University of Würzburg
Institute of Computer Science
Research Report Series

A Survey on Quality of Experience of HTTP Adaptive Streaming

Michael Seufert[1], Sebastian Egger[2], Martin Slanina[3], Thomas Zinner[1], Tobias Hoßfeld[1], and Phuoc Tran-Gia[1]

Report No. 490 April 2014

[1] Institute of Computer Science, University of Würzburg
Am Hubland, D-97074 Würzburg, Germany
seufert@informatik.uni-wuerzburg.de

[2] FTW Telecommunication Research Center Vienna
Donau-City-Straße 1, A-1220 Vienna, Austria
egger@ftw.at

[3] Department of Radio Electronics, Brno University of Technology
Technicka 12, CZ-616 00 Brno, Czech Republic
slaninam@feec.vutbr.cz

NOTICE: An extended version of this technical report including more insights into network-related aspects of HAS was published in "IEEE Communications Surveys & Tutorials" (DOI: 10.1109/COMST.2014.2360940).
Please do not cite this technical report, but cite the peer reviewed and published paper as follows: *Michael Seufert, Sebastian Egger, Martin Slanina, Thomas Zinner, Tobias Hofeld, Phuoc Tran-Gia. "A Survey on Quality of Experience of HTTP Adaptive Streaming". IEEE Communications Surveys & Tutorials, vol. 17, no. 1, 2015.*

NOTICE: An extended version of this technical report was published in “IEEE Communications Surveys & Tutorials”. Please do not cite this technical report, but cite the peer reviewed and published paper as follows: *Michael Seufert, Sebastian Egger, Martin Slanina, Thomas Zinner, Tobias Hofeld, Phuoc Tran-Gia. “A Survey on Quality of Experience of HTTP Adaptive Streaming”. IEEE Communications Surveys & Tutorials, vol. 17, no. 1, 2015.*

A Survey on Quality of Experience of HTTP Adaptive Streaming

M. Seufert, T. Zinner, T. Hoßfeld, P. Tran-Gia
Institute of Computer Science, University of Würzburg
Am Hubland, D-97074 Würzburg, Germany
seufert@informatik.uni-wuerzburg.de

S. Egger
FTW Telecommunication Research Center Vienna
Donau-City-Straße 1, A-1220 Vienna, Austria
egger@ftw.at

M. Slanina
Department of Radio Electronics, Brno University of Technology
Technicka 12, CZ-616 00 Brno, Czech Republic
slaninam@feec.vutbr.cz

Abstract

Changing network conditions pose severe problems to video streaming in the Internet. HTTP adaptive streaming (HAS) is a technology employed by numerous video services which relieves these issues by adapting the video to the current network conditions. It enables service providers to improve resource utilization and Quality of Experience (QoE) by incorporating information from different layers in order to deliver and adapt a video in its best possible quality. Thereby, it allows to take into account end user device capabilities, available video quality levels, current network conditions, and current server load. For end users, the major benefits of HAS compared to classical HTTP video streaming are reduced interruptions of the video playback and higher bandwidth utilization, which both generally result in a higher QoE. Adaptation is possible by changing the frame rate, resolution, or quantization of the video, which can be done with various adaptation strategies and related client- and server-side actions. The technical development of HAS, existing open standardized solutions, but also proprietary solutions are reviewed in this article as fundament to derive the QoE influence factors which emerge as a result of adaptation. The main contribution is a comprehensive survey of QoE related works from human computer interaction and networking domains which are structured according to the QoE impact of video adaptation. To be more precise, subjective studies which cover QoE aspects of adaptation dimensions and strategies are revisited. As a result, QoE influence factors of HAS and corresponding QoE models are identified, but also open issues and conflicting results are discussed. Furthermore, the interaction between HAS and other clients, which is often ignored in the context of HAS, affects QoE influence factors and is consequently analyzed. This survey gives the reader an overview of the current state of the art and recent developments. At the same time it targets networking researchers who develop new solutions for HTTP video streaming or assess video streaming from a user centric point of view. Therefore, the article is a major step towards truly improving HAS.

Keywords: HTTP adaptive streaming, variable video quality, dynamic streaming, Quality of Experience

Contents

1 Introduction **3**

2 Quality of Experience Influence Factors of HTTP Video Streaming **5**
2.1 Initial Delay 5
2.2 Stalling 6
2.3 Adaptation 7

3 Current HTTP Adaptive Streaming Solutions **9**
3.1 Development and Milestones 9
3.2 The Technology Behind 9
3.3 Server-side Actions 13
3.4 Client-side Actions 15
3.5 Performance Studies 16

4 QoE Influence of Adaptation Strategy and Application Parameters **18**

5 Video Adaptation Dimensions **21**

6 QoE Influence of Video Adaptation Dimensions **23**
6.1 Image Quality Adaptation 24
6.2 Spatial Adaptation 25
6.3 Temporal Adaptation 25
6.4 Trade-offs Between Different Adaptation Dimensions 26

7 User Experience Related Impairments Beyond Video QoE **28**
7.1 Interactions between HAS player instances 28
7.2 Interactions between TCP and HAS clients 30

8 Conclusion **31**

1 Introduction

Nowadays, video is the most dominant application in the Internet. According to a recent study and forecast [1], global Internet video traffic accounted for 15 PB per month in 2012, which is 57% of all consumer traffic. By 2017, it is expected to reach 52 PB per month, which will then be 69 % of the entire consumer Internet traffic. Two thirds of all that traffic will then be delivered by content delivery networks like YouTube, which is already today one of the most popular Internet applications.

For a long period of time, YouTube has been employing a server-based streaming, but recently it introduced HTTP adaptive streaming (HAS) [2] as its default delivery/playout method. HAS requires the video to be available in multiple bit rates, i.e., in different quality levels/representations, and split into small segments each containing a few seconds of playtime. The client measures the current bandwidth and/or buffer status and requests the next part of the video in an appropriate bit rate such that stalling (i.e., the interruption of playback due to empty playout buffers) is avoided and the available bandwidth is best possibly utilized.

This trend can not only be observed with YouTube, which is a prominent example, but nowadays an increasing number of video applications employ HAS, as it has several more benefits compared to classical streaming. First, offering multiple bit rates of video enables video service providers to adapt the delivered video to the users' demands. As an example, a high bit rate video, which is desired by home users typically enjoying high speed Internet access and big display screens, is not suitable for mobile users with a small display device and slower data access. Second, different service levels and/or pricing schemes can be offered to customers. For example, the customers could select themselves which bit rate level, i.e., which quality level they want to consume. Moreover, adaptive streaming allows for flexible service models, such that a user can increase or decrease the video quality during playback if desired, and can be charged in the end of a viewing session exactly taking into account the consumed service levels [3]. Finally and most important, the current video bit rate, and hence the demanded delivery bandwidth, can be adapted dynamically to changing network and server/CDN conditions. If the video is available in only one bit rate and the conditions change, either the bit rate is smaller than the available bandwidth which leads to a smooth playback but spares resources which could be utilized for a better video quality, or the video bit rate is higher than the available bandwidth which leads to delays and eventually stalling, which degrades the Quality of Experience (QoE) severely (e.g., [4, 5]). Thus, adaptive streaming might improve the QoE of video streaming.

HAS is an evolving technology which is also of interest for the research community. The open questions are manifold and cover both the planning phase:

- How (i.e., with which parameters) to convert a source video to given target bit rates?
- Which dimension (image quality, spatial, temporal) to adapt?

and the operational phase:

- When (i.e., under which circumstances) to adapt?
- Which quality representation to request?

Moreover, the performance of existing implementations or proposed algorithms has to be evaluated and improvements have to be identified. This is especially important as today's mechanisms do not take into account the resulting QoE of end users. However, QoE is the most important performance metric as video services are expected to maximize the satisfaction of their users.

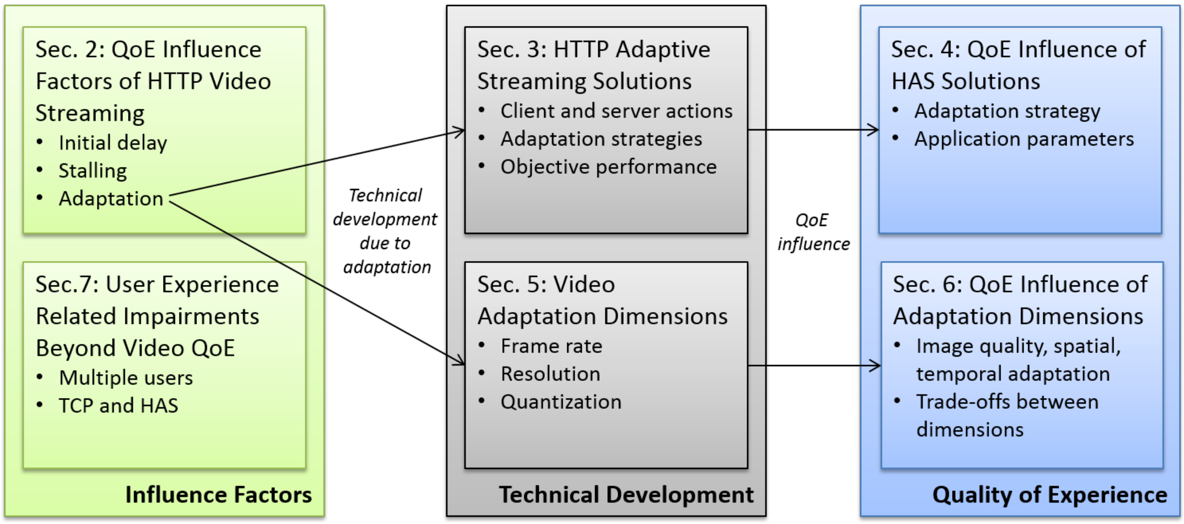

Figure 1: Structure of the article. Starting from the Quality of Experience of HTTP video streaming, technical possibilities of adaptation and the resulting influence on Quality of Experience are in the focus of this work.

Several subjective studies have been conducted in different disciplines and communities in order to investigate QoE aspects of adaptation. This work surveys these studies, identifies influence factors and QoE models, and dicusses challenges towards new mechanisms. Therefore, this article follows the structure as depicted in Figure 1: In Section 2 the main influence factors of QoE of HTTP video streaming, i.e., initial delay and stalling, are outlined. As changing video quality levels (i.e., adaptation) introduces new impacts on QoE, the remainder of the work will focus on adaptation. Section 3 presents the approach of HTTP adaptive streaming and describes the current state of the art. Section 4 describes the influences on QoE which arise from the employed adaptation strategy and parameters. The dimensions of adaptation which can be utilized for HAS are outlined in Section 5. As end users perceive the quality adaptation when using a HAS service, Section 6 surveys the influence of each dimension on QoE and describes possible trade-offs. Finally, Section 7 presents and discusses user experience related impairments beyond video QoE, and Section 8 concludes.

2 Quality of Experience Influence Factors of HTTP Video Streaming

HTTP video streaming (video on demand streaming) is a combination of download and concurrent playback. It transmits the video data to the client via HTTP where it is stored in an application buffer. After a sufficient amount of data has been downloaded (i.e., the video file download needs not to be complete yet), the client can start to play out the video from the buffer. As the video is transmitted over TCP, the client receives an undisturbed copy of the video file. However, there are a number of real world scenarios in which the properties (most importantly instantaneous throughput and latency) of a communication link serving a certain multimedia service are fluctuating. Such changes can typically appear when communicating through a best effort network (e.g., Internet) where the networking infrastructure is not under control of an operator from end to end, and thus its performance cannot be guaranteed. Another example is reception of multimedia content through a mobile channel, where the channel conditions are changing over time, due to fading, interferences, and noise. These network issues (e.g., packet loss, insufficient bandwidth, delay, and jitter) will decrease the throughput and introduce delays at the application layer. As a consequence, the playout buffer fills more slowly or even depletes. If the buffer is empty, the playback of the video has to be interrupted until enough data for playback continuation has been received. These interruptions are referred to as stalling or rebuffering.

In telecommunication networks, the Quality of Service (QoS) is expressed objectively by network parameters like packet loss, delay, or jitter. However, a good QoS does not guarantee that all customers experience the service to be good. Thus, Quality of Experience (QoE) – a concept of subjectively perceived quality – was introduced [6]. It takes into account how customers perceive the overall value of a service, and thus, relies on subjective criteria. For HTTP video streaming, [4, 7] showed in their results that initial delay and stalling are the key influence factors of QoE. However, changing the transmitted video quality as employed by HTTP adaptive streaming introduces a new perceptual dimension. Therefore, in this paper we will present detailed results on the influence of adaptation on subjectively perceived video quality.

2.1 Initial Delay

Initial delay is always present in a multimedia streaming service as a certain amount of data must be transferred before decoding and playback can begin. The practical value of the minimal achievable initial delay thus depends on the available transmission data rate and the encoder settings. Usually, the video playback is delayed more than technically necessary in order to fill the playout buffer with a bigger amount of video playtime in the receiver at first. The playout buffer is an efficient tool used to tackle short term throughput variations. However, the amount of initially buffered playtime needs to be traded off between the actual length of the corresponding delay (more buffered playtime = longer initial delay) and the risk of buffer depletion, i.e., stalling (more buffered playtime = higher robustness to short term throughput variations).

[8] show that the impact of initial delays strongly depends on the concrete application. Thus, results obtained for other services (e.g., web page load times [9], IPTV channel zapping time [10], UMTS connection setup time [11]) cannot easily be transfered to video streaming. However, those works presume a logarithmic relationship between waiting times and MOS (mean opinion score), which is a measure of subjectively perceived quality (QoE). [8] found fundamental differences between initial delays and stalling. They observed that initial delays are preferred to stalling by around 90% of users. The impact of initial delay on perceived quality is small and only depending on its length but not on video clip duration. In contrast to expected initial delays, which is waiting before the service and is well known from everyday usage of video applications, stalling invokes a sudden unexpected interruption within the service. Hence, stalling is processed differently by the human sensory system, i.e., it is perceived much worse [12].

Thus, for initial delays, the impact is not severe on QoE (in contrast to stalling). However, this is only valid if the user really intends to watch the video. For many user-generated contents, the users are browsing through videos in order to search for some contents they are interested in [13, 14]. In that case, initial delays should not be so high to be accepted by the user, however, the QoE related to video browsing is currently not investigated in research yet.

2.2 Stalling

Stalling is the stopping of video playback because of playout buffer underrun. If the throughput of the video streaming application is lower than the video bit rate, the playout buffer will empty. Eventually, insufficient data is available in the buffer and the playback of the video cannot continue. The playback is interrupted until the buffer contains a certain amount of video data. Here again, the amount of rebuffered playtime has to be traded off between the length of the interruption (more buffered playtime = longer stalling duration) and the risk of a shortly recurring stalling event (more buffered playtime = longer playback until potential next stalling event).

In [15], the authors showed that an increased duration of stalling decreases the quality. They also found that one long stalling event is preferred to frequent short ones. However, the position of stalling is not important. This last finding was refuted in [16] which showed that there is an impact of the position. In [17] the authors investigated both stalling and frame rate reduction. They showed that stalling is worse than frame rate reduction. Furthermore, they showed that stalling at irregular intervals is worse than periodic stalling. In [18], stalling is compared to quantization. The authors present a random neural network model to estimate QoE based on both parameters. In subjective studies they found that users are more sensitive to stalling than to an increase of quantization parameter in the video encoder, especially for lower values of the quantization parameter. The authors of [4] presented a model for mapping regular stalling patterns to MOS. They showed that there is an exponential relationship between stalling parameters and MOS. Moreover, they found that users tolerate one stalling event per clip as long as its duration remains in the order of a few seconds.

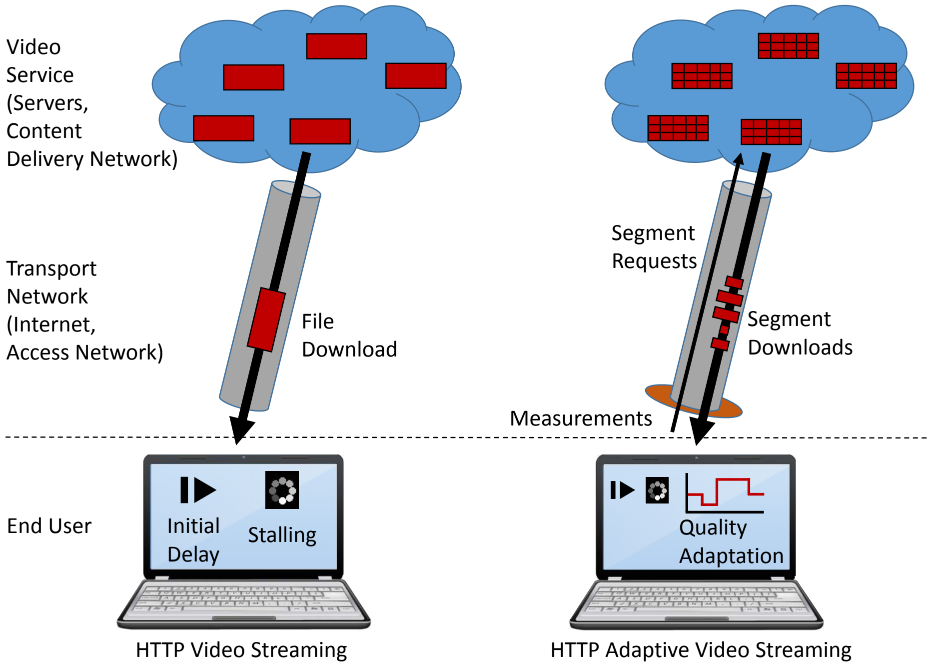

Figure 2: Comparison of HTTP video streaming and HTTP adaptive video streaming. End user is not aware of service or network influences, but only perceives initial delays, stallings, and quality adaptations.

2.3 Adaptation

HTTP adaptive streaming is based on classical HTTP video streaming but makes it possible to switch the video quality during the playback in order to adapt to the current network conditions. In Figure 2 both methods are juxtaposed. To make adaptation possible, the streaming paradigm had to be changed such that the client, who can measure his current network conditions at the edge of the network, controls which data rate is suitable for the current conditions. On the server side, the video is split into small segments and each of them is available in different quality levels/representations (which represent different bit rate levels). Based on network measurements, the adaptation algorithm at the client-side requests the next part of the video in the appropriate bit rate level which is best suited under current network conditions.

[19] compared adaptive and non-adaptive steaming under vehicular mobility and revealed that quality adaptation could effectively reduce stalling by 80% when bandwidth decreased, and was responsible for a better utilization of the available bandwidth when bandwidth increased. Also in non-mobile environments, HAS is useful because it avoids stalling to the greatest possible extent by switching the quality. The provisioning-delivery hysteresis presented in [20] confirms that it is better to control the quality, than to suffer from uncontrolled effects like stalling. The authors compared video streaming impairments due to packet loss to impairments due to resolution reduction. They

mapped objective results to subjectively perceived quality and found that the impact of the uncontrolled degradation (i.e., packet loss) on QoE is much more severe than the impact of a controlled bandwidth reduction due to resolution.

Thus, HAS is an improvement over classical HTTP video streaming as it aims to minimize uncontrolled impairments. However, compared to classical HTTP video streaming, another dimension, i.e., the quality adaptation, was introduced (see Figure 2). Therefore, we will present HAS in detail and review subjective studies on the influences of application layer adaptation on QoE of end users in the following sections. Adaptations on other layers (e.g., network traffic management, modification of content, CDN structure) will not be considered because end users eventually perceive only resulting initial delays, stallings, or quality adaptations when using a HAS service.

3 Current HTTP Adaptive Streaming Solutions

3.1 Development and Milestones

After the first launch of an HTTP adaptive streaming solution by Move Networks in 2006 [21, 22, 23], HTTP adaptive streaming was commercially rolled out by three dominant companies in parallel – as *Microsoft Silverlight Smooth Streaming* (MSS) [24] by Microsoft Corporation (2008), *HTTP Live Streaming* (HLS) [25] by Apple Inc. (2009) and *Adobe HTTP Dynamic Streaming* (HDS) [26] by Adobe Systems Inc. (2010). Despite their wide adoption and commercial success, these solutions are mutually incompatible, although they share similar technological background (see Section 3.2).

The first standardized approach to adaptive HTTP streaming was published by 3GPP in TS 26.234 Release 9 [27] in 2009 with the intended use in UMTS LTE (Universal Mobile Telecommunications System - Long Term Evolution) mobile communication networks. In the context of [27], the description of the adaptive streaming technique is quite general – the fundamental streaming principle is provided and only a brief description of the media format is given. The work of 3GPP continued by improving the adaptive streaming solution in close collaboration with MPEG [28] and, finally, the DASH (Dynamic Adaptive Streaming over HTTP) standard for general use of HAS was issued by MPEG in 2012 [29]. To date, the DASH specifications are contained in four parts, defining the media presentation description and segment formats, conformance and reference software, implementation guidelines, and segment encryption and authentication, respectively.

Apart from the standardization itself, it is also worth mentioning that in connection with DASH, an industry forum has been formed in order to enable smooth implementation of DASH in different services. Currently, the industry forum groups over 60 members, among which important players in the multimedia and networking market can be found [30]. One of the most important outputs of the industry forum is DASH-AVC/264 – a recommendation of profiles and settings serving as guidelines for implementing DASH with H.264/AVC video [31].

3.2 The Technology Behind

It has been mentioned in Section 3.1 that adaptive HTTP streaming solutions, provided as standardized or proprietary technologies by different companies, share a similar technological background. If we miss out the media and metadata formats, a common adaptive HTTP streaming solution architecture can look like the one shown in Figure 3 (inspired by [32]).

At first, the client makes a HTTP request to the server in order to obtain metadata of the different audio and video representations available by requesting the index file. The purpose of this index file is to provide a list of representations available to the client (e.g., available encoding bit rates, video frame rates, video resolutions, etc.) and a means to formulate HTTP requests for a chosen representation. The most important concept in adaptive HTTP streaming is that switching among different representations can occur at fixed, frequent time instants during the playback. To achieve this, the media

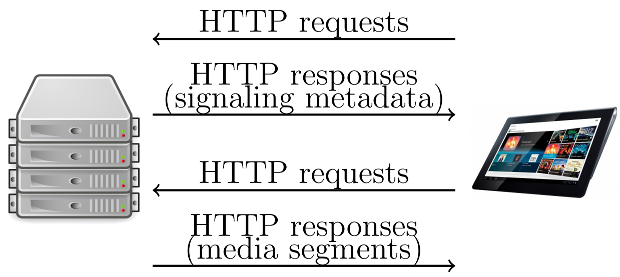

Figure 3: Principle of adaptive HTTP streaming.

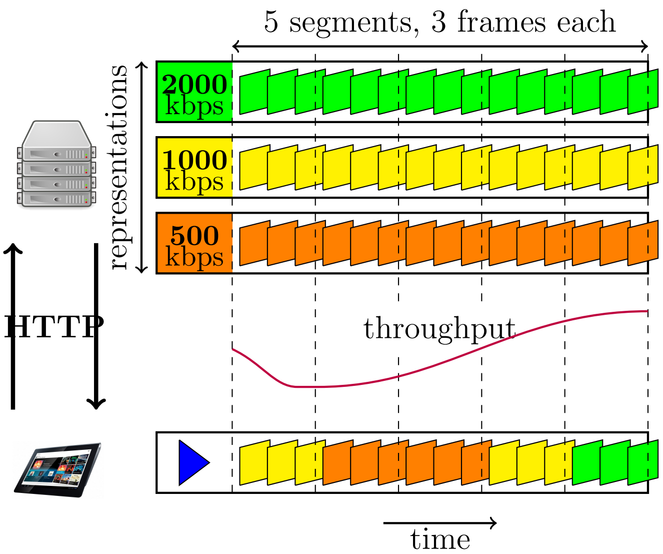

Figure 4: Video representations available for adaptive HTTP streaming.

corresponding to the respective representations is split up into parts of short durations (segments or chunks) typically 1 – 15 seconds long and either stored on the server as one file per segment (e.g., HLS) or extracted from a single file at runtime based on the client's request (e.g., DASH). To allow for smooth switching among different representations, the segments corresponding to different representations must be perfectly time (frame) aligned. This principle is illustrated in Figure 4.

The application control loop is shown in the upper part of Figure 5. Based on the measurement of relevant parameters (e.g., available bandwidth or receiver buffer fullness – more details are discussed in Section 3.4), the client's decision engine selects which representation to download next. In this work, the main focus is on the decisions of a single HAS instance and their impacts on QoE. However, there is a complex interplay of the control loop with other applications and the network which can also affect QoE. Therefore, the interactions between different HAS players and the interactions with the TCP congestion control loop are discussed in Section 7.

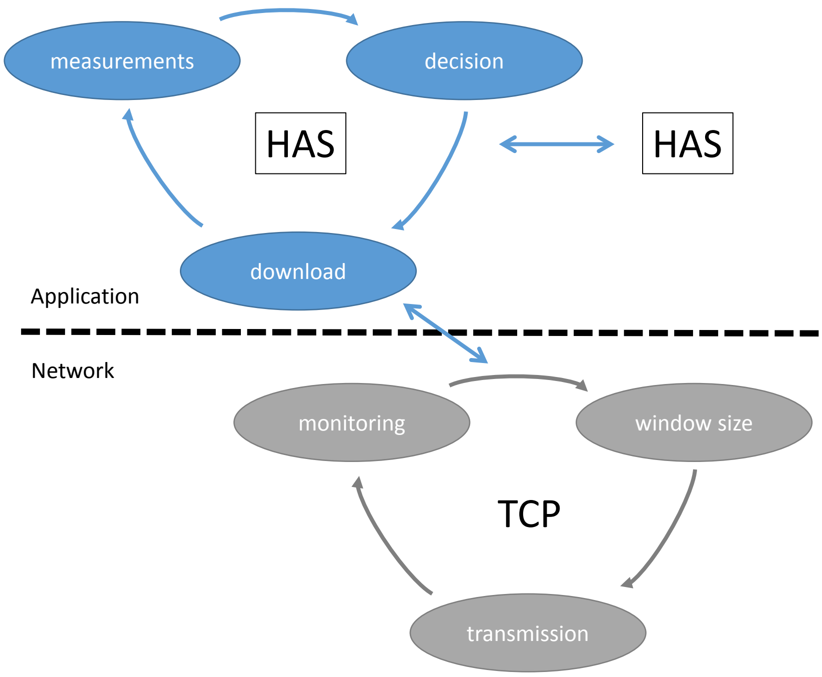

Figure 5: Control loop of HAS. Based on measurements the client decides which segment to download next. The control loop interacts with other applications (e.g., other HAS instances) and the network (e.g., TCP congestion control loop) which can also affect QoE.

Although the different HAS solutions share the basic principle as illustrated in Figure 3, they differ in a number of technical parameters. The main features of the proprietary and standard HAS solutions are summarized in Table 1 and Table 2, respectively. In the context of this paper, the following parameters are of high importance:

Table 1: Comparison of different *proprietary* HTTP adaptive streaming solutions.

Proprietary solution	Silverlight Smooth Streaming – MSS	HTTP Live Streaming – HLS	HTTP Dynamic Streaming – HDS
Owner	Microsoft	Apple	Adobe
Data description	Manifest (xml)	Playlist file (m3u8)	Manifest (f4m)
Video codec	H.264, VC-1	H.264	H.264, VP6
Audio codec	AAC, WMA	AAC (HE, LC), MP3, AC3	MP3, AAC
Format	fMP4 *.ismv + *.isma files	M2TS *.ts files	fMP4 *.f4f files
Segment length (typical)	2 s	10 s	2 – 8 s

Table 2: Comparison of different *standard* HTTP adaptive streaming solutions.

Standard solution	MPEG DASH	3GP DASH	HbbTV DASH
Owner	Standard	Standard	Recommendation
Data description	Media Presentation Description (XML)	Media Presentation Description (XML)	Media Presentation Description (XML)
Video codec	any	H.264	H.264
Audio codec	any	Enhanced aacPlus, AAC-LC, AAC-LTP Extended AMR-WB	HE-AAC, E-AC3
Format	MP4 or M2TS	3GPP File Format	MP4
Segment length (typical)	not specified	not specified	1 – 15 s

Codec. Although we can find codec agnostic (MPEG DASH) or variable codec (MSS, HDS) solutions, several systems are tailored to specific supported coding algorithms for both video and audio. Currently there is an obvious domination of H.264 for video, which is supported by all solutions in our scope, however it can be expected that the recently standardized H.265/HEVC (High Efficiency Video Coding) [33] will take part of its share in the coming years. For audio, the codec selection flexibility is generally higher. The implications of audio and video encoder selection and configuration are further discussed in Section 3.3.

Format. The currently dominant formats for media encapsulation are the MPEG-2 transport stream (M2TS [34]; used by HLS, DASH) and ISO Base Media File Format (MP4 [35]; used by DASH) or its derivative denoted to as fragmented MP4 (fMP4 [36, 37]; used by HDS and MSS).

MPEG-2 transport streams carry the data in packets of a fixed 184-byte length plus a 4-byte header. Each packet contains only one type of content (audio, video, data or auxiliary information). M2TS is commonly used for streaming, although its structure is not as flexible as in the case of MP4. MP4 organizes the audiovisual data in so-called boxes and treats different types of data separately. Further, there is a high flexibility in handling the auxiliary information (such as codec settings), which can be tailored to the needs of the application. As such, MP4 has been successfully adjusted to carry prioritized bit streams of scalable video over HAS [38], which we further consider in Section 3.3.

Segment length. The segment length used in a HAS system specifies the shortest video duration after which a quality (bitrate) adjustment can occur. Although some systems keep these values fixed (Table 1), the segment length can be left up to the individual implementation in many cases (Table 2). Again, we will further discuss the implications related to segment length design in Section 3.3.

It is clear from the principle of adaptive HTTP streaming that the decision engine responsible for selecting appropriate representations is running on the client side and needs to select the representation to request based on different criteria. These criteria can be measurements of downlink throughput, the actual video buffer status, device or screen properties, or context information (e.g., mobility). On the server side, in contrast, the most important decisions are done on the preparation of the content, i.e., what representations shall be provided, and on its delivery, e.g., the selection of the best CDN server for each request.

The behavior of the HAS system needs to obey the requirements of the actual use case. For example, in the context of a live system, the content is being made available at the server during the viewing and a low overall delay introduced by the system shall be achieved. This implies that the provided segment lengths should be small and the segments need to start downloading as soon as they appear on the server. For video on demand systems, on the contrary, a larger receiver buffer can be used together with longer segments to avoid flickering caused by frequent quality representation changes. The following sections discuss the actions the system or its designer can take on the server side and on the client side in order to efficiently adapt to the actual conditions while considering context requirements.

3.3 Server-side Actions

On the server side of an HTTP adaptive streaming system, the main concern is the preparation of the content, i.e., selection of available representations, and optimal encoding. This also includes proper selection of system parameters, such as the length of an encoded segment (where selectable).

The length of the video segment needs to obey two contradictory requirements. First, the segments need to be short enough to allow for fast reaction to changing network conditions. On the other hand, the segments need to be long enough to allow high coding efficiency of the source video encoder [39] and to keep the amount of overhead low (the impact of segment size on the necessary overhead is quantified in [40]). Clearly, these two requirements form an optimization problem which needs to be considered at the server side during content preparation.

In [41], the length of video segments to be offered to the client is optimized based on the content, so that I-frames and representation switches are placed at optimal positions, e.g., video cuts. Such an approach led to approximately 10% decrease of the required bit rate for a given video image quality. This work is followed by [42], where variable segment lengths across different representations are considered – it is proposed that for higher bit rates, longer segments are used in order to improve coding efficiency.

Among the server-side actions is also the selection of compression algorithms for audiovisual content (in cases where it is not fixed by the system specification). A recent comparison of different video compression standards [43] justifies the very wide spread use of H.264/AVC (Advanced Video Coding) encoding [44] for video as shown in Tables 1 and 2, although codec flexibility, available in several proprietary and standard solutions, is a clear advantage due to the emerging highly efficient HEVC (High Efficiency Video Coding) standard [43].

Apart from single-layer codecs like H.264/AVC or HEVC where only different representations (i.e., different files) of the same video can be switched, multi-layer codecs can be used which enable bitstream switching. Such features were also introduced in AVC (cf. SP/SI synchronization/switching frames [45]) and date back to the MPEG-2 standard [46] which needed a big overhead to achieve scalability in times when processors were hardly fast enough.

A modern multi-layer codec is Scalable Video Coding (SVC) which is an amendment to AVC and offers temporal, spatial, and image quality scalability [47]. This means, SVC allows for adaptation of frame rate and content resolution, and switching between different levels of image quality. It makes use of difference coding of the video content such that data in lower layers can be used to predict data or samples of higher layers (so called enhancement layer). In order to switch to a higher layer, only the missing difference data have to be transmitted and added. This is the major difference to adaptation with single-layer codecs that quality can be increased incrementally in case of spare resources. In contrast with single-layer codecs, a whole new segment has to be downloaded, and already downloaded lower quality segments have to be discarded. The trade-off is the overhead introduced by multi-layer codecs. This means, for example, that a SVC file of a video of a certain bit rate is larger compared to an AVC file of the same video and same bit rate. However, for adaptation with AVC, for each quality level a separate video file is needed, whereas with SVC, a single file is sufficient.

By applying a hierarchical coding schema SVC allows for the selection of a suitable sub-bitstream for the on-the-fly adaptation of the media bitstream to device capabilities and current network conditions. A valid sub-bitstream contains at least the AVC-compatible base layer and zero or more enhancement layers. Note that all enhancement layers depend on the base layer and/or on the previous enhancement layer(s) of the same scalability dimension. The subjective evaluation of various SVC configurations is well known [48, 49] but papers dealing with DASH and QoE focus on AVC, and the integration of DASH and SVC is only evaluated using objective metrics [50] and simulations [51, 52] so far.

The architecture of HTTP adaptive streaming systems allows for optimization of the server load. In [53], for instance, the authors propose a scheme for balancing the load among several servers through altering the addresses in the DASH manifest files.

It is obvious that the main concern in the server-side actions is the selection of appropriate coding for the different available quality levels. This includes not only the selection of the compression algorithm and its settings, but also the decision on adaptation dimensions to be employed. An overview of the available adaptation dimensions will follow in Section 5.

3.4 Client-side Actions

On the client side, the most important decisions are, which segments to download, and when to start with the download and how to manage the receiver video buffer. The adaptation algorithm (decision engine) should select the appropriate representations in order to maximize the QoE, which can be achieved in several different ways. The most common approach is to estimate the instantaneous channel bandwidth and use it as decision criteria.

The receiver video buffer size is dealt with in [54], where the authors perform an analysis of receiver buffer requirement for variable bit rate encoded bit streams. They found that the optimal buffer length depends on the bitstream characteristics (data rate and its variance) as well as network characteristics and, of course the desired video QoE represented by initial delay and rebuffering probability.

A recent work [55] reviews the available bitrate estimation algorithms and describes the active and passive bitrate measurement approaches. The passive measurement requires no additional probe packets to be inserted in the network, which results in no additional overhead at the expense of less accurate results. No additional overhead is the reason why passive tools are generally used for available bitrate estimation in HAS. The author of [55] classifies the passive measurement approaches to *cross layer-based*, where the protocol stack is modified to obtain packet properties (e.g., [56]) and *model-based*, which employ throughput modeling such as [57] for wireless LAN networks using TCP or UDP transport protocols. The following paragraphs describe different adaptation algorithms generally based on passive measurement of available bit rate or segment download time.

Depending on the actual use case and scenario, different adaptation strategies can be employed to adapt to the varying available bitrate. In [58] the authors compare several segment request strategies in HAS for live services. The analysis uses passive measurement of segment arrival times, aiming at evaluating the video startup delay, end to end delay, and the time available for segment download through the analysis of different initial delay adjustment, the time to start downloading the next segment and the way to handle missing segments or their parts. It is shown that different strategies exhibit different behavior, and the adjustment needs to reflect network conditions and desired QoE priorities of the system.

For live and video on demand services, [59] and [60] describe a decision engine based on Markov Decision Process (MDP) using the estimated bit rate as input. [61] proposes a rate adaptation algorithm based on smoothed bandwidth changes measured through segment fetch time, whereas in [62], the authors propose an adaptation engine based on the dynamics of the available throughput in the past and the actual buffer level to select the appropriate representation. At the same time, the algorithm adjusts the required buffer level to be kept in the next run. Also in [63] an algorithm for single-layer content (e.g., AVC) of constant bit rate is presented which selects representations according to current bandwidth, current buffer level, and the average bit rate of each segment. For multi-layer content (e.g., SVC), [64] describes the Tribler algorithm, which relies on thresholds of downloaded segments, and [65] proposes the BIEB algorithm, which is also an SVC-based strategy computing segment thresholds based on size ratios between quality levels. Note that with multi-layer strategies different quality levels of the same time slot can be requested independently. Single-layer strategies can also request different representations of the same time slot, but only one can be used for decoding, and already downloaded other representations will be discarded.

It is important to mention that the presented algorithms select among the available representations just based on technical parameters like bandwidth or bit rate, but do not take the expected quality perceived by the end user into account.

3.5 Performance Studies

HTTP adaptive streaming uses the TCP transport protocol, which, although reliable, introduces higher overhead and delays compared to the simpler UDP, broadly used for video services earlier. As there are fundamental differences between TCP and UDP, studies have been done on justifying the use of TCP for video transmission. In [66], for instance, the authors employ discrete-time Markov models to describe the performance of TCP for live and stored media streaming without adaptation. It is found that the delivery is generally fluent when TCP throughput is roughly twice the video bit rate, i.e., there is a significant system overhead as the expense for reliable transmission.

A number of studies have been published aiming at comparison of the existing HAS solutions, both proprietary and standardized, in terms of performance. In [63], the authors compared MSS, HLS, HDS and DASH in a vehicular environment. They found that the best performance, represented by average achieved video bit rate and the number of switches among representations, was achieved by MSS among proprietary solutions and by Pipelined DASH among all the candidates. The idea behind Pipelined DASH is that several segments can be requested at a time in contrary to standard DASH. Pipelining is beneficial in vehicular and mobile scenarios, where packet loss might result in a poor usage of the available resources in case only one TCP connection is established. The drawback is that pipelining requires appropriate sending buffer control, i.e., server complexity is increased. More details on pipelining performance can be found in [67]. In [40], the performance of DASH for live streaming is studied. An analysis of performance with respect to segment size is provided, quantifying the impact of the HTTP protocol and segment size on the end to end delay.

In [65], a performance comparison of the adaptation algorithms described in [62, 63, 64, 65] was conducted. The traffic patterns used for the evaluation were recorded in realistic wired and vehicular mobility situations. In terms of average playback quality and bandwidth utilization, BIEB [65] and Tribler [64] could outperform the other algorithms significantly. Both algorithms delivered a high average playback quality to the user, but Tribler had to switch to a different quality nine times more often than BIEB. The algorithm of [62] showed better results than BIEB in some aspects, as it has a lower quality switching frequency and a better network efficiency because no data is unnecessarily downloaded and bandwidth is wasted. However, compared to the size of the movie, the segments discarded by BIEB were negligible. In the vehicular scenario, BIEB outperformed the other algorithms, but no performance results were provided so far for other scenarios.

Also other optimization criteria for algorithm performance assessment have been used like PSNR [68] or pseudo-subjective measures like engagement [69, 70]. These criteria are often assumptions regarding QoE impact, which date from earlier studies and have not been questioned nor verified with respect to their QoE appropriateness [71]. Thus, dedicated studies on the impact of adaptation strategies and application parameters on QoE will be presented in the following section. These results should be taken into account when designing a QoE-aware HAS algorithm.

4 QoE Influence of Adaptation Strategy and Application Parameters

A QoE model for adaptive video streaming which can be used for automated QoE evaluation is described in [79]. The authors found that adaptation strategy related parameters (stalling, representation switches) have to be taken into account and that they have to be considered on a larger time scale (up to some minutes) than video encoding related parameters (resolution, frame rate, quantization parameter, bit rate) which only influence in the order of a few seconds. In [5], QoE metrics for adaptive streaming, which are defined in 3GPP DASH specification TS 26.247, are presented. They include HTTP request/response transactions, representation switch events, average throughput, initial delay, buffer level, play list, and MPD information. However, the conducted QoE evaluation considers only stalling as most dominating QoE impairment. Other results regarding the QoE influence of adaptation parameters are summarized in Table 3 and will now be presented in more detail.

[19] investigated how playout buffer threshold and video segment size influence the number of stalling events. They found that a small buffer of 6 s is sufficient to achieve a near un-interrupted streaming experience under vehicular mobility. Further increasing the buffer size would lead to an increased initial delay and could also be an issue for memory constrained mobile devices. With video segment size, there is a trade-off between short segment sizes resulting in many small files which have to be stored for multiple bit rates of each video. Larger segment sizes, however, may not be sufficient to adapt to rapid bandwidth fluctuations especially in vehicular mobility and lead to more stalling. However, this effect can be balanced by increasing the buffer threshold, i.e., the amount of data which is buffered before the video playback starts. Thus, the authors stated explicitly that it is important to configure the buffer threshold in accordance with the used video segment size. Also [72] confirmed by simulations that a longer adaptation interval, i.e., longer time between two possible quality adaptations, leads to higher quality levels of the video and fewer quality changes. However, the number of stalling events and total delay increase. Additionally, [73] revealed an impact on players' concurrent behavior, such that large segment sizes allow for a high network utilization but have negative effects on fairness. These aspects beyond pure video QoE will be discussed in more detail in Section 7.

[75] showed that the active adaptation of the video bit rate improved or decreased the video quality according to the switching direction, but downgrading had a stronger impact on QoE than increasing the video bit rate. Thus, the authors argue that there could be a degradation caused by the switching itself. [77] investigated the adaptation of image quality for layer-encoded videos. They found that the frequency of adaptation should be kept as small as possible. If a variation cannot be avoided, its amplitude should be kept as small as possible. Thus, a stepwise decrease of image quality was rated slightly better than one single decrease. Also [76] compared smooth to abrupt switching of image quality. They confirmed that down-switching is generally considered annoying. Abrupt up-switching, however, might even increase QoE as users might be

Table 3: Effects of application parameters settings on adaptation and on subjectively perceived video quality.

No	**Adaptation** [19]	Yes
- More stalling - Worse bandwidth utilization		- Less stalling - Better bandwidth utilization
Small	**Buffer Size** [19]	Large
- More stalling - Less initial delay - Less memory requirements	Buffer size of 6s is sufficient	- Less stalling - More initial delay - Increased memory requirements
Small	**Adaptation Interval (Segment Size)** [19]	Large
- Less stalling - More files - Lower coding efficiency - Worse quality - More switches - Shorter delay - Lower network utilization		- More stalling - Fewer files - Higher coding efficiency [39] - Better quality [72] - Less switches [72] - Higher delay [72] - Higher network utilization [73] - Negative impact on fairness [73]
Low	**Adaptation Frequency** [74]	High
- Better than constant low quality - Decreasing further has no effect	Switching is degradation itself [75]	- Annoying
Down	**Adaptation Direction** [75]	Up
- Stronger impact	QoE according to direction	- Smaller impact
Low	**Adaptation Amplitude** [74]	High
- Image quality not detectable - Frame rate down to 15 fps not detectable - Resolution down to half of original size not detectable	Most dominant factor	- Low acceptance - Abrupt up-switching might increase QoE [76]
	Additional Effects	
	Memory effect [77]	- Higher quality in the end leads to higher QoE
	Base layer [77, 78]	- Higher base layer allows for longer impairments to be accepted

happy to notice the visual improvement. [77] found that a higher base (i.e., lowest quality) layer resulted in higher perceived quality which implies that segments which raise the base layer should rather be downloaded instead of improving on higher quality layers. Finally, they also observed a strong memory effect, i.e., higher quality in the end of a video clip leads to higher QoE. In [78] the impact of image quality adaptation on SVC videos is shown for a base layer and one enhancement layer. The authors found that a higher base layer quality allowed for longer impairments to be accepted. The duration of such impairments had linear influence on the perceived quality. For their 12 s video clips the influence of the number of impairments was only significant between one and two impairments, while the interval between impairments did not seem to have any significant influence.

In [80], the authors present an approach which overrides client adaptation decisions in the network in order to optimize QoE globally or for a group of users. However, this way of 'adaptation' is going beyond single user optimization as discussed within this section. A discussion on network level adaptation issues and related user experience problems follows in Section 7.

[74] investigated flicker effects for SVC videos, i.e., rapid alternation of base layer and enhancement layer, in adaptive video streaming to handheld devices. They identified three effects, namely, the period effect, the amplitude effect, and the content effect. The period effect, i.e., the frequency of adaptation, manifested itself such that high frequencies (adaptation interval less than 1 s) are perceived as more annoying than constant low quality. At low frequencies (adaptation interval larger than 2 s), quality was better than constant low quality, but saturated when decreased further. The amplitude was the most dominant factor for the perception of flicker as artifacts become more apparent. However, image quality adaptation was not detectable for most participants at low amplitudes. Also for temporal adaptation, changes between frame rates of 15 fps and 30 fps were not detected by half of the users. Only a reduction of image quality from 24 QP above 32 QP or frame rate reduction below 10 fps brought significant flicker effects which resulted in low acceptance for high frequencies. For spatial adaptation the authors indicated that the change of resolution should not exceed half the original size in order to deliver a generally acceptable quality. Finally, the content played a significant role in spatial and temporal adaptation. For image quality reduction, no significant effect could be found. The authors concluded that videos with complex spatial details were particularly affected by resolution reduction, while videos with complex and global motion required high frame rates for smooth playback. Thus, the perceived quality is affected differently by the different adaptation dimensions (i.e., image quality, spatial, or temporal adaptation). Section 5 will present these dimensions in detail and their influences on QoE will be outlined in Section 6.

5 Video Adaptation Dimensions

In order to follow the requirement of providing video content at different bitrates for HTTP adaptive streaming, one or several adaptation dimensions can be utilized. In the following paragraphs, we describe the possible adaptation dimensions and provide a real world example of the bit rate reduction efficiency of each approach. The real world example is based on encoding 20 seconds long sequences with different content (sport - 200 m sprint, cartoon - a clip from the Sintel movie [81], action - a car chasing scene from the movie “Knight and Day”) encoded with H.264/AVC using different frame rate (from 25 fps down to 2.5 fps), resolution (from 1920×1080 progressive down to 128×72 progressive) and quantization parameter (from 30 up to 51). All other encoding parameters remained unchanged during the experiment (the x264 codec implementation was used, high profile, level 4.0, adaptive GOP length up to 2 seconds).

Video frame rate based bit rate adaptation is relying on decreasing the temporal resolution of a video sequence, i.e., encoding a lower number of frames per second. The typical efficiency of such an approach is shown in Figure 6. The original frame rate of a progressive-scanned video sequence corresponding to 100 % is 25 fps in our real world example. In order to achieve 80 % of the original bit rate, one needs to reduce the frame rate to approximately 65 %. In such a case, motion in video is no longer perceived as smooth and the perceived quality degradation is significant [82].

Resolution based bit rate adaptation decreases the number of pixels in the horizontal and/or vertical dimension of each video frame. The corresponding efficiency of such an approach is shown in Fig. 7[1]. The steeper descend (compared to Figure 4) of the curves is quite advantageous – even a small decrease of frame resolution leads to a significant reduction of required bit rate (e.g., 80 % of the original bit rates is achieved by decreasing the frame size to approximately 85 % in both directions).

Quantization based bit rate adaptation adjusts the lossy source encoder in order to reach the desired bit rate. In H.264/ AVC, the available values of the quantization parameter (QP) are between 0 (lossless coding) and 51 (coarse quantization with poor visual quality of a sequence). Fig. 8 shows the bit rate descend for quantization parameter increasing from 30 up to 50. The steep decrease of bit rate in around QP 30 is getting flatter for QP values close to 50, which is quite natural as there is a certain amount of information in the encoded bit stream carrying data other than just quantized transform coefficients (e.g., prediction mode signalization, motion vector values, etc.).

The adaptation dimensions mentioned in this section can be further extended as described in [83], where the author proposes a three-level model to describe the user satisfaction. Apart from transcoding, which is essentially the core operation for HAS content preparation, [83] also mentions transmoding, i.e., conversion among different modalities – audio, video, or text, as an alternative approach to adaptation.

[1] Resolution was changed in both the vertical and the horizontal dimension in order to keep the aspect ratio unchanged.

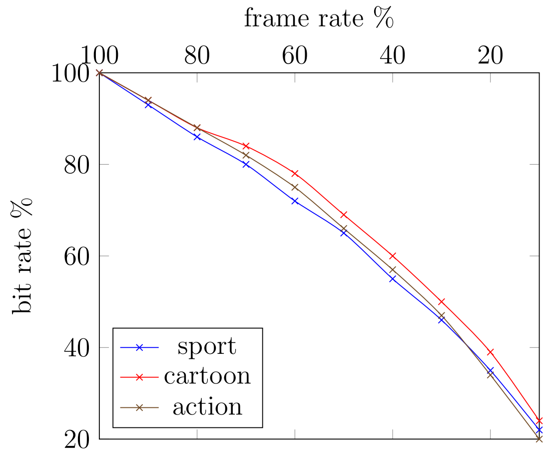

Figure 6: Adaptation through frame rate reduction.

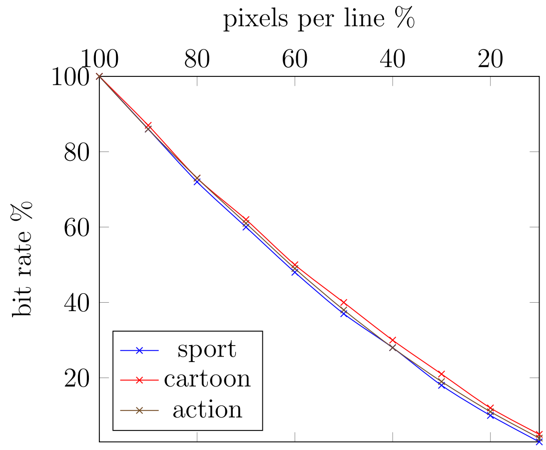

Figure 7: Adaptation through resolution reduction.

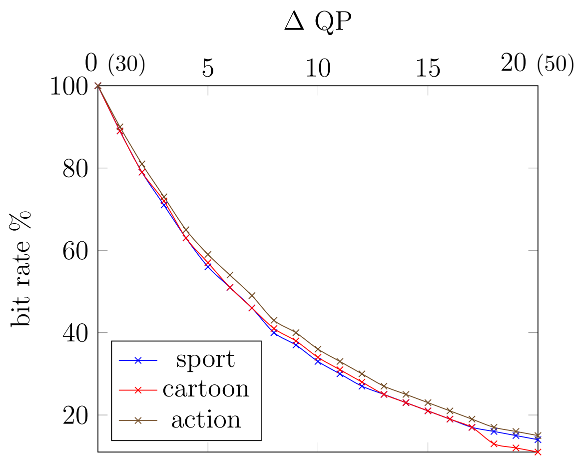

Figure 8: Adaptation through adjustment of transform coefficient quantization.

6 QoE Influence of Video Adaptation Dimensions

In this section, we review related work with respect to Quality of Experience of HTTP adaptive streaming. Therefore, we mainly focus on studies based on subjective user tests as these represent the gold standard for QoE assessment. Please note that most of these studies were not specifically designed for HAS, but the results were transferred where possible. However, some general issues still remain which are not addressed by research so far, e.g., long video QoE tests in the order of 10 minutes, which is a typical duration for user-generated content.

Several works (e.g., [84, 85, 86]) conclude that there is a content dependency regarding quantitative effects of adaptation. Especially spatial and temporal information of the video clips determine how the effects of adaptations are perceived. Thus, in this section no absolute results can be presented as they differ for each single video, but the main focus will be the presentation of general, qualitative effects of adaptation on QoE. Thereby, the different adaptation dimensions and their respective influence will be outlined first, and links to QoE models will be provided. A summary of this survey can be found in Table 4, but more details are given in the following three subsections. Afterwards, trade-offs between the different dimensions will be highlighted.

Table 4: Main QoE findings of single dimension adaptation. More references can be found in the detailed description in Section 6.

Dimension	**Main Findings**	**QoE Function**
Image Quality	- Image quality decrease leads to lower QoE [87] - Encoder has most significant effect [89] - There is a logistic relationship between coded bit rate and QoE [90]	- Objective per-frame/per-chunk metrics [88]
Spatial	- Lower resolution leads to lower QoE [91] - Higher resolution can be a drawback with increased level of distortion [93]	- Resolution, temporal and spatial information [92]
Temporal	- Stalling is worst degradation [17] - Lower frame rate leads to lower QoE [94] - QoE of decreased frame rate is highly dependent of content motion [84]	- Number and length of stalling events [4] - Frame rate and spatial information [92]

6.1 Image Quality Adaptation

[89] found that the perceptual quality of a decoded video is significantly affected by the encoder type. They confirmed work from 2005 in which the authors of [95] showed that the video quality produced by the H.264 codec was the most satisfying, rated higher than RealVideo8 and H.263. Also for low bitrates H.264 outperforms H.263 and also MPEG-4 [96]. The fact that encoder type significantly affects QoE was also shown for scalable video codecs in [86] which investigated adaptation with H.264/SVC and wavelet-based scalable video coding. [41] proposed an improved approach for encoding and segmentation of videos for adaptive streaming using H.264, which reduces the needed bit rate up to 30% without any loss in quality. Conversely, this means that by using their approach, the video can be encoded with a higher image quality for the same target bit rate.

In [75], the performance of both H.264 and MPEG-4 was investigated under different mobile network conditions (WiFi and HSDPA) and video bit rates. Therefore, the authors implemented their own on-the-fly video codec changeover and bit rate switching algorithm. With WiFi connection, H.264 is perceived better than MPEG-4 especially for low video bit rates. In contrast, with HSDPA connection, MPEG-4 yields better results for all bit rates. A change of codec during playback from H.264 to MPEG-4 always degraded the user perception. However, a change in the other direction was perceived as an improvement for low bit rates. Furthermore, the authors found that a bit rate decrease, i.e., a decrease of image quality, resulted in a decreased quality, but a bit rate increase is not always better for QoE as the switch itself might be perceived as an impairment.

[97], [90], and [87] investigated different video bit rates for different codecs and showed that an increased video bit rate leads to an increased video quality. In particular in [90], a logistic function describes the relationship between coded bit rate and subjective video quality for H.264, stating that with increasing bit rate the video quality increases but eventually saturates. Thus, a further increase of bit rate does not result in a higher perceived quality.

Changing the quantization parameter of H.264 video streams is in the focus of [18]. The authors found that QoE falls slowly when the quantization parameter starts to increase, i.e., the video bit rate decreases and the image quality gets worse. Only after reaching a high quantization parameter the perceived quality drops faster. [89] found that in order to reach a good or excellent QoE the pixel bit rate should be around 0.1 bits per pixel when H.264 is used. If other information like frame rate or frame size are unavailable, pixel bit rate can serve as a rough quantitative gauge for QoE. [98] used linear models and per-chunk metrics to predict the MOS of video sequences with image quality adaptations. [88] showed that the perceived quality of HTTP video streams with dynamic quantization parameter adaptation can be predicted by temporal pooling of objective per-frame metrics. The authors state that a simple method like the mean of quality levels of a HAS video sequence delivers already very decent prediction performance.

6.2 Spatial Adaptation

[99, 91] found that spatial resolution was the key criteria for QoE for small screens. Low resolutions contributed to enhanced eyestrain of the subjects. However, acceptability of spatial resolutions was also tied to shot types. [89] found that in general higher MOS is associated with higher spatial resolution. However, they showed that for the same video bit rate, a video with higher spatial resolution is perceived worse. This is due to a lower pixel bit rate value (cf. above) which causes severe intra-frame degradations especially for sequences with large spatiotemporal activities. In [93] the impact of resolution on subjective quality is investigated by comparing HDTV and SDTV sequences. They found that QoE increases with increasing resolution for slightly distorted images. However, larger image size becomes a drawback when the level of distortion increases as artifacts are more prevalent and disturbing in HD. In this case observers tend to prefer SD as this reduces the visual impact of the distortions. In [92] a model for mapping resolution to MOS is presented. The considered resolutions ranged from SD to SQCIF. The authors show that MOS is a function of resolution, spatial information, and temporal information. Moreover, they found that temporal information is more important than spatial information for their model.

6.3 Temporal Adaptation

In [94], the effect of frame dropping on perceived video quality was investigated and a model for predicting QoE was proposed. It was shown that frame dropping has a negative impact on QoE and the quality impairment is depending on motion and content of the sequence. Moreover, the authors found that periodic frame dropping, i.e., decrease of frame rate, is less annoying than an irregular discarding of frames. [100] investigated the influence of frame rate on the acceptance of video clips of different temporal nature and importance of auditory and visual components. The authors showed that reducing the frame rate generally has a negative influence on the users' acceptance of the video clips. However, low temporal videos, i.e., videos with little motion, were affected more by lower frame rates than high temporal videos. This finding was confirmed by [84] which investigated different frame rates for dynamic content. They showed that reducing the frame rate generally leads to a lower user satisfaction, but it does not proportionally reduce the users' understanding and perception of the video. Instead they found a complex link between understanding and perception, i.e., users have difficulty to absorb audio, video, and textual information concurrently. Thus, highly dynamic clips for which it is difficult to assimilate all information, such as sports or action clips, were unaffected by reduced frame rate. On the other hand, a static news clip which can be understood easily as the most important information is delivered by audio, suffers from reduced frame rate because the lack of lip synchronization is clearly visible. Also [101] reconfirmed the findings that a reduction of frame rate has only little impact on high motion videos. [102] confirmed these findings in a study on MPEG1 subjective quality which additionally took into account packet loss, and stated that the MOS was good if the frame rate was more than about 10 frames per second. More recently, [103] stated that the threshold of subjective acceptability is around 15 fps. [104] found that the optimal frame rate for a given bit rate depends on the type of motion in a sequence. They found that videos with jerky motion benefit from increased image quality at lower frame rates. Clips with smoother (i.e., less jerky) motion were generally insensitive to changes in frame rate. [17] found that only for very low frame rates the quality decreased as the impairment duration increased. For medium or high frame rates the quality was similar whether the frame rate reduction occurred during the entire video or during a short part of the video. Thus, they state that there is no quality gain by re-increasing frame rate after a temporary drop. In [92] a model for mapping frame rate to MOS is given. MOS can be expressed as a function of frame rate and spatial information. Adding temporal information did not improve the model performance.

6.4 Trade-offs Between Different Adaptation Dimensions

The three dimensions presented not only allow for single dimension adaptation but also for combined quality changes in multiple dimensions. Several studies consider trade-offs between different adaptations and will be presented in this section. Table 5 summarizes the main findings and links to the corresponding works.

[71] claimed that there exists an encoding which maximizes the user-perceived quality for a given target bit rate which can be extended to an optimal adaptation trajectory for a whole video stream. In their work the authors focus on the adaptation of MPEG-4 video streams within a two-dimensional adaptation space defined by frame rate and spatial resolution. They showed that a two-dimensional adaptation which reduces both resolution and frame rate outperforms adaptation in one dimension. Comparing clips of similar average bit rates, it is shown that reduction of frame rate is perceived worse than reduction of resolution. In [85], the authors researched optimal combinations of image quality and frame rate for given bit rates. They found that until image quality improves to an acceptable level, it should be enhanced first. Once it has improved adequately, temporal quality should be improved. Especially spatially complex videos require a high image quality first, while videos with high motion require a higher frame rate at a lower bit rate, which was already discussed in the previous subsection. Also in [105], for a given bit rate trade-offs between frame rate and image quality were presented. A trend was found that for decreasing video bit rate also the optimal frame rate decreases. The authors show that for different video bit rates there exist switching points which define multiple bit rate regions requiring a different optimal frame rate for adaptation. [106] investigated trade-offs between frame rate and quantization for soccer clips. The authors found that participants were more sensitive to reductions in frame quality than to reduced frame rate. Especially for small screen devices, a higher quantization parameter removed important information about the players and the ball. In contrast, a low frame rate of 6 fps was accepted 80% of the time although motion was not perceived as being smooth. The experimental results obtained by [107] show that image quality is valued higher by test users (which were able to choose which distortion step they preferred) than temporal resolution of the content for low bitrate videos. [108] confirmed that for fast foreground motion like soccer reducing frame rate was preferred to reducing frame

quality. However, for fast camera or background motion a high frame rate is better because disturbing jerkiness can be detected more easily which results in lower QoE.

In [101], trade-offs between resolution and frame quality are investigated. They found that a small resolution (without upscaling) and high image quality is preferred to a large resolution and low frame quality for a given bit rate. [86] compared different combinations of resolution, frame rate, and pixel bit rate, which resulted in similar average video bit rates. They found that at low bit rates a larger resolution is preferred and thus frame rate should be decreased. At high bit rates, frame rate is more important and pixel bit rate should be decreased to achieve a high perceived quality. [89] maximized the QoE by selecting an optimal combination of frame rate and frame size under limited bandwidth, i.e., video bit rate. They found that in general resolution should be kept low. For videos with a high frame difference and variance, also frame rate should be low (which implies a high pixel bit rate). Instead frame rate should be high for content with low temporal activity in order to achieve a high QoE.

Table 5: Trade-offs between different adaptation dimensions. A $\triangleright$ B means that dimension A is more important than dimension B, i.e., a degradation of A is worse than a degradation of B.

Major Finding	Detailled Description
Stalling $\triangleright$ Adaptation [5]	- Stalling $\triangleright$ Initial Delay [89] - Stalling $\triangleright$ Image Quality [18] - Stalling $\triangleright$ Frame Rate [17]
Image Quality $\triangleright$ Frame Rate	- Spatially complex videos [85] - Slow motion videos [101] - Small screen devices [106] - Fast foreground motion [108] - High frame difference and variance [89]
Frame Rate $\triangleright$ Image Quality:	- Fast camera or background motion [108] - High bit rates [86] - Low temporal activity [89]
Image Quality $\triangleright$ Resolution	- Different types of videos [101]
Frame Rate $\triangleright$ Resolution	- Different types of videos [71]
Resolution $\triangleright$ Frame Rate:	- Low bit rates [86]

7 User Experience Related Impairments Beyond Video QoE

Within this section we discuss challenges which affect the QoE beyond video playback for users of HAS enabled devices or users in networks in which HAS clients are established. This discussion includes several network related issues that arise from the particular behavior of video adaptation algorithms established in HAS clients, and their interplay with network level optimization algorithms. Within Section 3.2 this interplay between application level control loop and network level control loop has already been introduced and depicted in Figure 5. Typically such an interplay can introduce numerous different issues, especially as more network level control loops beyond TCP can be involved, e.g., in wireless networks. However, we only address those issues that affect the QoE relevant dimensions, as discussed in previous sections. Therefore, this section is very selective in terms of network issues and hence incomplete. A more thorough discussion of network related HAS problems can be found in [109, 68]

The problems described here reach beyond pure video based quality, as they also tackle availability and stability of the single network connection, the respective HAS client is utilizing, as well as the utilized network infrastructure at large. We first discuss issues between concurrent HAS clients and the impact of certain HAS configurations regarding interworking between several clients, and second describe the impact HAS client behavior exerts on the TCP protocol.

If several adaptive players share a network connection the following questions have been identified (cf. [109]) to be of particular interest:

- Can the players share the available bandwidth in a stable manner, without experiencing oscillatory bitrate transitions?
- Can they share the available bandwidth in a fair manner?
- How does the number of competing streams affect stability and fairness? And how do different adaptive players compete with each other?
- How does a player compete with TCP bulk transfers?

In the following, we will address the first three questions from above whereas the last question will be discussed in more detail at the end of this section.

7.1 Interactions between HAS player instances

One major issue for competing HAS players within a network is stability (in terms of quality levels) and the amount of switching events. This results from the oscillation between different available bandwidth estimates. It has been shown by several studies that such a behavior is invoked when more than one instance of HAS players compete for bottleneck bandwidth [109, 110]. Depending on the amount of available bandwidth and the time the different players join the stream the reaction is different. In [109] the authors have shown that in a two player setup the client joining the stream first grabs the highest available video bandwidth the network supports. Following, the second client joining starts with a low video bandwidth and tries to increase his video bandwidth subsequently. However, the throughput needed for the higher video bandwidth can not be provided by the network as the first client utilizes this bandwidth already, hence the buffer levels of the second client depletes, and the video adaptation algorithm switches to a lower video bandwidth. This leads to a permanent oscillation between different video bandwidths, and a large number of quality switching events for the second client, respectively. A similar finding is reported in [73] for a larger number of player instances that also results in a high number of quality oscillation for all players. Their results have also identified that the adaptation algorithms of different player realizations do maintain larger buffers in case of the Netflix implementation which results in over-estimation of available bandwidth due to good buffer fill levels. Contrary, the Silverlight Smooth streaming solution reacts too late on short-term available bandwidth peaks and then tries to maintain the high bandwidth setting too long which leads to occasional stalling events which should be avoided through adaptive streaming technologies. Associated to a high number of quality adaptation is also the problem of the delay required to converge to the final bit rate. This is a problem especially for short viewing sessions like 2 - 3 min clips where the client will not reach the maximum available bandwidth due to the above described behavior (cf. [110]).

Another challenge associated with the behavior of competing players is fairness between different HAS clients. If one player on network bottleneck grasps a large share of the bandwidth before the other players join the stream it will be privileged throughout its whole streaming session while the other players on the network compete for the remaining bandwidth as shown in [109, 110] and with a rising number of competing clients the unfairness increases as well [73]. However, the problem is not only an unfair distribution of network resources but as a result the bandwidth utilization within the bottleneck network also drops [111]. It has also been shown by [109] that this behavior is not based on TCP's well known unfairness towards connections of different RTT's but rather a result of the competing HAS clients. Although, in the case of TCP fairness issues, solutions to prevent unfair behavior between competing clients do exist, they are obviously not yet applied to HAS client control mechanisms.

As long as these client control mechanisms are not applied, such unfair behavior can be countered by a federated instance that overrides client adaptation decisions. Thereby, comparable quality levels for all clients globally in the network would be ensured, as discussed in Section 4 and shown by [80].

The above mentioned issues are strongly connected to the HAS parameterization of number of parallel streams, segment size and inter-request gap. An in-depth analysis by [73] has shown that segment size influences the players' concurrent behavior severely. While large segment sizes allow high network utilization, they negatively impact fairness between the players on the other hand. In addition the authors in [73] have identified the inter-request gap between the different segments as influencing factors in terms of fairness. The higher the inter-request gap between the segments the more likely is a fair allocation of bandwidth between the different players. These results on segment size are closely connected to the question of *how to derive the optimum between buffer size and number of switching events as well as the ability to react on network changes?* [70] while maintaining user perceived quality on a network level.

Overall, the related work paints no clear picture on how to design an effective rate-adaptation logic for a complex and demanding application (video streaming) that has to function on top of a complex transport protocol (TCP). The interactions between two feedback loops (rate-adaptation logic at the application layer and TCP congestion control at the transport layer as depicted in Figure 5) are not yet understood well.

7.2 Interactions between TCP and HAS clients

Beyond the above discussed influence of different HAS clients amongst each other there is also the problem of their influence on other applications using the network connection. One of the main problems arising is the interaction between aggressive HAS client behavior that periodically requests small files (video segments) over HTTP. This causes TCP to overestimate the bandwidth delay product of the transmission line and results in a buffer bloat effect as shown in [112], which in turn leads to queuing delays reaching up to one second and being over 500 ms for about 50% of the time. Having a one way queuing delay close to one second makes it almost impossible to use the bottleneck link for anything else but video transmission. This is particularly concerning as they also showed that active queue management (AQM) techniques, a widely believed solution to this problem, do not manage to eliminate large queuing delays [112].

This section has shown that beyond pure video QoE the egoistic behavior of current adaptive video strategies leads to unstable network conditions, an unfair distribution of network resources and under-utilization of these resources. These do not only impact the HAS clients but also severely impact other applications by large queuing delays. Despite the numerous work on the identification of these issues there are currently only network optimized counter strategies around but no solutions that try to solve these issues with respect to optimizing user perceived quality of the video stream.

8 Conclusion

In this work, the evolving research field of HAS was surveyed. Possibilities for adaptation in different dimensions and their gains were discussed. Current solutions, their development and milestones, their technology, and their adaptation parameters were presented. As these solutions are not QoE-driven so far and only offer what can be called a "best-effort QoE", this work outlined the influence of adaptation on QoE. It covered single dimension adaptation as well as trade-offs between different dimensions, and the influence of the adaptation strategy.

From investigating adaptation strategy parameters, it could be found that stalling, initial delay, memory requirements, and bandwidth utilization heavily depend on buffer size and chunk size (adaptation interval). However, a buffer of 6 s was shown to be sufficient for most bandwidth conditions. Moreover, the adaptation frequency should be rather low, as switching is a degradation itself. The most dominating factor is the adaptation amplitude, for which a high amplitude (i.e., a detectable quality change) results in low acceptance and perceived quality. Apart from these parameters, also memory effect and base layer quality influences the QoE of users.

From several studies reported in related work, it could be confirmed that HAS clearly outperforms classical streaming as it significantly reduces stalling which is considered to be the worst quality degradation. For each adaptation dimension, main findings and QoE functions have been presented. Multi-dimensional adaptation outperforms single dimension adaptation, and thus, should be considered in future HAS mechanisms. The order of importance of the different adaptation dimension is image quality before frame rate and finally resolution, i.e., a decrease of image quality is perceived worst. Although this order seems to be valid for most video contents, there exist some special cases in which the order can be different.

Beyond the impact of adaptation on pure video QoE, we also showed that user related impairments such as network stability issues and high RTT's can be caused by the aggressive behavior of HAS clients. Additionally, the interrelation between different player instances causes bad QoE and does also impact the performance of TCP on network level.

In this work we discussed numerous related work on the Quality of Experience of HAS in order to foster future research and development of new mechanisms. We showed that current HAS solutions only decide on adaptation based on bandwidth measurements. Following, the resulting QoE which is affected by adaptation is not optimal. Thus, future solutions should be QoE-driven and also take QoE influences into account when deciding which quality adaptation to perform.

References

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2012-2017," tech. rep., Cisco, 2013.

[2] J. Roettgers, "Don't Touch That Dial: How YouTube is Bringing Adaptive Streaming to Mobile, TVs," 2013.

[3] A. Sackl, P. Zwickl, and P. Reichl, "The Trouble with Choice: An Empirical Study to Investigate the Influence of Charging Strategies and Content Selection on QoE," in *Proceedings of the 9th International Conference on Network and Service Management (CNSM)*, (Zurich, Switzerland), 2013.

[4] T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, and P. Tran-Gia, "Quantification of YouTube QoE via Crowdsourcing," in *Proceedings of the IEEE International Workshop on Multimedia Quality of Experience - Modeling, Evaluation, and Directions (MQoE 2011)*, (Dana Point, CA, USA), 2011.

[5] O. Oyman and S. Singh, "Quality of Experience for HTTP Adaptive Streaming Services," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 20–27, 2012.

[6] P. Le Callet, S. Möller, and A. Perkis, eds., *Qualinet White Paper on Definitions of Quality of Experience (2012)*, (Lausanne, Switzerland), European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), 2012.

[7] R. K. P. Mok, E. W. W. Chan, X. Luo, and R. K. C. Chan, "Inferring the QoE of HTTP Video Streaming from User-Viewing Activities," in *Proceedings of the ACM SIGCOMM Workshop on Measurements Up the STack (W-MUST)*, (Toronto, ON, Canada), 2011.

[8] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea," in *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX 2012)*, (Yarra Valley, Australia), 2012.

[9] S. Egger, P. Reichl, T. Hoßfeld, and R. Schatz, "Time is Bandwidth? Narrowing the Gap between Subjective Time Perception and Quality of Experience," in *Proceedings of the IEEE International Conference on Communications (ICC 2012)*, (Ottawa, ON, Canada), 2012.

[10] R. E. Kooij, A. Kamal, and K. Brunnström, "Perceived Quality of Channel Zapping," in *Proceedings of the Communication Systems and Networks (CSN 2006)*, (Palma de Mallorca, Spain), 2006.

[11] S. Egger, T. Hoßfeld, R. Schatz, and M. Fiedler, "Waiting Times in Quality of Experience for Web Based Services," in *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX 2012)*, (Yarra Valley, Australia), 2012.

[12] A. Raake and S. Egger, "Quality and Quality of Experience," in *Quality of Experience: Advanced Concepts, Applications and Methods* (S. Möller and A. Raake, eds.), Springer, 2014.

[13] A. Finamore, M. Mellia, M. M. Munafò, R. Torres, and S. G. Rao, "YouTube Everywhere: Impact of Device and Infrastructure Synergies on User Experience," in *Proceedings of the 2011 Internet Measurement Conference*, (Berlin, Germany), 2011.

[14] L. Chen, Y. Zhou, and D. M. Chiu, "Video Browsing - A Study of User Behavior in Online VoD Services," in *Proceedings of the 22nd International Conference on Computer Communications and Networks (ICCCN)*, (Nassau, Bahamas), 2013.

[15] Y. Qi and M. Dai, "The Effect of Frame Freezing and Frame Skipping on Video Quality," in *Proceedings of the 2nd International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2006)*, (Pasadena, CA, USA), 2006.

[16] T. N. Minhas and M. Fiedler, "Impact of Disturbance Locations on Video Quality of Experience," in *Proceedings of the 2nd Workshop of Quality of Experience for Multimedia Content Sharing (QoEMCS 2011)*, (Lisbon, Portugal), 2011.

[17] Q. Huynh-Thu and M. Ghanbari, "Temporal Aspect of Perceived Quality in Mobile Video Broadcasting," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 641–651, 2008.

[18] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of Experience Estimation for Adaptive HTTP/TCP Video Streaming Using H. 264/AVC," in *Proceedings of the 2012 IEEE Consumer Communications and Networking Conference (CCNC)*, (Las Vegas, NV, USA), 2012.

[19] J. Yao, S. S. Kanhere, I. Hossain, and M. Hassan, "Empirical Evaluation of HTTP Adaptive Streaming Under Vehicular Mobility," in *Proceedings of the 10th International IFIP TC 6 Networking Conference: Networking 2011*, (Valencia, Spain), 2011.

[20] T. Zinner, T. Hoßfeld, T. N. Minash, and M. Fiedler, "Controlled vs. Uncontrolled Degradations of QoE – The Provisioning-Delivery Hysteresis in Case of Video," in *Proceedings of the 1st Workshop of Quality of Experience for Multimedia Content Sharing (QoEMCS 2010)*, (Tampere, Finland), 2010.

[21] Move Networks, "Move Networks," 2010.

[22] A. Zambelli, "A History of Media Streaming and the Future of Connected TV," *The Guardian*, 2013.

[23] D. F. Brueck and M. B. Hurst, "Apparatus, System, and Method for Multi-bitrate Content Streaming," 2010.

[24] A. Zambelli, "Smooth Streaming Technical Overview," tech. rep., Microsoft Corporation, 2009.

[25] Apple Inc., "HTTP Live Streaming Overview," 2013.

[26] Adobe Systems Inc., "HTTP Dynamic Streaming," 2013.

[27] European Telecommunications Standard Institute (ETSI), "Universal Mobile Telecommunication System (UMTS); LTE; Transparent End-to-end Packet-switched Streaming Service (PSS); Protocols and Codecs (3GPP TS 26.234 version 9.1.0 Release 9)," 2009.

[28] European Telecommunications Standard Institute (ETSI), "Universal Mobile Telecommunication System (UMTS); LTE; Transparent End-to-end Packet-swiched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH) (3GPP TS 26.247 version 1.0.0 Release 10)," 2010.

[29] International Standards Organization/International Electrotechnical Commission (ISO/IEC), "23009-1:2012 Information Technology – Dynamic Adaptive Streaming over HTTP (DASH) – Part 1: Media Presentation Description and Segment Formats," 2012.

[30] DASH Industry Forum, "For Promotion of MPEG-DASH," 2013.

[31] DASH Industry Forum, "Guidelines for Implementation: DASH-AVC/264 Interoperability Points," 2013.

[32] T. C. Thang, Q.-D. Ho, J.-W. Kang, and A. Pham, "Adaptive Streaming of Audiovisual Content Using MPEG DASH," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 1, pp. 78–85, 2012.

[33] International Standards Organization/International Electrotechnical Commission (ISO/IEC), "23008-2:2013 Information Technology – High Efficiency Coding and Media Delivery in Heterogeneous Environments – Part 2: High Efficiency Video Coding," 2013.

[34] International Standards Organization/International Electrotechnical Commission (ISO/IEC), "13818-1:2000 Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: Systems," 2000.

[35] International Standards Organization/International Electrotechnical Commission (ISO/IEC), "14496-12:2005 Information Technology – Coding of Audio-visual Objects – Part 12: ISO Base Media File Format," 2005.

[36] T. Siglin, "Unifying Global Video Strategies: MP4 File Fragmentation for Broadcast, Mobile and Web Delivery," tech. rep., Transitions, Inc., 2011.

[37] Adobe Systems Inc., "F4V/FLV Technology Center," 2013.

[38] I. Kofler, R. Kuschnig, and H. Hellwagner, "Implications of the ISO Base Media File Format on Adaptive HTTP Streaming of H.264/SVC," in *Proceedings of the 2012 IEEE Consumer Communications and Networking Conference (CCNC)*, (Las Vegas, NV, USA), 2012.

[39] K. Ramkishor, T. S. Raghu, K. Siuman, and P. S. S. B. K. Gupta, "Adaptation of Video Encoders for Improvement in Quality," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2003)*, (Bangkok, Thailand), 2003.

[40] T. Lohmar, T. Einarsson, P. Frojdh, F. Gabin, and M. Kampmann, "Dynamic Adaptive HTTP Streaming of Live Content," in *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2011)*, (Lucca, Italy), 2011.

[41] V. Adzic, H. Kalva, and B. Furht, "Optimizing Video Encoding for Adaptive Streaming over HTTP," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 397–403, 2012.

[42] J. Lievens, S. Satti, N. Deligiannis, P. Schelkens, and A. Munteanu, "Optimized Segmentation of H.264/AVC Video for HTTP Adaptive Streaming," in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, (Ghent, Belgium), 2013.

[43] J. Ohm, G. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, 2012.

[44] International Standards Organization/International Electrotechnical Commission (ISO/IEC), "14496-10:2012 Information Technology – Coding of Audio-visual Objects – Part 10: Advanced Video Coding," 2012.

[45] M. Karczewicz and R. Kurceren, "The SP-and SI-frames Design for H. 264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 637–644, 2003.

[46] International Standards Organization/International Electrotechnical Commission (ISO/IEC), "13818-2:1996 Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: Video," 1996.

[47] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[48] J.-S. Lee, F. De Simone, N. Ramzan, Z. Zhao, E. Kurutepe, T. Sikora, J. Ostermann, E. Izquierdo, and T. Ebrahimi, "Subjective Evaluation of Scalable Video Coding for Content Distribution," in *Proceedings of the 18th ACM International Conference on Multimedia*, (Florence, Italy), 2010.

[49] T. Oelbaum, H. Schwarz, M. Wien, and T. Wiegand, "Subjective Performance Evaluation of the SVC Extension of H.264/AVC," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP 2008)*, (San Diego, CA, USA), 2008.

[50] C. Müller, D. Renzi, S. Lederer, S. Battista, and C. Timmerer, "Using Scalable Video Coding for Dynamic Adaptive Streaming over HTTP in Mobile Environments," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012)*, (Bucharest, Romania), 2012.

[51] R. Huysegems, B. De Vleeschauwer, T. Wu, and W. Van Leekwijck, "SVC-Based HTTP Adaptive Streaming," *Bell Labs Technical Journal*, vol. 16, no. 4, 2012.

[52] Y. Sanchez, T. Schierl, C. Hellge, T. Wiegand, D. Hong, D. D. Vleeschauwer, W. V. Leekwijck, and Y. L. Loudec, “Efficient HTTP-based Streaming Using Scalable Video Coding,” *Signal Processing: Image Communication*, vol. 27, no. 4, pp. 329–342, 2012.

[53] Z. Li, M. Sbai, Y. Hadjadj-Aoul, A. Gravey, D. Alliez, J. Garnier, G. Madec, G. Simon, and K. Singh, “Network friendly video distribution,” in *Network of the Future (NOF), 2012 Third International Conference on the*, pp. 1–8, Nov. 2012.

[54] T. Kim, N. Avadhanam, and S. Subramanian, “Dimensioning Receiver Buffer Requirement for Unidirectional VBR Video Streaming over TCP,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP 2006)*, (Atlanta, GA, USA), 2006.

[55] T. Arsan, “Review of Bandwidth Estimation Tools and Application to Bandwidth Adaptive Video Streaming,” in *Proceedings of the 9th International Conference on High-Capacity Optical Networks and Emerging/Enabling Technologies (HONET 2012)*, (Istanbul, Turkey), 2012.

[56] Z. Yuan, H. Venkataraman, and G.-M. Muntean, “iBE: A Novel Bandwidth Estimation Algorithm for Multimedia Services over IEEE 802.11 Wireless Networks,” in *Proceedings of the 12th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services (MMNS 2009): Wired-Wireless Multimedia Networks and Services Management*, (Venice, Italy), 2009.

[57] Z. Yuan, H. Venkataraman, and G. Muntean, “MBE: Model-Based Available Bandwidth Estimation for IEEE 802.11 Data Communications,” *IEEE Transactions on Vehicular Technology*, vol. 61, no. 5, pp. 2158–2171, 2012.

[58] T. Kupka, P. Halvorsen, and C. Griwodz, “An Evaluation of Live Adaptive HTTP Segment Streaming Request Strategies,” in *Proceedings of the IEEE 36th Conference on Local Computer Networks (LCN 2011)*, (Bonn, Germany), 2011.

[59] C. C. Wüst and W. F. J. Verhaegh, “Quality Control for Scalable Media Processing Applications,” *Journal of Scheduling*, vol. 7, no. 2, pp. 105–117, 2004.

[60] D. Jarnikov and T. Ozcelebi, “Client Intelligence for Adaptive Streaming Solutions,” in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2010)*, (Singapore), 2010.

[61] C. Liu, I. Bouazizi, and M. Gabbouj, “Rate Adaptation for Adaptive HTTP Streaming,” in *Proceedings of the 2nd Annual ACM Conference on Multimedia Systems (MMSys 2011)*, (Santa Clara, CA, USA), 2011.

[62] K. Miller, E. Quacchio, G. Gennari, and A. Wolisz, “Adaptation Algorithm for Adaptive Streaming over HTTP,” in *Proceedings of the 19th International Packet Video Workshop (PV 2012)*, (Munich, Germany), 2012.

[63] C. Müller, S. Lederer, and C. Timmerer, “An Evaluation of Dynamic Adaptive Streaming over HTTP in Vehicular Environments,” in *Proceedings of the 4th Workshop on Mobile Video (MoVID 2012)*, (Chapel Hill, NC, USA), 2012.

[64] S. Oechsner, T. Zinner, J. Prokopetz, and T. Hoßfeld, “Supporting Scalable Video Codecs in a P2P Video-on-Demand Streaming System,” in *Proceedings of the 21th ITC Specialist Seminar on Multimedia Applications - Traffic, Performance and QoE (ITC-SS21)*, (Miyazaki, Japan), 2010.

[65] C. Sieber, T. Hoßfeld, T. Zinner, P. Tran-Gia, and C. Timmerer, “Implementation and User-centric Comparison of a Novel Adaptation Logic for DASH with SVC,” in *Proceedings of the IFIP/IEEE International Workshop on Quality of Experience Centric Management (QCMan 2013)*, (Ghent, Belgium), 2013.

[66] B. Wang, J. Kurose, P. Shenoy, and D. Towsley, “Multimedia Streaming via TCP: An Analytic Performance Study,” *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 4, no. 2, p. 16:116:22, 2008.

[67] H. F. Nielsen, J. Gettys, A. Baird-Smith, E. Prud'hommeaux, H. W. Lie, and C. Lilley, "Network Performance Effects of HTTP/1.1, CSS1, and PNG," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 4, p. 155166, 1997.

[68] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware Resource Allocation and Packet Scheduling for Video Transmission over Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 749–759, 2007.

[69] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the Impact of Video Quality on User Engagement," in *Proceedings of the ACM SIGCOMM*, (Toronto, ON, Canada), 2011.

[70] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "A Quest for an Internet Video Quality-of-experience Metric," in *Proceedings of the 11th ACM Workshop on Hot Topics in Networks (HotNets-XI)*, (Redmond, WA, USA), 2012.

[71] N. Cranley, P. Perry, and L. Murphy, "User Perception of Adapting Video Quality," *International Journal of Human-Computer Studies*, vol. 64, no. 8, pp. 637–647, 2006.

[72] O. Abboud, T. Zinner, K. Pussep, S. Al-Sabea, and R. Steinmetz, "On the Impact of Quality Adaptation in SVC-based P2P Video-on-Demand Systems," in *Proceedings of the 2nd Annual ACM Conference on Multimedia Systems (MMSys 2011)*, (Santa Clara, CA, USA), 2011.

[73] R. Kuschnig, I. Kofler, and H. Hellwagner, "An Evaluation of TCP-based Rate-control Algorithms for Adaptive Internet Streaming of H.264/SVC," in *Proceedings of the 1st Annual ACM Conference on Multimedia Systems (MMSys 2010)*, (Phoenix, AZ, USA), 2010.

[74] P. Ni, R. Eg, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Flicker Effects in Adaptive Video Streaming to Handheld Devices," in *Proceedings of the 19th ACM International Conference on Multimedia (MM 2011)*, (Scottsdale, AZ, USA), 2011.

[75] B. Lewcio, B. Belmudez, A. Mehmood, M. Wältermann, and S. Möller, "Video Quality in Next Generation Mobile Networks – Perception of Time-varying Transmission," in *Proceedings of the 2011 IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR 2011)*, (Naples, FL, USA), 2011.

[76] M. Grafl and C. Timmerer, "Representation Switch Smoothing for Adaptive HTTP Streaming," in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013)*, (Vienna, Austria), 2013.

[77] M. Zink, J. Schmitt, and R. Steinmetz, "Layer-encoded Video in Scalable Adaptive Streaming," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 75–84, 2005.

[78] Y. Pitrey, U. Engelke, M. Barkowsky, R. Pépion, and P. Le Callet, "Subjective Quality of SVC-coded Videos with Different Error-patterns Concealed Using Spatial Scalability," in *Proceedings of the 3rd European Workshop on Visual Information Processing (EUVIP 2011)*, (Paris, France), 2011.

[79] C. Alberti, D. Renzi, C. Timmerer, C. Mueller, S. Lederer, S. Battista, and M. Mattavelli, "Automated QoE Evaluation of Dynamic Adaptive Streaming over HTTP," in *Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX 2013)*, (Klagenfurt, Austria), 2013.

[80] N. Bouten, J. Famaey, S. Latré, R. Huysegems, B. D. Vleeschauwer, W. V. Leekwijck, and F. D. Turck, "QoE Optimization Through In-network Quality Adaptation for HTTP Adaptive Streaming," in *Proceedings of the 8th International Conference on Network and Service Management (CNSM 2012)*, (Las Vegas, NV, USA), 2012.

[81] Blender Foundation, "Sintel, the Durian Open Movie Project," 2010.

[82] Y.-F. Ou, Y. Zhou, and Y. Wang, "Perceptual Quality of Video with Frame Rate Variation: A Subjective Study," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010)*, (Dallas, TX, USA), 2010.

[83] F. Pereira, "Sensations, Perceptions and Emotions Towards Quality of Experience Evaluation for Consumer Electronics Video Adaptations," in *Proceedings of the 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2005)*, (Scottsdale, AZ, USA), 2005.

[84] G. Ghinea and J. P. Thomas, "QoS Impact on User Perception and Understanding of Multimedia Video Clips," in *Proceedings of the 6th ACM International Conference on Multimedia*, (Bristol, UK), 1998.

[85] R. K. Rajendran, M. Van Der Schaar, and S.-F. Chang, "FGS+: Optimizing the Joint SNR-temporal Video Quality in MPEG-4 Fine Grained Scalable Coding," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2002)*, (Scottsdale, AZ, USA), 2002.

[86] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 882–893, 2011.

[87] M.-N. Garcia and A. Raake, "Parametric Packet-layer Video Quality Model for IPTV," in *Proceedings of the 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA 2010)*, (Kuala Lumpur, Malaysia), 2010.

[88] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, "To Pool or not to Pool: A Comparison of Temporal Pooling Methods for HTTP Adaptive Video Streaming," in *Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX 2013)*, (Klagenfurt, Austria), 2013.

[89] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional Perceptual Quality Assessment for Low Bit-rate Videos," *IEEE Transactions on Multimedia*, vol. 10, no. 7, pp. 1316–1324, 2008.

[90] K. Yamagishi and T. Hayashi, "Parametric Packet-layer Model for Monitoring Video Quality of IPTV Services," in *Proceedings of the IEEE International Conference on Communications (ICC 2008)*, (Beijing, China), 2008.

[91] H. Knoche, J. D. Mccarthy, and M. A. Sasse, "How Low Can You Go? The Effect of Low Resolutions on Shot Types in Mobile TV," *Multimedia Tools and Applications*, vol. 36, no. 1-2, pp. 145–166, 2008.

[92] L. Janowski and P. Romaniak, "QoE as a Function of Frame Rate and Resolution Changes," in *Proceedings of the 3rd International Workshop on Future Multimedia Networking (FMN 2010)*, (Krakow, Poland), 2010.

[93] P. Le Callet, S. Péchard, S. Tourancheau, A. Ninassi, and D. Barba, "Towards the Next Generation of Video and Image Quality Metrics: Impact of Display, Resolution, Contents and Visual Attention in Subjective Assessment," in *Proceedings of the 2nd International Workshop on Image Media Quality and its Applications (IMQA 2007)*, (Chiba, Japan), 2007.

[94] R. R. Pastrana-Vidal, J.-C. Gicquel, C. Colomes, and H. Cherifi, "Frame Dropping Effects on User Quality Perception," in *Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004)*, (Lisbon, Portugal), 2004.

[95] S. Jumisko-Pyykkö and J. Häkkinen, "Evaluation of Subjective Video Quality of Mobile Devices," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, (Singapore), 2005.

[96] S. Winkler and C. Faller, "Maximizing Audiovisual Quality at Low Bitrates," in *Proceedings of the 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2005)*, (Scottsdale, AZ, USA), 2005.

[97] O. Verscheure, P. Frossard, and M. Hamdi, "User-oriented QoS Analysis in MPEG-2 Video Delivery," *Real-Time Imaging*, vol. 5, no. 5, pp. 305–314, 1999.

[98] J. De Vriendt, D. De Vleeschauwer, and D. Robinson, "Model for Estimating QoE of Video Delivered Using HTTP Adaptive Streaming," in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, (Ghent, Belgium), 2013.

[99] H. Knoche, J. D. McCarthy, and M. A. Sasse, "Can Small be Beautiful?: Assessing Image Resolution Requirements for Mobile TV," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, (Singapore), 2005.

[100] R. T. Apteker, J. A. Fisher, V. S. Kisimov, and H. Neishlos, "Video Acceptability and Frame Rate," *IEEE MultiMedia*, vol. 2, no. 3, pp. 32–40, 1995.

[101] D. Wang, F. Speranza, A. Vincent, T. Martin, and P. Blanchfield, "Toward Optimal Rate Control: A Study of the Impact of Spatial Resolution, Frame Rate, and Quantization on Subjective Video Quality and Bit Rate," in *Proceedings of the SPIE Visual Communications and Image Processing (VCIP 2003)*, (Lugano, Switzerland), 2003.

[102] T. Hayashi, S. Yamasaki, N. Morita, H. Aida, M. Takeichi, and N. Doi, "Effects of IP Packet Loss and Picture Frame Reduction on MPEG1 Subjective Quality," in *Proceedings of the 3rd IEEE Workshop on Multimedia Signal Processing*, (Copenhagen, Denmark), 1999.

[103] J. Chen and J. Thropp, "Review of Low Frame Rate Effects on Human Performance," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 37, pp. 1063–1076, 2007.

[104] M. Masry, S. S. Hemami, W. M. Osberger, and A. M. Rohaly, "Subjective Quality Evaluation of Low-bit-rate Video," in *Proceedings of the SPIE Photonics West 2001: Electronic Imaging*, (San Jose, CA, USA), 2001.

[105] Y. Wang, S.-F. Chang, and A. C. Loui, "Subjective Preference of Spatio-temporal Rate in Video Adaptation Using Multi-dimensional Scalable Coding," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2004)*, (Taipei, Taiwan), 2004.

[106] J. D. McCarthy, M. A. Sasse, and D. Miras, "Sharp or Smooth?: Comparing the Effects of Quantization vs. Frame Rate for Streamed Video," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2004)*, (Vienna, Austria), 2004.

[107] J. Korhonen, U. Reiter, and J. You, "Subjective Comparison of Temporal and Quality Scalability," in *Proceedings of the 3rd International Workshop on Quality of Multimedia Experience (QoMEX 2011)*, (Mechelen, Belgium), 2011.

[108] N. Van den Ende, H. De Hesselle, and L. Meesters, "Towards Content-aware Coding: User Study," in *Proceedings of the 5th European Interactive TV Conference (EuroITV 2007)*, (Amsterdam, The Netherlands), 2007.

[109] S. Akhshabi, A. C. Begen, and C. Dovrolis, "An Experimental Evaluation of Rate-adaptation Algorithms in Adaptive Streaming over HTTP," in *Proceedings of the 2nd Annual ACM Conference on Multimedia Systems (MMSys 2011)*, (Santa Clara, CA, USA), 2011.

[110] R. Houdaille and S. Gouache, "Shaping HTTP Adaptive Streams for a Better User Experience," in *Proceedings of the 3rd Annual ACM Conference on Multimedia Systems (MMSys 2012)*, (Chapel Hill, NC, USA), 2012.

[111] J. Esteban, S. A. Benno, A. Beck, Y. Guo, V. Hilt, and I. Rimac, "Interactions Between HTTP Adaptive Streaming and TCP," in *Proceedings of the 22nd ACM Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV 2012)*, (Toronto, ON, Canada), 2012.

[112] A. Mansy, B. Ver Steeg, and M. Ammar, "SABRE: A Client Based Technique for Mitigating the Buffer Bloat Effect of Adaptive Video Flows," in *Proceedings of the 4th Annual ACM Conference on Multimedia Systems (MMSys 2013)*, (Oslo, Norway), 2013.