Signal Processing: *Image Communication* 39 (2015) 432–443

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Quality of Experience of adaptive video streaming: Investigation in service parameters and subjective quality assessment methodology

Samira Tavakoli [a,*], Kjell Brunnström [b,c], Jesús Gutiérrez [a], Narciso García [a]

[a] *Universidad Politécnica de Madrid, Spain*
[b] *Acreo Swedish ICT AB, Sweden*
[c] *Mid Sweden University, Sweden*

ARTICLE INFO

Available online 15 May 2015

Keywords:
Adaptive video streaming
Quality of Experience
Time-varying video quality
Subjective evaluation

ABSTRACT

The usage of HTTP adaptive streaming (HAS) has become widely spread in multimedia services. Because it allows the service providers to improve the network resource utilization and user's Quality of Experience (QoE). Using this technology, the video playback interruption is reduced since the network and server status in addition to capability of user device, all are taken into account by HAS client to adapt the quality to the current condition. Adaptation can be done using different strategies. In order to provide optimal QoE, the perceptual impact of adaptation strategies from point of view of the user should be studied. However, the time-varying video quality due to the adaptation which usually takes place in a long interval introduces a new type of impairment making the subjective evaluation of adaptive streaming system challenging. The contribution of this paper is two-fold: first, it investigates the testing methodology to evaluate HAS QoE by comparing the subjective experimental outcomes obtained from ACR standardized method and a semi-continuous method developed to evaluate the long sequences. In addition, influence of using audiovisual stimuli to evaluate the video-related impairment is inquired. Second, impact of some of the adaptation technical factors including the quality switching amplitude and chunk size in combination with high range of commercial content type is investigated. The results of this study provide a good insight toward achieving appropriate testing method to evaluate HAS QoE, in addition to designing switching strategies with optimal visual quality.

1. Introduction

In recent years a great increase of video services has taken place which provided the ease of access for the consumer from almost any type of device and different location. This has made video streaming as the most dominant application in the Internet and this dominance is expected to grow even further within the near future. One of the main advances in this aspect is using HTTP streaming over TCP as delivery method which is used by most of the video content delivery services such as YouTube and Netflix. In contrast to the traditional streaming over UDP, where packet losses result in audiovisual distortions, the packet retransmission feature of TCP ensures reliable delivery of the video content. Buffering of the content at the client side further allows to overcome the network resource limitations in a short time scale and assures a continuous playout of the video content. However, this delivery method

* Corresponding author.

http://dx.doi.org/10.1016/j.image.2015.05.001
0923-5965/

is prone to the temporal impairments such as long initial delay (in case where large playout buffers have to be filled initially) or stalling (i.e. playback interruption due to the empty playout buffer).

To overcome these problems, several service providers have implemented HTTP adaptive streaming (HAS). HAS makes it possible to switch the video quality during the playback in order to adapt to the current network conditions. In HAS, the video content is available in multiple bitrate (called adaptive streams or *representations*) on the server-side which may differ in terms of spatial resolution, temporal resolution (framerate), encoding quantization, as well as combination thereof. Each representation consists of independently decodable segments of a few seconds of video, termed *chunk*. The characteristics of the representations are stored at the server-side in a file (or *manifest*) providing the client the required information for the adaptation process. After filling the buffer with the initial bitrate, the video playout is started. During the playback, the adaptation algorithm in the client-side measures the current bandwidth and/or buffer status in order to decide on the appropriate bitrate for the next video chunk request such that the available bandwidth is utilized best and stalling is avoided. Nevertheless, in terms of user perceived quality, another dimension, i.e. time-varying quality switching is introduced.

Apart from the benefit of dynamically adapting the current video bitrate to the available bandwidth, employing HAS provides further advantages compared to the classical video streaming. For instance, offering multiple bitrates of the video enables service providers to adapt the delivered video to the users with different demands and network/device accessibility. Furthermore, based on the available video quality, different pricing schemes and service levels can be offered to the customers. All these advantages have made employing this technology more popular. Nevertheless, in regard to the adaptation behavior that should be decided in the client side, there is still no clear guideline about the performance of different scenarios in terms of the user's Quality of Experience (QoE).

Up to now, several research works in the HAS area have been conducted which can be differentiated along technical and perceptual based quality assessment. Technical analysis such as [1] mainly focuses on inquiring switching strategies to optimize the bandwidth utilization and other network related parameters. Whilst the perceptual based analysis concerns about the QoE impact of adaptation related parameters. Up to now, research has been mostly concentrated on technical aspects of HAS, but in order to optimize the user's QoE, it is crucial to deliberate more on the perceptual aspect.

There are different potential QoE influence factors related to client switching behavior which have been already addressed in the previous studies. Two of the most common ones are the *switching frequency* (i.e. number of switches per adaptation event[1]) and *switching amplitude* (i.e. difference between the current and target quality levels in an adaptation event).

About the switching frequency, results presented in [2–6] show that frequency of the adaptation should be kept as low as possible. In the same time, the study presented in [3] indicates that if the duration spent on the high quality level is sufficiently long, higher switching frequencies do not significantly degrade the QoE. Considering the switching amplitude, most of the previous studies [3–7] conclude that gradual multiple variations are preferred over rapid variations. Nevertheless, as highlighted in [4], this conclusion may not be applied to the scenarios where the quality levels' difference is very small.

Another possible QoE influence factor could be the *chunk length* considered for switching the quality between different representations. There are some concerns in regard to selection of chunk length which should be considered in different applications. For instance, in certain scenarios like a live broadcast, using long chunk size may not be suitable as switching granularity is more considerable. Employing small chunks improves the client reaction time to network bandwidth variations but also increases the activity on the client side. To the best of our knowledge, studying the QoE impact of chunk size has not received much attention yet.

Characteristics of the content is another possible influence factor on the user's perception of quality switches. Studied presented in [3,8] revealed that the effect of spatial and temporal switching varies depending on the content type. This was found to be even true for switches with the same amplitude so that it is difficult to spot quality oscillation when there are frequent scene changes while in steady shots they are more noticeable.

One of the common approaches to evaluate the impact of visual distortions on the user's perceived QoE is through subjective assessments. Extensive research related to subjective studies of audiovisual quality has brought up several testing methodologies to obtain reliable results for development of multimedia technologies. There are different international recommendations provided by standardization organizations such as ITU-R BT.500 [9] and ITU-T P.910 [10] which give guidelines to assess the quality of television pictures. However, the novelties of adaptive streaming technology and corresponding visual degradations require research for a new assessment methodology that allows obtaining reliable conclusions regarding the HAS users' visual experience.

The most common methodologies, like Absolute Category Rating (ACR) [10], recommend the use of short test video sequences of around 15 sec after which the observers provide their ratings. However, in adaptive streaming there are switching behaviors whose effect takes longer time. Therefore, longer test sequences may be more appropriate to study these cases. Also, as presented in [11], traditional testing methods may not accurately predict the perceptual quality, because the relative impact of impairment types would change with the setting of subjective test. This means, it is not clear if the perceptual quality of adaptation event solely evaluated using the ACR method would be the same as actual scenario when the adaptaion is occurred during a longer sequence. Another standardized

[1] Adaptation event denotes the period of video playback when the quality switchings from the current level to the target level occur.