the end of the scale. This approach, however convenient for the analysis of the results, opens the question as to how people would map their internal representation of 'Good' and 'Bad' on this scale.

Another type of scale is the Just Noticeable Difference (JND) scale. In (Watson, 2000) Watson proposes this scale for Video Quality. The benefit of the proposed JND method is that there is no rating involved, which circumvents the drawbacks of the previous methods. JND relies on executing comparison tests on two samples consecutively, while the examined parameter intensity increases or decreases in one of the samples. As soon as the user notices the difference, the method defines this distance as 1 JND on the subjective scale. An example of this approach would be to scale the difference in quality of a video, due to different encoding bit-rate. Users in the test tend to notice the difference with a certain probability, as result of the imperfections of the HVS and the cognitive processes. Therefore the method finds the most likely JND scale of the parameter by maximizing the likelihood of all of the responses to the tests, executing the test set multiple times with different people. JND avoids the variability and bias of the rating scales, because the method relies on the capability of the participants to discriminate between two levels of quality, which comes quite easy to the viewers. When the videos are shown next to each other, most viewers will discriminate between them with high accuracy. However, if the videos are viewed independently they might find them both with the same subjectively quality. As a consequence, direct comparisons like this tend to sort the stimuli intensity rather than scale them well. Furthermore, the examiner needs to be able to very finely tune the tested parameter, in our example produce videos with given bit-rate. Therefore, this method cannot scale a directly predetermined set of parameter values.

Benefiting from the accuracy of the pair-wise comparison, usually referred to as two-alternative-forced-choice (2AFC), is also the MLDS method. MLDS stands for Maximum Likelihood Difference Scaling, and is also a method that scales the differences in quality, as in JND. However, this method presents the viewer with two pairs of stimuli, rather than one pair. The participant is then asked to discriminate between the two ranges of quality given by the two pairs. Because the discrimination is done on the ranges, the difference scale presents an actual scale of differences in quality. Maloney and Yang present the MLDS method on subjective analysis of image quality in (Maloney & Yang, 2003). Later in (Charrier et al., 2010) MLDS is used for measuring the performance of different IQA.

Because of its superior performance the authors of this chapter have also used MLDS for subjective estimation of quality in video (Menkovski, et al., 2011a). MLDS as well as the other difference scaling methods does not deliver absolute values of quality as the MOS scale does (as 'good' or 'fair'), but it only gives relative difference in the quality. This relative difference in quality, on the other hand, represents the utility of the tested parameter on the visual quality. As this is commonly the case in estimating QoE, the tester needs to optimize the service so that the delivered quality is justified by the spent resources. In these cases the benefits of MLDS in accuracy cannot be challenged by the direct results of rating.

The main drawback in the use of MLDS is that it requires much more testing than pair-wise comparisons for JND or rating tests. This drawback is minimal when the comparison is on images that require short time, but it becomes more significant on videos. To address this problem we have worked on an adaptive MLDS approach in (Menkovski et al., 2011b). In this approach the pairs for the MLDS tests are not selected randomly, but through a guided approach based on the previous responses in a iterative manner.

4. Multimedia encoding technologies and their effect on the QoE

Video encoding is an active area of research and development, with a large number of video encoding algorithms produced so far. These algorithms vary in compression efficiency and performance. However, currently the established player is based on the H.264 standard that came out of the collaboration between the ISO MPEG and the ITU-T video coding expert groups (Ostermann et al., 2004). This type of coding techniques divides the image in macro blocks. Each macro block consists of three components, i.e. the Y – luminance and the two chrominance components, Cr and Cb. The luminance component is kept at a higher spatial resolution than the chrominance components due to the fact that the HVS is more sensitive to luminance.

The macro blocks are then coded into Intra mode or Inter mode. In the Intra mode, all the blocks of the given frame are fully coded into an I-frame. The I-frame as such does not depend on any previous or subsequent frames. In the Inter mode, P-frames or B-frames are coded, where only differences from previous or subsequent frames are coded. Additionally, motion of the macro blocks is estimated with motion vectors in order to minimize the need to encode data. Main reasons for degradation of the quality and in turn the QoE are restrictions on the bit-rate, which cause the blockiness effect. The more efficient the encoder is the better quality it will produce on lower bit-rates.

In comparison to other coding algorithm (Wiegand et al., 2003), H.264 shows superior performance over the whole range of bit-rate constraints. This performance gains are paid

by increased complexity compared to the other codecs. Additionally this highly compressed bit-stream is more sensitive to errors. A subjective study by Pinson et al. (Pinson, Wolf, & Cermak, 2010) shows that H.264 suffers from higher degradation of quality than MPEG2 on the same amount of errors in the transmission channel.

4.1 Transcoding

The purpose of transcoding, or re-encoding, is to convert a previously encoded media signal (video and/or audio) into another one with different format. This is a commonly used procedure in mobile streaming due to the need to adapt the content to different types of devices. Media transcoding is most efficient when the source stream is of the highest possible quality. Ideally, the input signal is directly the original one as captured from the recorder. The parameters of the re-encoding process, such as frame size, frame rate and bit rate can also be adapted to the various available resources. In this way, the clients connected through a broadband connection can be served with the best possible quality, while clients on lower speed connections can still have access to a lower quality version of the video.

Since the encoders are lossy, each conversion adds degradation to the media quality. Compression artifacts are cumulative, so the number of transcoding steps should be kept to a minimum. In most applications the main reason for transcoding is bit-rate reduction. The complete process of re-encoding would require the use of a pixel-domain approach, in which the original signal is fully decoded, processed as necessary and then re-encoded according to the constraints of the new wanted stream. While this is always possible to do, and sometimes necessary, this process is usually very costly and difficult to be implemented efficiently in a real-time environment.

Two types of transcoding systems can be defined: open-loop and closed-loop (Assuncao & Ghanbari, 1998). Open-loop systems are the simplest ones, and require a very limited elaboration of the input stream because only the encoded DCT (Discreet Cosine Transform) coefficients are modified. The roughest technique consists of discarding all the coefficients below a certain threshold level, or above a certain frequency. The number of coefficients removed can be adjusted to obtain the required final bit-rate. However, the open-loop systems introduce a major source of distortion called drift, which lead to a visual blurring. This is due to incoherence between the pictures used for prediction by the encoder and the decoder. As a matter of fact, the new P-frames are directly derived from the old P-frames, without considering the inaccuracies introduced by the re-quantization. This way, a continuous drop of quality is accumulated over predicted frames. The closed-loop approach tries to solve this problem by approximating the complete decoder-encoder architecture. The stream is still not fully re-encoded, but a feedback loop containing a frame buffer is used in order to compensate for the transcoding distortion, so that it won't be propagated into the successive frame.

In addition to bit-rate reduction, transcoding is also used for spatial resolution downscaling. Reuse of motion parameters and MacroBlock information can help in reducing encoding complexity and consequently transcoding efficiency. A simple way to implement spatial reduction is to filter the input signal in order to retain only low-frequency coefficients, and then reconstruct new reduced MacroBlocks with the retained ones. Pixel averaging is a typical technique for spatial resolution downsampling in which an n-by-n region of pixels is

encoded with a single pixel having a color value of the average of all of the pixels. This technique results in blurriness.

In case of temporal transcoding, some of the frames that were originally present in the video are dropped. The necessary task in this kind of transcoding is the re-estimation of motion vector. Bilinear interpolation techniques can be repeated iteratively to approximate all skipped frames. Other alternatives are: Forward Dominant Vector Selection, Telescopic Vector Composition and Activity-Dominant Vector Selection (Ahmad et al., 2005).

When the video is transmitted through a channel with high error rates, like a wireless network, the transcoding process can be optimized using stronger protection mechanisms in order to compensate these losses and supply the user with a better quality of experience while maintaining the same required rate. By adding more redundancy, the total video size is augmented, so it's necessary to reduce the encoding quality in order to retain the same bit-rate. A trade-off between resiliency and effective bit-rate needs to be estimated in order for the overall delivered quality to be maximized.

4.2 Scalable video coding

Scalable video Coding (SVC) is an extension of the H.264/MPEG-4 Advanced Video Coding standard, which introduces a special layered encoding; this is similar to progressive JPEG, used for image compression and transmission over the Web. A video encoded in SVC format is created composing multiples sub-streams derived from the original video signal. Each sub-stream can be transmitted independently from the others. However, in order to reconstruct the video the client has to start from the base layer and then sequentially use the available improvement layers. This way, starting from a single encoded video stream is possible to achieve a multi bit-rate streaming service. In contrast, other currently used technologies require separate encoding of each individual stream at different bit rate. This operation needs to be done once, if the streams are stored as independent files, or each time the content is required if done on-the-fly. In the first case, for a typical streaming service, between 3 and more than 10 copies of each video are created and stored on the transmitting source, introducing a great redundancy. While the storage costs are becoming less of a factor, this approach greatly increases the initial efforts and the ongoing maintenance complexity of the system. In the second case, we remove the storage requirements, but the encoding process has significantly higher computing cost. SVC tries to reduce those costs while keeping encoding efficiency and an efficient granularity.

Using SVC brings a penalty of increased bit-rate by 10 to 20% compared to the highest bit-rate needed for a single-layer H.264 encoding. However, if we want to achieve the same flexibility with single-layer encoded streams we would need to pay much higher cost in storage. For example, using 6 streams with bit-rates ranging from 0.15 to 3 Mbps would require a total of 6.25 Mbps for the normal encoding ($0.15 + 0.3 + 0.5 + 0.8 + 1.5 + 3$ Mbps) and only 3.6 Mbps for SVC (-42%).

There are four main classes of scalability in the SVC definition (Schwarz, Marpe, & Thomas Wiegand, 2007):

- temporal: reducing the temporal resolution (number of frames per second)
- spatial: reducing the spatial resolution (number of pixels per spatial region, frame size)

- fidelity: also called Signal to Noise Ratio (SNR) or quality scalability, reducing the fidelity of the video (coarsely quantized pixels)
- a fourth class of scalability is obtained by a combination of the other three.

Temporal scalability is the result of the introduction of hierarchical B-frames. In the classical prediction structure, B-frames are derived only from preceding and subsequent I/P frames. In the new scalable model, B-frames are used to predict other B-frames of enhancing layers in a cascade way. For instance, the base layer could be composed of I and P frames, which are used to predict the B frames of the first temporal enhancement layer. In turn, B-frames of the first layer are used to predict B-frames on the second layer (together with key frames of the base layer), and so on. As a result, the total number of B-frames between two key frames (of type I or P) is $2^k - 1$. Hierarchical B-frames can be organized in different orders. A simple sequence is to iteratively put one B-frame of a higher layer between two predicting frames of the lower layer. A consequence of this kind of hierarchization of B-frames is that the first frame to be predicted and decoded is the central one, while in the classical approach it would usually be the first one of the group.

In the spatial scalability, intra and inter-layer prediction mechanisms are exploited. Each following layer corresponds to one of the supported increasing spatial resolutions. Again, the base layer is used to encode the following enhancement layers in a bottom-up fashion. An intuitive constraint is that the resolution cannot decrease in enhancing layers. An important property of SVC is that each spatial layer is decodable using a single motion compensation loop, keeping a low complexity level. Spatial scalability generally performs better when high-resolution material is used as input.

Quality scalability is based on an improvement of the concept of coarse-grain quality scalable coding, called medium-grain quality scalability. The improving signal contained in each successive enhancing layer is re-quantized using a finer quantization step compared to the previous layer. Motion compensation is done using only key frames composing the base layer. Both spatial and quality scalability are sources of loss of encoding efficiency compared to a single-layer system. Several analyses of SVC performances have been conducted. In (Van der Auwera et al., 2008), traffic characteristics of the various scalability modes proposed by SVC are compared with MPEG-4 Part 2.

Even though, SVC is still a little more costly in terms of encoding efficiency or overhead, the gap with H.264/AVC single layer coding is small enough to be considered a good alternative in most cases. Furthermore, rate-distortion comparison shows that SVC outperforms almost all other video encoding technologies currently available (Wien et al., 2007). Another positive property of SVC is its strong error resilience, which results in a graceful degradation in quality when the transmission is impaired by packet losses (Wiegand et al., 2009). Some layers are more important than others for the final perceived quality of the video displayed. Based on this assumption, a stronger protection (redundancy) should be used for these layers to allow for better degradation on quality.

4.3 Multiple description coding

The idea behind Multiple Description Coding is quite similar to Scalable Coding, where a single source is encoded in multiple streams with different quality. The more of these streams are received by the final user, the better the quality of the decoded signal is. Each

single stream is called ‘description’ and can be ideally self decoded. Since the purpose of MDC is to contrast packet losses and sudden link failures by increasing the redundancy, each description should be sent through different paths, even if this reduces routing efficiency. Using multiple paths can however have additional positive impacts in traffic dispersion and load balancing, which is useful to prevent or mitigate congestions. The benefits of MDC come obviously at the price of using more bits to achieve the same quality than using a ‘single description’ encoding (Y. Wang et al., 2005). The various descriptions are not necessarily required to have the same bit-rate, allowing a certain degree of adaptability to specific transmission channels.

Different ways exist to create MDs (multiple descriptions). The simplest one is to divide somehow the source data into several subsets and compress them separately to produce different descriptions. Interpolation is then used to decode any received combination of descriptions. A classic example is to separate odd and even samples, for example alternate frames, obtaining two subsets at half the rate of the original signal. Three decoders can then be used by the receiver: two for each separate description and one in case both the streams are properly received. Generally N descriptions would then require $2^{(N-1)}$ decoders to be decoded. A different approach is to use successive refinements to reduce the total number of decoders. In this case, one decoder is used when a single stream is received, a different one is used to decode two descriptions and so on. Another possibility is to repeat some part of data in every description created. Not all information has the same usefulness, so the most important parts are replicated more than the others. This approach is called Unequal Error Protection.

5. Multimedia streaming technologies and their effect on QoE

In multimedia streaming the content is continuously received over a transmission channel as it is reproduced to the viewer. This technique offers viewing of live content with reduced delay and offline content without the need to store the entire content locally before viewing. The cost for these benefits is that there is a possibility for additional degradation in quality coming from the transmission process. The degradation comes from delays or untimely reception of the content or from errors in the transmission.

There are different methods for video streaming. These methods implement protocols that deal with the issues of transmission of multimedia content in different manner. Today’s streaming solutions either take place over the Internet or other private or proprietary networks that implement the IP protocol stack. In these environments the streaming protocol is either implemented over a UDP or a TCP protocol. In both cases the streaming content is segmented in data packets and transmitted over the network. The UDP protocol does not provide any guarantees about the delivery nor does any transmission control. Protocols that work over UDP need to have their own transmission control and error recovery mechanisms. Protocols that utilize TCP benefit from its error correction mechanisms, however these mechanisms can add unacceptable delays for streaming multimedia.

A commonly used protocol for multimedia streaming over UDP is the Real-time Transport Protocol (RTP). The protocol works in conjunction with the RTCP (Real-time control protocol) and the RTSP (Real-time streaming protocol). The RTP is responsible for the

transport of the audio visual content. RTCP provides statistics monitors the transmission for the purpose of Quality of Service estimation and management. RTSP provides an application-level control over the playback of the content and works over a TCP channel (Schulzrinne, 1996)(Schulzrinne, 1998)(Huitema, 2003).

The underlying UDP transmission in RTP provides for simple and efficient transport of the content. The protocol presents minimal overhead in the transmission and can operate even in a unidirectional transmission channel for the RTP only implementation. This type of transmission is usually desirable when the application requires low latency or has very limited buffering requirements. The cost of this simplicity is that UDP does not offer any guarantees for delivery, which leaves the RTP susceptible to errors and data loss, if no higher level protection mechanisms are implemented.

Degradation in quality due to transmission errors and data loss in IP networks is a problematic that has been significantly examined in the literature (Kanumuri et al., 2006)(Mu et al., 2009). Transmission of RTP packets over TCP is also standardized, but is not used very commonly. The mechanisms of TCP provide guarantees for delivery of the data, however these mechanisms are implemented through time-outs and retransmissions that can add significant delay to the overall delivery. Particularly, in error-prone channels TCP transmission leads to excessive delays and connection failures. These problems result in playback freezes, skipping of content segments or interruptions. These types of impairments are severe and could possibly be alleviated by error concealment mechanisms of the player, if the control over error correction is left to the application layer. In offline video playback these problems can be reduced to some extent with large buffers. The drawback of this approach is that the playback startup is delayed in the beginning, proportionally to the amount of data in the buffer.

With the proliferation of video on the Web the progressive download method has gained significant popularity. Part of the popularity of progressive download comes from its compatibility with HTTP technologies that underpin the Web. The method is a multimedia content transmission over TCP and HTTP. The difference with traditional download is that the content is being displayed to the viewer while being downloaded. This method relies on the TCP guarantees for data delivery and on significant buffer size to deal with the delays coming from the TCP transmission. Since progressive download does not adapt and is not designed for particular transmission channel characteristics, usually a significant portion of the content needs to be downloaded before successful playback can be executed. The amount of data that needs to be buffered is crucial to the delivered QoE for this method. The degradation in QoE in progressive download comes mainly from the amount and frequency of freezes in the playback and their duration as well as the delay in the video reproduction startup. The effect of freezes on quality is evaluated in (Huynh-Thu & Ghanbari, 2010).

As we have previously stated the popularity of progressive download comes from its compatibility with Web server technology. Another streaming method that builds on this compatibility but tries to address some of the issues of progressive download is adaptive streaming or stream switching. This adaptive streaming paradigm is recently gaining a lot of interest in web based services. Its success is due to many factors. First of all, it can be simply built on top of existing and widespread open technologies like HTTP. Dedicated and optimized servers exist for adaptive streaming, however it can be completely implemented using normal web servers already deployed. One of the main advantages of using plain HTTP is that it is usually permitted even in the most restricted environments, passing

through firewalls and proxy servers. Another advantage derives from the cacheability of web contents. Since videos are seen as normal web content, intermediate caches placed between the original server and end user will allow for a significant reduction of network load, bringing contents closer to the consumer.

In progressive streaming, multiple versions of a single video may be available to users, diversified according to bit-rate, and as consequence, quality level. At the beginning of streaming session, the user is required to select the most suitable version that meets its requirements and available resources. In case of considerable variation in bandwidth, the viewer may experience frequent freezes due to buffer under-runs. Moreover, in progressive streaming usually the client tries to greedily get and cache in the buffer as much content as possible, regardless of the current position in the video playback. This attitude can waste substantial network resources, because many videos are not watched all the way to the end.

To overcome those issues, while still retaining the simplicity of progressive streaming, the HTTP Adaptive Streaming approach has been proposed, and is currently deployed by several media content providers. The idea is to encode and store multiple versions of each single video at different bitrates, and let the client choose instant by instant the best suitable version to download. To do so, every stream is fragmented in chunks of fixed length, typically in the range between 1 and 10 seconds. The reference frame at the beginning of each chunk is synchronized, so that switching between different streams does not create glitches. The client keeps a limited buffer and does not try to get more chunks than necessary to fill it. The available bandwidth can be estimated by the download time required for previous chunks. When changes in network conditions are detected, a different version of the video can be selected for the next chunk. As discussed in the previous sections, frequent changes in video quality should be avoided because they degrade the QoE.

Since HTTP is a pull-based protocol, each client is required to autonomously manage its streaming session by actively requesting each chunk. This means that servers are relieved from keeping track of the status of each client. Moreover, each client can potentially implement its own adaptive algorithm. The list of available streams and chunks must be known by the client. To do so, a manifest file is published by the server, and this is first file that a client requests and later parses before starting its streaming session.

One of the most important Content Delivery Network service available today is Akamai. The performances of its adaptive streaming implementation were analyzed in (De Cicco & Mascolo, 2010). Each video appears to be encoded in five different bit-rate versions. The client communicates with the server passing variables and commands through POST messages. The frame rate is never modified, so the adaptation mechanism relies only on the quality level of the video. The control algorithm is executed on average each two seconds. Nevertheless, when a sudden increase of the available bandwidth occurs, the time required by the transition to fully match the new bandwidth is roughly 150 seconds. This large actuation delay can also be noticed in case of drop of the bandwidth, leading to short interruptions in video playback.

A study of client-side intelligence for adaptive streaming was conducted in (Jarnikov & Özçelebi, 2011). A system using reward and penalty parameters is used in order to personalize the streaming strategy. Experiments showed that considering the past two minutes of network conditions is enough to have a sufficient feedback for the adaptation algorithm. Increasing this period leads to a more conservative strategy that tends to keep a

lower quality level. Comparisons with Apple's implementation shows that the proposed solution can deliver video at quality level at least as good as the Apple solution, and usually better exploit the available bandwidth at the cost of more changes of quality levels. It is also pointed out that the Apple implementation sometimes wastes bandwidth, requesting more than once the same temporal chunk at various quality levels.

A system that uses bit-stream switching and a customized congestion control for adaptive streaming over RTP/UDP is proposed in (Schierl & Wiegand, 2004). The pre-buffering time is fixed at 1 second, and the maximum signaling overhead due to feedback mechanisms is kept below 5% of the total data sent. Experimental results shows that bit-stream switching, combined with temporal scalability (drops of frames), provide the best performances, compared to only temporal scalability.

6. Conclusions

The QoE methods carry quite a big significance for any multimedia streaming service. They provide for efficient service management by balancing resource consumption with delivered quality. Nevertheless, estimation of the QoE is no small task. QoE is a multifaceted and complex metric that aims to quantify the experienced quality. A multitude of factors affect it, some of which may not be directly measurable in specific systems. This chapter presents a discussion on the factors that affect QoE in multimedia services, with emphasis on mobile streaming. It also presents a survey of the methods for estimating the QoE, or more particularly the video quality, from both a subjective and an objective nature. Focusing more on multimedia streaming the chapter continues with a discussion on how different technologies for encoding and streaming video content influence the delivered quality.

In conclusion of this discussion it becomes evident that while there are many QoE models, most of them are significantly constrained to specific aspects. As such they only work well for their attended purpose (Mu & Mauthe, 2009). A more general model that would capture most of the factors that attribute to the experience is still missing. In light of this we need to consider QoE as an evolving domain where the models and methods progress with the advent of new technologies. The lack of generality of the models is mainly because there is still a need for better understanding of the many aspects of QoE and how they influence each other. Significant development in the objectively measurable factors is already underway for some time, with a fruitful community for video and audio quality estimation as well as transport factors that affect quality. However, there is a need to improve our understanding of the psychological aspects of QoE as well. Just an example of the importance of the psychological factors is shown by the results a recent study (Kortum & M. Sullivan, 2010), which shows how the desirability of the content plays crucially important role on the subjective perception of quality.

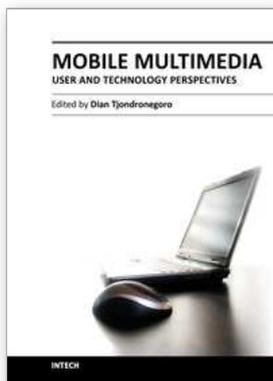
7. References

- Ahmad, I., Wei, X., Sun, Y., & Zhang, Y. Q. (2005). Video transcoding: an overview of various techniques and research issues. *Multimedia, IEEE Transactions on*, 7(5), 793–804.
- Apteker, R. T., Fisher, J. A., Kisimov, V. S., & Neishlos, H. (1995). Video acceptability and frame rate. *IEEE Multimedia*, 2(3), 32–40. doi:10.1109/93.410510

- Assuncao, P. A. A., & Ghanbari, M. (1998). A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bit streams. *Circuits and Systems for Video Technology, IEEE Transactions on*, 8(8), 953-967.
- Van der Auwera, G., David, P. T., Reisslein, M., & Karam, L. J. (2008). Traffic and quality characterization of the H. 264/AVC scalable video coding extension. *Advances in Multimedia*, 2008(2), 1-27.
- Avcıbaşı, İ., Sankur, B., & Sayood, K. (2002). Statistical evaluation of image quality measures. *Journal of Electronic Imaging*, 11, 206.
- Brooks, P., & Hestnes, B. (2010). User measures of quality of experience: why being objective and quantitative is important. *Network, IEEE*, 24(2), 8-13.
- Charrier, C., Knoblauch, K., Moorthy, A. K., Bovik, A. C., & Maloney, L. T. (2010). Comparison of image quality assessment algorithms on compressed images. *SPIE conference on Image quality and System Performance (to appear, 2010)*.
- De Cicco, L., & Mascolo, S. (2010). An experimental investigation of the Akamai adaptive video streaming. *HCI in Work and Learning, Life and Leisure*, 447-464.
- Gunawan, I. P., & Ghanbari, M. (2008). Efficient reduced-reference video quality meter. *Broadcasting, IEEE Transactions on*, 54(3), 669-679.
- Hekstra, A. P., Beerends, J. G., Ledermann, D., De Caluwe, F. E., Kohler, S., Koenen, R. H., & Rihs, S. (2002). PVQM-A perceptual video quality measure. *Signal processing: Image communication*, 17(10), 781-798.
- Huitema, C. (2003). Real time control protocol (RTCP) attribute in session description protocol (SDP).
- Huynh-Thu, Q., & Ghanbari, M. (2010). Modelling of spatio-temporal interaction for video quality assessment. *Signal Processing: Image Communication*, 25(7), 535-546.
- Jain, R. (2004). Quality of experience. *IEEE Multimedia*, 11(1), 96-95.
- Jarnikov, D., & Özçelebi, T. I. (2011). Client intelligence for adaptive streaming solutions. *Signal Processing: Image Communication*.
- Kanumuri, S., Cosman, P. C., Reibman, A. R., & Vaishampayan, V. A. (2006). Modeling packet-loss visibility in MPEG-2 video. *Multimedia, IEEE Transactions on*, 8(2), 341-355.
- Kortum, P., & Sullivan, M. (2010). The effect of content desirability on subjective video quality ratings. *Human factors: the journal of the human factors and ergonomics society*, 52(1), 105.
- Krantz, D. H. (1968). A theory of context effects based on cross-context matching* 1. *Journal of Mathematical Psychology*, 5(1), 1-48.
- Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching* 1. *Journal of Mathematical Psychology*, 9(2), 168-199.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). Foundations of measurement, vol. 1: Additive and polynomial representations. *New York: Academic*.
- Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, 3(8).
- Menkovski, V., Exarchakos, G., & Liotta, A. (2010). Online QoE prediction. *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on* (pp. 118-123). Presented at the Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on. doi:10.1109/QOMEX.2010.5517692
- Menkovski, V., Exarchakos, G., Liotta, A., & Sanchez, A. C. (2010). Estimations and Remedies for Quality of Experience in Multimedia Streaming. *Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC), 2010*

- Third International Conference on* (pp. 11-15). Presented at the Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC), 2010 Third International Conference on. doi:10.1109/CENTRIC.2010.14
- Menkovski, V., Exarchakos, G., Liotta, A., & Sánchez, A. C. (2010). Measuring Quality of Experience on a Commercial Mobile TV Platform. *Advances in Multimedia (MMEDIA), 2010 Second International Conferences on* (pp. 33-38). Presented at the Advances in Multimedia (MMEDIA), 2010 Second International Conferences on. doi:10.1109/MMEDIA.2010.12
- Menkovski, V., Oredope, A., Liotta, A., & Sánchez, A. C. (2009). Predicting quality of experience in multimedia streaming. *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia* (pp. 52-59).
- Menkovski, Vlado, Exarchakos, Georgios, & Liotta, Antonio. (2011a). The value of relative quality in video delivery. *Journal of Mobile Multimedia*, 7(3).
- Menkovski, Vlado, Exarchakos, Georgios, & Liotta, Antonio. (2011b). Adaptive testing for video quality assessment. Presented at the Quality of Experience of multimedia content sharing, Lisbon, Portugal.
- Mu, M., & Mauthe, A. (2009). An interview with video quality experts. *ACM SIGMM Records Issue 4, December 2009*.
- Mu, M., Gostner, R., Mauthe, A., Tyson, G., & Garcia, F. (2009). Visibility of Individual Packet Loss on H. 264 Encoded Video Stream-A User Study on the Impact of Packet Loss on Perceived Video Quality. *Annual Multimedia Computing and Networking, San Jose, USA*.
- Mued, L., Lines, B., Furnell, S., & Reynolds, P. (2003). The effects of audio and video correlation and lip synchronization. *Campus-Wide Information Systems*, 20(4), 159-166. doi:10.1108/10650740310491333
- Ostermann, J., Bormans, J., List, P., Marpe, D., Narroschke, M., Pereira, F., Stockhammer, T., et al. (2004). Video coding with H.264/AVC: tools, performance, and complexity. *IEEE Circuits and Systems Magazine*, 4(1), 7- 28. doi:10.1109/MCAS.2004.1286980
- Pinson, M. H., & Wolf, S. (2004). A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3), 312- 322. doi:10.1109/TBC.2004.834028
- Pinson, M. H., Wolf, S., & Cermak, G. (2010). HDTV subjective quality of H. 264 vs. MPEG-2, with and without packet loss. *Broadcasting, IEEE Transactions on*, 56(1), 86-91.
- Recommendation, ITU-R BT. 500-11 (2002). "Methodology for the Subjective Assessment of the Quality of Television Pictures," Recommendation ITU-R BT. 500-11. *ITU Telecom. Standardization Sector of ITU*.
- Reibman, A. R., & Vaishampayan, V. (2003). Quality monitoring for compressed video subjected to packet loss. *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on* (Vol. 1, p. I-17).
- Schierl, T., & Wiegand, T. (2004). H. 264/avc rate adaptation for internet streaming. *Packet Video Workshop*.
- Schulzrinne, H. (1996). RTP: A transport protocol for real-time applications.
- Schulzrinne, H. (1998). Real time streaming protocol (RTSP).
- Seshadrinathan, K., & Bovik, A. C. (2009). Motion-based perceptual quality assessment of video. *Proc. SPIE-Human Vision and Electronic Imaging*.
- Seshadrinathan, K., Soundararajan, R., Bovik, A. C., & Cormack, L. K. (2010). Study of subjective and objective quality assessment of video. *Image Processing, IEEE Transactions on*, 19(6), 1427-1441.

- Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of. *Journal of Mathematical Psychology*, 24(1), 21–57.
- Shepard, R. N. (1978). On the status of direct psychophysical measurement. *Minnesota studies in the philosophy of science*, 9, 441–490.
- De Simone, F., Naccari, M., Tagliasacchi, M., Dufaux, F., Tubaro, S., & Ebrahimi, T. (n.d.). Subjective assessment of H. 264/AVC video sequences transmitted over a noisy channel. *Quality of Multimedia Experience, 2009. QoMEX 2009. International Workshop on* (pp. 204–209).
- Steinmetz, R. (1996). Human perception of jitter and media synchronization. *Selected Areas in Communications, IEEE Journal on*, 14(1), 61–72.
- Tran, H. A., & Mellouk, A. (2010). QoE Model Driven for Network Services. In E. Osipov, A. Kasser, T. M. Bohnert, & X. Masip-Bruin (Eds.), *Wired/Wireless Internet Communications* (Vol. 6074, pp. 264–277). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from <http://www.springerlink.com/content/7159q510q378m234/>
- Venkataraman, M., Chatterjee, M., & Chattopadhyay, S. (2009). Evaluating quality of experience for streaming video in real time. *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE* (pp. 1–6).
- Wang, Y., Reibman, A. R., & Lin, S. (2005). Multiple description coding for video delivery. *Proceedings of the IEEE*, 93(1), 57–70.
- Wang, Z., Lu, L., & Bovik, A. C. (2004). Video quality assessment based on structural distortion measurement. *Signal processing: Image communication*, 19(2), 121–132.
- Watson, A. B. (2000). Proposal: Measurement of a JND scale for video quality. *IEEE G-2.1. 6 Subcommittee on Video Compression Measurements*.
- Watson, Andrew B., Hu, J., & McGowan, J. F. (2001). Digital video quality metric based on human vision. *Journal of Electronic Imaging*, 10(1), 20. doi:10.1117/1.1329896
- Wiegand, T., Noblet, L., & Rovati, F. (2009). Scalable video coding for IPTV services. *Broadcasting, IEEE Transactions on*, 55(2), 527–538.
- Wiegand, T., Schwarz, H., Joch, A., Kossentini, F., & Sullivan, G. J. (2003). Rate-constrained coder control and comparison of video coding standards. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7), 688–703.
- Wien, M., Schwarz, H., & Oelbaum, T. (2007). Performance analysis of SVC. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(9), 1194–1203.
- Winkler, S., & Mohandas, P. (2008). The evolution of video quality measurement: from PSNR to hybrid metrics. *Broadcasting, IEEE Transactions on*, 54(3), 660–668.
- Winkler, Stefan. (2009). On the properties of subjective ratings in video quality experiments. IN *PROC. INTERNATIONAL WORKSHOP ON QUALITY OF MULTIMEDIA EXPERIENCE (QOMEX)*. Retrieved from <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.160.3958>
- Wu, D., Hou, Y. T., Zhu, W., Zhang, Y. Q., & Peha, J. M. (2001). Streaming video over the Internet: approaches and directions. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(3), 282–300.
- Zink, M., Künzel, O., Schmitt, J., & Steinmetz, Ralf. (2003). Subjective Impression of Variations in Layer Encoded Videos. IN: *INTERNATIONAL WORKSHOP ON QUALITY OF SERVICE*, 137--154.



Mobile Multimedia - User and Technology Perspectives

Edited by Dr. Dian Tjondronegoro

ISBN 978-953-307-908-0

Hard cover, 156 pages

Publisher InTech

Published online 20, January, 2012

Published in print edition January, 2012

As multimedia-enabled mobile devices such as smart phones and tablets are becoming the day-to-day computing device of choice for users of all ages, everyone expects that all mobile multimedia applications and services should be as smooth and as high-quality as the desktop experience. The grand challenge in delivering multimedia to mobile devices using the Internet is to ensure the quality of experience that meets the users' expectations, within reasonable costs, while supporting heterogeneous platforms and wireless network conditions. This book aims to provide a holistic overview of the current and future technologies used for delivering high-quality mobile multimedia applications, while focusing on user experience as the key requirement. The book opens with a section dealing with the challenges in mobile video delivery as one of the most bandwidth-intensive media that requires smooth streaming and a user-centric strategy to ensure quality of experience. The second section addresses this challenge by introducing some important concepts for future mobile multimedia coding and the network technologies to deliver quality services. The last section combines the user and technology perspectives by demonstrating how user experience can be measured using case studies on urban community interfaces and Internet telephones.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Vlado Menkovski and Antonio Liotta (2012). QoE for Mobile Streaming, Mobile Multimedia - User and Technology Perspectives, Dr. Dian Tjondronegoro (Ed.), ISBN: 978-953-307-908-0, InTech, Available from: <http://www.intechopen.com/books/mobile-multimedia-user-and-technology-perspectives/qoe-for-mobile-streaming>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.