Perceptual Quality of HTTP Adaptive Streaming Strategies: Cross-Experimental Analysis of Multi-Laboratory and Crowdsourced Subjective Studies

Samira Tavakoli[a], Sebastian Egger[b], Michael Seufert[c] Raimund Schatz[b], Kjell Brunnström[d], and Narciso García[a]

[a]Universidad Politécnica de Madrid, Spain, [b]Austrian Institute of Technology, Austria, [c]University of Würzburg, Germany, [d]Acreo Swedish ICT AB, Sweden

Abstract—Today's packet-switched networks are subject to bandwidth fluctuations that cause for degradation of user experience of multimedia services. In order to cope with this problem, HTTP Adaptive Streaming (HAS) has been proposed in recent years as a video delivery solution for the future Internet and being adopted by an increasing number of streaming services such as Netflix and Youtube. HAS enables service providers to improve users' Quality of Experience (QoE) and network resource utilization by adapting the quality of the video stream to the current network conditions. However, the resulting time-varying video quality caused by adaptation introduces a new type of impairment and thus novel QoE research challenges. Despite of various recent attempts to investigate these challenges, many fundamental questions regarding HAS perceptual performance are still open. In this paper, the QoE impact of different technical adaptation parameters including chunk length, switching amplitude, switching frequency and temporal recency are investigated. In addition, the influence of content on perceptual quality of these parameters is analyzed. To this end, a large number of adaptation scenarios have been subjectively evaluated in four laboratory experiments and one crowdsourcing study. Statistical analysis of the combined dataset reveals results that partly contradict widely held assumptions and provide novel insights in perceptual quality of adapted video sequences, e.g. interaction effects between quality switching direction (up/down) and switching strategy (smooth/abrupt). The large variety of experimental configurations across different studies ensures the consistency and external validity of the presented results that can be utilized for enhancing the perceptual performance of adaptive streaming services.

Index Terms—HTTP adaptive streaming, QoE, perceptual quality, laboratory study, subjective study, crowdsourcing, cross-experimental study

I. Introduction

ACCORDING to a recently published traffic forecasting study [1], global Internet video traffic accounted for 57% of all consumer traffic in 2012 and will even rise to 69% in 2017. Two thirds of this traffic will be delivered through HTTP-based streaming services like Netflix or comparable online platforms with TCP as underlying transport protocol. A common characteristic of all aforementioned multimedia services is that their service are run 'best effort' over the

S. Tavakoli and N.García are with the Grupo de Tratamiento de Imágenes, E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain (email: sat/narciso@gti.ssr.upm.es).

S. Egger and R. Schatz are with the department of Technology Experience, AIT Austrian Institute of Technology, Donau-City Str. 1, 1220 Vienna, Austria (email: sebastian.egger/raimund.schatz@ait.ac.at).

M. Seufert is with the Chair of Communication Networks, University of Würzburg, D-97074, Würzburg, Germany (email: seufert@informatik.uni-wuerzburg.de).

K. Brunnström is with Acreo Swedish ICT AB, Box 1070, SE-164 25, Kista, Sweden (email: kjell.brunnstrom@acreo.se).

Internet. However, changing channel conditions in today's wireless network is inevitable. Although packet loss can be prevented by the use of TCP, network congestion and different radio conditions lead to varying network bandwidth. If the available bandwidth falls below the video bitrate, long initial delay or stalling (i.e., the interruption of playback due to empty playout buffers) will eventually occur which severely deteriorates the users' Quality of Experience (QoE) [2].

In order to cope with this problem, several Internet video services have switched to HTTP Adaptive Streaming (HAS) as their default delivery method. HAS requires the video to be available on the server in an adaptation set which consists of video representations of different bitrates. Each representation has been split into small chunks each containing a few seconds of the playtime. After the buffer was filled up with the video in an initial bitrate, the HAS client starts the video playout. Subsequently, it measures the current bandwidth and/or buffer status and requests the next chunks of the video in an appropriate bitrate, such that stalling is avoided and available bandwidth is utilized best.

Apart from the benefit of dynamically adapting the quality to the available bandwidth, employing HAS provides further advantages compared to the classical streaming approaches. For instance, offering multiple video bitrates enables service providers to adapt the delivered video to users with different demands, devices and network accessibility. Furthermore, based on the available video quality, different pricing schemes and service levels can be offered to the customers. These advantages have made HAS as a popular approach for distribution of multimedia over the future Internet. Nevertheless, using this method also results in a new type of degradation in terms of users' QoE: time-varying video quality. Despite the several studies in this area, there are many open questions concerning adaptation parameters, switching behaviors and related factors and their influence on user's perception.

This paper targets a set of these open research questions (RQ) in order to identify the influence parameters on perceptual quality of switching strategies:

RQ1: With regard to the up- and down-switching, what is the QoE gain of smooth switching over abrupt switching? Does chunk length has any effect on QoE? What is the impact of the switching amplitude (defined as bitrate difference between the current and target quality level)

RQ2: Does more frequent quality switching result in worse QoE than less frequent switching? What is the influence of the last quality level in terms of recency effects?

RQ3: What is the impact of content type on the QoE of different switching strategies?

RQ4: Is it better to switch the quality or stay at a constant low quality level?

In order to answer these questions, we gathered a large dataset spanning a wide range of adaptation settings evaluated in multi-laboratory and crowdsourcing environments, followed by an extensive cross-experimental analysis.

The remainder of the paper is organized as follows. Section II discusses the current state of research in HAS QoE domain specially related to the targeted RQs. The description of the dataset and experimental setups are presented in Section III. Results of the cross-experimental analysis are presented in Section IV and discussed in Section V. The paper's conclusions are drawn in Section VI.

II. RELATED WORK

Recent work on performance analysis of HAS solutions can be categorized into technical and perception based quality assessment. Technical analysis such as [3] mainly concentrates on theoretical analysis of quality switching strategies aiming to optimize the bandwidth utilization or other network related parameters. In contrast, perception based analysis is more user-centric aiming to analyze the QoE impact of different adaptation strategies and related parameters or combinations of different adaptation dimensions and their resulting QoE performance. In this paper we will focus on perception based analysis.

Influence of the adaptation related parameters on viewers' perceptual quality have been already shown by previous studies. The study presented in [4] claims that "quality switches are perceived as a degradation itself". However, quality switches are often inevitable due to changing bandwidth conditions. In this situation, with a given adaptation set (where each representation could differ in terms of encoding quantization setting, spatial/temporal resolution, and/or audio bitrate), two key aspects must be considered to design a perceptually efficient adaptation behavior: *frequency* and *amplitude* of the quality switching events.

Concerning switching *frequency*, different factors can influence on end-user QoE such as: 1) the number of switches in each adaptation event due to using short/long chunk length, 2) the length of adaptation event (which also varies because of using short/long chunk length), and 3) the number of adaptation events occurring during the playback. In that respect, [5] investigated the adaptation of image quality for layer-encoded videos. According to this study, adaptation frequency should be kept as small as possible. From another side, the experimental results presented in [6] showed that higher switching frequencies are not penalized in terms of QoE if sufficient time is spent on a high quality level. The study presented in [7] which considered relatively high quality levels (from 2 to 8 Mbps) shows that quality oscillations are hardly perceptible when quality differences between layers are not pronounced.

With respect to switching *amplitude*, two problems can arise: 1) if there is any difference between smooth (i.e. stepwise change from the current to the target quality level) and abrupt switching, and 2) what is the impact of the bitrate

difference between the current and target quality level video QoE. In addition, the impact of these factors in up-switching (increasing the quality) and down-switching (decreasing the quality) should be considered. The results presented in [5], [8], [9] showed that inserting intermediate levels between quality drops was favorable for the perceived quality, compared to switching directly to the target quality level. However, the results presented in [10] showed that, although down-switching is generally considered annoying, abrupt up-switching might even increase the QoE as users might be happy to notice the visual improvement.

Even if optimal adaptation strategies in terms of switching amplitude and frequency could be found, one fundamental question remains open: is it better to switch the quality level or would it be better to try maintaining a certain constant (even low) quality level in order to minimize the impairment caused by the switching itself? A number of studies have addressed the *"to switch or not to switch" dilemma* [6], [11]–[13], resulting that constant quality is usually preferred to time-varying quality. In particular: constant (even lower) quality is preferred to decreasing the quality [11]; constant or nearly constant quality is preferable to frequently varying quality even if the mean quality is lower [6], [12]. However, some exceptions that contradict the aforementioned results must be considered here as well: more than one study pointed out that "if the constant quality is too low, adaptation to the better quality could be preferred" [6], [14].

With regard to perceptual quality of adaptation strategies, the impact of user-related psychological factors should not be neglected. This includes the *memory and recency effects* of the human behavioral response while continuously evaluating the time-varying video quality [15]. Recency refers to the human brain's preference to attach higher importance to recent stimuli. Therefore, the duration of presenting a specific quality level has influence on the occurrence of recency effects. In that respect, [16] investigated the recency effect of the last quality level in the adaptation event in addition to the recency time, which was defined as the time since the last quality adaptation. Their results showed that ending with low or high quality, in addition to the time duration spent experiencing the high quality do not have significant QoE impact. However, this finding might be due to the specific characteristic of the switching patterns the authors used in their study.

Previous studies have addressed the influence of contextual and objective characteristics of the content on viewers' quality perception [17]–[19]. Therefore, the question arises whether perceptual effect of the aforementioned factors would be similar for different *types of video content*. In this regard the result of [6], [12] showed that the effect of different switching dimensions (specifically spatial and temporal quality switching) varies depending on the content type, even for comparable switching amplitudes: "while it is difficult to spot quality oscillations when there are frequent scene changes, they are more noticeable in steady shots and in strong edges". Considering the influence of objective characteristics of the content in terms of spatial and temporal complexities, the result in [20] showed that "the perceived quality of adaptation when playing back a content with high spatial and low

temporal amount of activity is significantly lower compared to the other content types".

Summarizing, although many research questions on QoE of adaptation have been already tackled in previous research, still many of these questions have remained open or not appropriately resolved due to (i) a limited number of tests conducted to address a question, (ii) shortcomings of the reported studies (such as missing information in the respective publication), or evident limitations in the considered set of test conditions, or (iii) contradictory outcomes with respect to identical research questions [21]. The targeted research questions in this study (cf. RQs in Section I) is motivated by this shortcoming in order to determine the perceptual influence of video adaptation parameters within HAS system.

III. STUDY DESCRIPTION

Table I summarizes the information about the subjective experiments that constitute the dataset under study. In the following subsections, first we describe the technical settings and test materials used for implementing the different adaptation scenarios, followed by describing the test setups and procedures of individual experiments.

TABLE I: The subjective experiments behind the dataset. Detailed information is provided in Subsection III-B.

#	Experiment	Methodology	#Subj	Ref
1	Acreo-Laboratory	ACR	20	
2	UPM-Laboratory	UPM method-Audio	21	LabI
3		UPM method-NoAudio	22	
4	FTW-Laboratory	ACR	34	LabII
5	FTW-Crowdsourcing	ACR	576	CSIII

A. Test materials and conditions

In order to investigate research questions RQ1-4, an exhaustive set of video sequences with defined switching patterns was needed. The different type of switching patterns used throughout the studies are presented in Figure 1.

For the comparison of abrupt versus smooth switching strategies, the two video sequences to be compared against each other shared the same lower (Q_i) and higher quality level (Q_{i+k}), with i indicating the lower quality level of the respective sequence and k indicating the number of quality level change for reaching to the higher one. In the case of abrupt switching, the quality change occurred in the middle of sequence duration, whereas for the smooth switching after every chunk one quality change took place until reaching to the target level. Since human perception of quality switching can be different with respect to the switching direction, abrupt and smooth switching test sequences were constructed for both up- and down-switching directions as demonstrated in the left and right side of the upper portion of Figure 1. By considering different quality profiles for the current and target quality levels different switching amplitudes were also provided.

For the comparison of high and low frequent switching strategies, the quality was oscillated between Q_i and Q_{i+k} as depicted in the lower portion of Figure 1. In order to study the

recency effect due to the last presented quality level, sequences starting and ending with Q_i as well as sequences starting and ending with Q_{i+k} were constructed for evaluation. Special care was taken to keep the time on the highest and lowest quality level constant across the high/low frequent switching sequences in order to eliminate the impact of different dwell times on high/low quality levels as reported in [16].

To construct these scenarios, different representation sets in H.264/AVC format were prepared in each experiments that their encoding parameters are presented in Table II. For LabI and LabII experiments (cf. Table I) the quality was adapted on the video compression dimension by varying the quantization parameter (QP), whereas for CSIII study the spatial video quality was varied (cf. *QP* and *Res* in Table II).

For LabI studies, the encoding setting used in practice for the living-room platform was considered. Four quality levels from 600 kbps to 5000 kbps were produced using Rhozet Carbon Coder. Two chunk lengths (2 sec and 10 sec) were chosen for up- and down-switching the quality in abrupt and smooth way. In total, 8 adaptation scenarios (cf. LabI in Table III) or Hypothetical Reference Circuits (HRC) were considered to provide the Processed Video Sequences (PVS).

For LabII and CSIII studies, mobile access scenario was selected, hence the chosen video bitrates ranged from 128 kbps up to 2400 kbps. The quality representations were prepared using FFmpeg encoder. The chunk lengths (2 sec, 4 sec, 5 sec and 10 sec) were chosen to be inline with current HAS solutions [22] and to yield integer multiplies for the envisaged video sequence lengths. Overall, 12 up- and down-switching patterns and 20 quality oscillation scenarios (HRC) were considered for evaluation (cf. LabII and CSIII in Table III and Table IV).

An important aspect for facilitating comparison of large numbers of quality profiles is immediate visibility of the switching characteristics (e.g. smooth switching, spatial resolution changed, chunk length, etc.) used for the scenario under investigation. Therefore, we derived a coding scheme that allows to identify the underlying switching characteristics of a video sequence in a simple way. An example code is shown in Figure 2. The first character of the code allows to distinguish between the different quality profiles to be compared (D/I for decreasing/increasing scenarios, O for quality oscillation). Then, the second character denotes which switching strategy
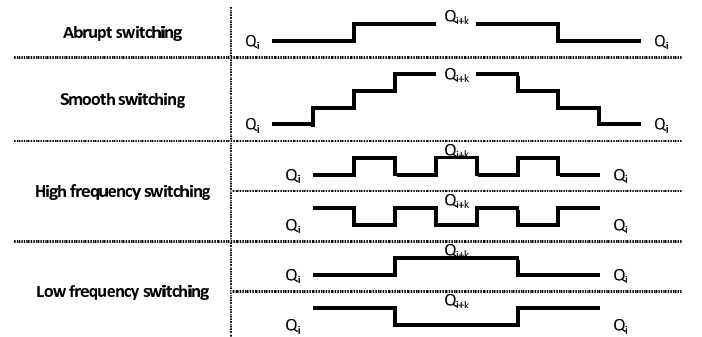

Fig. 1: Quality profiles used throughout different studies. Q_i and Q_{i+k} denote the quality levels (cf. Table II) used within the respective profile.

TABLE II: Video encoding parameters for the chunks used in different studies. According to the encoder setting used in LabI, the 'adaptive' QP (between 22 to 32) based on the content complexity was assigned. For Lab II, the QPs were chosen in accordance with the target bitrate.

Ref	Level	QP	FR(fps)	Res	Target BR
LabI	4cI	adapt	24	1280x720	5000
	3cI	adapt	24	1280x720	3000
	2cI	adapt	24	1280x720	1000
	1cI	adapt	24	1280x720	600
LabII	3cII	33	25	1280x720	1400
	2cII	40	25	1280x720	700
	1cII	47	25	1280x720	350
CSIII	5sIII	26	25	1280x720	2400
	4sIII	26	25	854x480	1250
	3sIII	26	25	640x360	800
	2sIII	26	25	426x240	430
	1sIII	26	25	256x144	195

TABLE III: Quality profiles used for the comparison between abrupt and smooth switching when increasing and decreasing the video quality (RQ1). *Chnk* and *Dur* are in sec.

Status	Behavior	Chnk	Dur	Q-Level	Code	Ref
Decreasing	Smooth	2	14	4cI-1cI	DS2cI41	LabI
		10	40		DS10cI41	
		10	20	3cII-1cII	DS10cII31	LabII
		5	20	5sIII-1sIII	DS5sIII51	CSIII
		5	20	5sIII-2sIII	DS5sIII52	CSIII
	Abrupt	2	14	4cI-1cII	DA2cI41	LabI
		10	40		DA10cI41	
		10	20	3cII-1cII	DA10cII31	LabII
		5	20	5sIII-1sIII	DA5sIII51	CSIII
		5	20	5sIII-2sIII	DA5sIII52	CSIII
Increasing	Smooth	2	14	1cI-4cI	IS2cI14	LabI
		10	40		IS10cI14	
		10	20	1cII-3cII	IS10cII13	LabII
		5	20	1sIII-5sIII	IS5sIII15	CSIII
		5	20	2sIII-5sIII	IS5sIII25	CSIII
	Abrupt	2	14	1cI-4cI	IA2cI14	LabI
		10	40		IA10cI14	
		10	20	1cII-3cII	IA10cII13	LabII
		5	20	1sIII-5sIII	IA5sIII15	CSIII
		5	20	2sIII-5sIII	IA5sIII25	CSIII

TABLE IV: Quality profiles constructed for comparison between high and low frequent quality oscillation and to identify perceptual differences for starting and ending with high/low quality in respect to recency effects (RQ2). *Chnk* and *Dur* are in sec.

Freq	Chnk	Dur	#Switch	Q-Level	Code	Ref
H	4	28	2	(5-3-5)sIII	OH4sIII53	CSIII
L		28	5		OL4sIII53	
H	4	28	2	(5-2-5)sIII	OH4sIII52	
L		28	5		OL4sIII52	
H	4	28	2	(5-1-5)sIII	OH4sIII51	
L		28	5		OL4sIII51	
H	10	70	2	(3-2-3)cII	OH10cII32	LabII
L		70	5		OL10cII32	
H	10	70	2	(3-1-3)cII	OH10cII31	
L		70	5		OL10cII31	
H	4	28	2	(3-5-3)sIII	OH4sIII35	CSIII
L		28	5		OL4sIII35	
H	4	28	2	(2-5-2)sIII	OH4sIII25	
L		28	5		OL4sIII25	
H	4	28	2	(1-5-1)sIII	OH4sIII15	
L		28	5		OL4sIII15	
H	10	70	2	(2-3-2)cII	OH10cII23	LabII
L		70	5		OL10cII23	
H	10	70	2	(1-3-1)cII	OH10cII13	
L		70	5		OL10cII13	
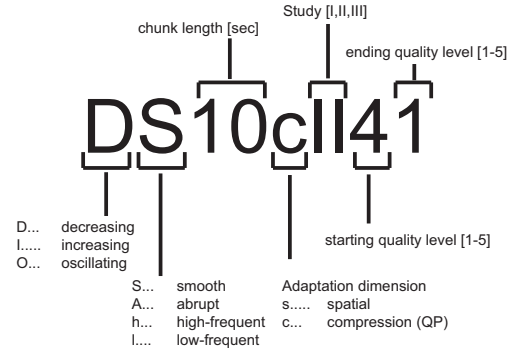
Fig. 2: Description of the code structure for identifying characteristics of the adaptation test condition.

is applied (S/A for smooth/abrupt and L/H for low/high frequent). The third character reflects the chunk length in seconds and digit character describes the encoding dimension utilized (c/s for compression/spatial). The fifth character then denotes which adaptation set used throughout the sequence (cf. *Ref* in Table II). Finally, the sixth and seventh characters express the quality at the beginning and the end of the sequence in order(ranging from [1-5] depending on the levels available in the adaptation set as depicted in Table II column *Level*). The codes constructed by this scheme facilitate to map the individual results presented in the plots (see Section IV) to the underlying adaptation characteristics.

Table III represents the adaptation profiles constructed for the comparison between smooth and abrupt switching strategies. The quality levels described in column *Q-Level* stem from the respective adaptation set in Table II, whereas column *Code* contains the corresponding code according to the coding scheme introduced above. In the first two lines it can be seen, that both adaptation profiles contain similar quality change

(from 4cI to 1cI) but do contain different chunk lengths (which in turn results in different durations; cf. column *Dur*). In the code this is reflected by the third digit In order to compare the perceptual effect of different parameters in up- and down-switching use cases, test scenarios are separated in increasing and decreasing groups with otherwise identical switching characteristics.

Table IV represents the profiles for the comparison between high and low frequent switching strategies. For low frequent profiles only two quality switches took place whereas for high frequent switching five quality changes were performed. In addition to studying the low vs. high frequent switching, studying the recency effect was taken into account. Therefore, all adaptation profiles were created in such a way to start and also end on high quality level (e.g. 5-3-5) as well as on the low quality level (e.g. 3-5-3). Furthermore, studying the perceptual effect of switching amplitude was also considered to account and therefore varied accordingly (e.g. 5-3-5 vs. 5-1-5).

Several authors (e.g. [23]) reported the influence of content on the perceived quality which might hence influence the per-

No Impairment **0**	PVS_1	VS_1 **1**	...	...	PVS_m	VS_m **m**

Fig. 3: UPM test methodology. PVS and VS stand on 'processed video sequence' and 'voting segment' in order. '0' printed in the corner of the first segment's frames has no degradation indicating the start of the test. Randomized order of 6 min long test sequences were presented to the subjects.

ception of quality switches as well. Taking this into account, a large set of video content type was included in the studies that their characteristics are summarized in Table V.

B. Details of studies

1) LabI: Targeting RQ1, RQ3 and RQ4 in addition to investigating the impact of test methodology on observers' assessments, three experiments were conducted in different environments, which have been described in [20], and denoted as LabI in this study. The PVSs were identical in all experiments but evaluated through different testing approaches.

Two of the experiments were carried out in the laboratory of Universidad Politécnica de Madrid (UPM), Spain, by employing the methodology that has been developed to evaluate the quality degradation in long sequences [25]. Figure 3 shows an example of the test sequence used in these experiments. Using this method, subjects continuously viewed the 6 min long sequences including the sequential PVS with intervention of a non-impaired segment which was considered as Voting Segment (VS). During VS, that its frames had a printed number indicating the former PVS number, the test subjects were asked to rate the overall quality of the previous PVS considering the Absolute Category Rating (ACR) 5-graded scale [24]. To study the influence of audio presence on evaluation of test scenarios, one of the experiments was done by presenting only the video stimulus of the test sequence and the other one in the presence of audio (called UPM-NoAudio and UPM-Audio).

The third experiment was conducted in Acreo Swedish ICT's laboratory, Sweden, in which the randomized order of the PVSs (cf. Figure 3) were presented one after another following ACR methodology and using the Video Quality Experts Group (VQEG) player [26]. The observers were rated the quality of the PVSs using the same scales as in the UPM experiments. In order to allow for cross-lab comparison, the ambient and all the hardware and software in Acreo were adjusted similar to the UPM complying with the ITU-R Rec. BT.500-11 [27]. A 46" Hundai S465D display was used with the native resolution of 1920x1080 . The viewing distance was set to four times the display height. The TVs peak white luminance was 177cd/m2 and the illumination level of the room was 20 lux.

Prior to the test session, the test subjects were screened for visual acuity and color vision . Later on, the test instruction and the rating scale presented in the observers' native language (mainly Spanish in UPM, Swedish in Acreo, and English for the international observers of each experiment) were given. After reading the instruction, a training session was conducted to familiarize the observers with the test procedure.

The observers came from different countries but mostly from

Spain and Sweden. About 70% of them had computer science background (student, employed) and around 30% of them had already participated in a subjective test. After post-screening of the subjective data in accordance with the latest recommendations from Video Quality Experts Group [28], the scores of 21 observers from UPM-Audio (6 female and 15 male, age from 27 to 50) , 22 observers from UPM-NoAudio (5 female and 17 male, age from 24 to 54) and 23 observers from Acreo experiment (7 female and 16 male, age from 18 to 68) were considered for evaluation. The number of remaining observers was in accordance to ITU-R recommendation [27].

After collecting the subjective scores for the perceptual quality of the PVSs from each experiment, the Mean Opinion Scores (MOS) and 95% confidence interval (CI) of their statistical distribution were calculated. In the comparison between the MOS obtained from UPM-Audio and UPM-NoAudio experiments (effect of the audio presence), as well as the combined UPM data and Acreo data (effect of the testing methodology), no significant difference was found according to the ANalysis Of VAriance (ANOVA) test. That means that the main effect of the audio presence and testing methodology was not significant (ANOVA result for each effect in order: $p = 0.63$ and $p = 0.31$). In addition, not any PVS was significantly differently rated in the three experiments. Based on this result, the three datasets were merged into one set to further analysis the targeted HAS technical parameters [20].

2) LabII and CSIII: The comprehensive technical report of these studies has been provided in [29]. Besides using different evaluation methodology, different adaptation sets were also used in two studies: In LabII, the video quality was varied along the compression dimension , whereas in CSIII the spatial resolution of the videos was varied. The quality profiles used are reported in Table III and Table IV. Both studies utilized all four up- and down- switching profiles shown in Figure 1 in order to answer RQ1 and RQ2. Data for RQ3 was only collected in CSIII and data for RQ4 only in LabII.

The laboratory study (LabII) was executed in Telecommunications Research Center Vienna (FTW) in accordance to ITU-R Rec. BT.500-11 [27] with the following differences: As we were targeting an online browser based scenario for HAS, the test persons were watching the videos on 15.4" laptop screen (HP Elitebook 8530W and 1680x1050 screen resolution) in a distance of approximately three times the screen height (65 cm). We also chose the computer based watching scenario to be able to compare the results to the crowdsourcing results as workers there only use computer screens rather then TV screens. For the setting of the screen, a peak luminance of $194 \, \text{cd/m}^2$ was set and the environmental luminance on the screen was set to $85 \, \text{lux}$. As rating scale, an ACR 5-graded scale was used. Presentation of the sequences with the subsequent rating procedure was performed according to [27]. Before beginning of the test, subjects were screened for the visual acuity. The majority of the test participants were from Austria, 40% university subjects from different area and the rest were employed people with different occupations. After checking the reliability of provided scores according to [27], scores from 34 subjects (18 female and 14 male, mean/median age 38.1/32 respectively) were used for further analysis.

TABLE V: Characteristics of the source video content used in different studies. In columns *SI* and *TI*, the spatial and temporal information of the content as formulated in [24] are presented and in *Original Format*, the original resolution and frame rate.

Ref	Type-Code	SI	TI	Original Format	Description
LabI	Movie1	56.41	47.90	1080p/24fps	Action, adventure, with some scene in smooth motion, some with group of walking people, some with camera panning
	Movie2	48	70.65	1080p/24fps	Drama, romance, mostly with the smooth motion in the static background, some scene with group of dancing people in bright ambient
	Movie3	53,26	71,40	1080p/ 24fps	Action, Si Fi, with the rapid changes in some sequences, cloudy atmosphere
	Sport1	59.10	38.45	1080p/50fps	Soccer, average motion, wide angle camera sequences, uniform camera panning
	Doc	41.77	67.49	1080p/50fps	Sport documentary, mostly with handheld shooting camera
	News	55.57	59.96	1080p/50fps	Spanish news, some scenes with static shooting camera with one/two standing/sitting people; some outdoor scenes, rest of the scenes with camera pan
	Music	46.81	52.89	1080p/50fps	Music concert, high movement of the singer with some sudden scene change
LabII & CSIII	Sport2	62.46	36.33	1080p/24 fps	Sport, field athletics , high vertical motion (panning), uniform color in the of the racetrack and lots of spectators in the upper portion of the clip
	Movie4	49.26	57.37	1080p/24 fps	Action, high speed pursuit in a city, high motion component with lots of scene changes, central picture portion rather constant

For the crowdsourcing study (CSIII), we followed the recommendations in [30] for the execution of video QoE assessment in crowdsourcing environments. To ensure arrival and play-out of non-distorted videos, the test sequences were downloaded to the local cache of the respective crowd worker. In order to identify reliable subjects upfront of the video rating task we included a gamified visual task where subjects were asked to identify certain moving shapes and numbers in two screen images and to click on them. Furthermore, watching times and clicking behaviour were monitored, and questions with respect to the displayed numbers were asked. Too long watching time and numerous clicks on not present shapes were recorded (for details see [31]) together with not correct answers. Only subjects clicking on existing shapes and answering correctly were allowed to proceed to the next (video rating) stage.

After the successful download of the video, the video playout was started in fullscreen mode and followed by the quality question on an ACR 5-graded scale as used in the LabII study. Furthermore, reliability questions as described in [30] together with user behaviour related parameters (playback time, focus time, toggling of fullscreen mode, pausing the video plyer etc.) were captured in order to online computing the subjects reliability. If they passed this reliability check, they were able to go through another rating (and reliability check) cycle and get paid for it. In total a maximum of three videos could be rated by each subject (for the details and procedure of reliability computation see [31]). The crowdsourcing campaign was run 25 days on microworkers.com. In total 673 micro workers participated in the study and issued 1593 ratings. After the reliability check, 576 reliable subjects remained (290 of them rated only one video and 286 of them rated two or more videos) with a total of 1377 reliable ratings. As we had 46 test video sequences, this resulted in minimum of 29 ratings per condition.

The subjective scores of both laboratory and crowdsourcing experiments were eventually grouped based on the HRCs and accordingly the MOS and 95% CI were calculated.

IV. RESULTS

In the following, first the existing results from each study which address the aforementioned research questions are presented in individual bar plots. Then, the statistical analysis of each study factors considering the results of individual study as well as the whole dataset are depicted.

RQ1: With regard to the up- and down-switching, what is the QoE gain of smooth switching over abrupt switching? Does chunk length have any effect on QoE? What is the impact of switching amplitude (defined as bitrate difference between the current and target quality level)?
In order to investigate these questions, various scenarios were considered and evaluated in LabI and LabII studies (cf. Table III). Since the abrupt/smooth switching scenarios were tested using different chunk length, we analyzed the chunk length effect in this part (although it could be also considered as part of the RQ2). Figure 4 and Figure 5 show the MOS obtained for these scenarios averaged over the relevant PVSs. It can be observed that smooth switching does not necessarily provide better perceptual quality compared to the abrupt way. In order to consistently analyze this effect, the F-test statistic and corresponding p-value from one-way ANOVA of individual scenarios were obtained which are presented in Table VI and Table VII. It can be seen that only for some of the scenarios smooth switching provided significantly better QoE compared to the abrupt switching. These results are in contrast to several results from related work such as [10].

To assess the effect size (partial Eta-squared[1]) of switching behavior, chunk length and switching amplitude, the ANOVA was conducted considering all the scenarios from both studies. In regard to up-switching, no significant difference between the abrupt and smooth change was found. Only in the case of down-switching a 'small effect'[2] was observed (cf. Table VIII). Considering the influence of the chunk length, significant results with only 'small effect' were obtained for both up- and down-switching. With respect to switching amplitude, a significant result was obtained with a 'medium effect' for the up-switching and a 'small effect' for the down-switching.

---

[1]Partial Eta-squared, η_p^2, describes the magnitude of an effect of which ascertain the *practical significance* of statical significance. Even a statically significant result obtained for a factor 'may not' be practically important if the effect size is too small [32].

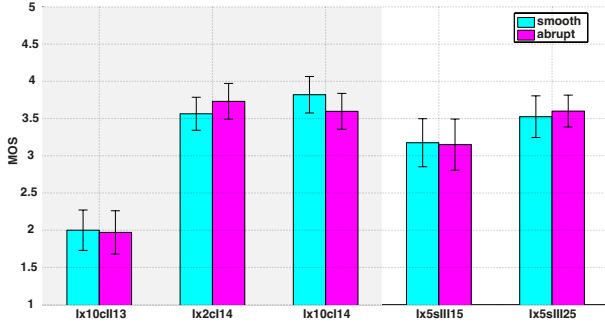[2]Results corresponded to the effect size are presented between comas.

Fig. 4: Increasing (LabI & LabII): Smooth switching does not provide significantly better QoE than abrupt switching. This observation is based on applying different chunk lengths.
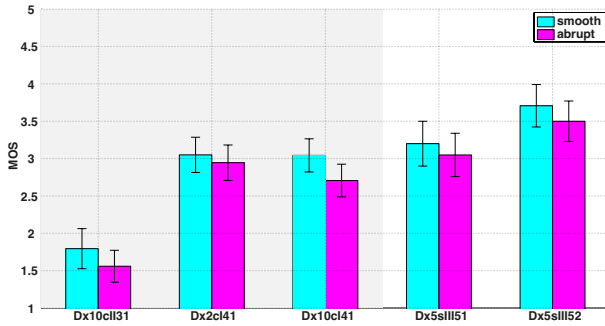

Fig. 5: Decreasing (LabI & LabII): Smooth switching does not provide significantly better QoE than abrupt switching. This observation is based on applying different chunk lengths.

TABLE VI: Statistical test results about the effect of abrupt vs. smooth up-switching scenarios presented in Figure 4.

HRC	Ix10cII13	Ix10cI14	Ix2cI14	IxsIII15	Ix5sIII25
F	0.022	19.328	12.443	0.011	0.184
p	0.880	1.19E-05	0.000	0.914	0.668

TABLE VII: Statistical test results about the effect of abrupt vs. smooth down-switching scenarios presented in Figure 5.

HRC	Dx10cII31	Dx10cI41	Dx2cI41	DxsIII51	Dx5sIII52
F	1.944	42.630	4.120	0.535	1.136
p	0.167	$< 10^{-4}$	0.042	0.466	0.289

RQ2: Does more frequent quality switching result in worse QoE than less frequent switching? What is the influence of the last quality level in terms of recency effects?

For this comparison, scenarios were implemented by oscillating between two quality levels (see high/low frequency switching in Figure 1 for a graphical representation of these profiles and Table IV for the respective switching characteristics). In Figure 6 results from the laboratory study (LabII) are depicted. The video sequences used in this study were adapted on the compression dimension. The analysis of the result showed no significant difference between high and low frequent quality adaptation across all conditions as well as the individual pair comparison shown in Table IX. However, considering switching amplitude in this scenarios, a significant 'medium effect' was observed.

Figure 7 shows the results from the crowdsourcing study (CSIII), where quality was varied along the spatial dimension.

TABLE VIII: Statistical results about the effect of the parameters in RQ1-RQ3 obtained from the one-way ANOVA over the whole dataset. The wavy underlined, dashed underlined and bold numbers in the right column correspond to the significant parameters with small, medium and large effects in order.

| RQ	Study parameter	F	p	η_p^2
RQ1	Smooth/Abrupt-Increasing	0.47	0.492	0.001
	Smooth/Abrupt-Decreasing	38.09	$< 10^{-4}$	0.012
	Chunk length-Increasing	5.23	0.005	0.003
	Chunk length-Decreasing	29.24	$< 10^{-4}$	0.019
	Amplitude-Increasing	75.33	$< 10^{-4}$	0.070
	Amplitude-Decreasing	45.60	$< 10^{-4}$	0.043
RQ2	Switching frequency-Oscillation	1.33	0.248	0.001
	Amplitude-Oscillation	87.55	$< 10^{-4}$	0.042
	Recency effect	76.49	$< 10^{-4}$	**0.337**
RQ3	Content-Increasing	56.81	$< 10^{-4}$	**0.132**
	Content-Decreasing	70.19	$< 10^{-4}$	**0.158**
	Content-Abrupt/Smooth-Inc	2.09	0.04	0.006
	Content-Abrupt/Smooth-Dec	4.61	$< 10^{-4}$	0.013
	Spatial information	6.98	0.008	0.001
	Temporal information	366.29	$< 10^{-4}$	0.057
	Spatiotempral information	134.50	$< 10^{-4}$	0.063
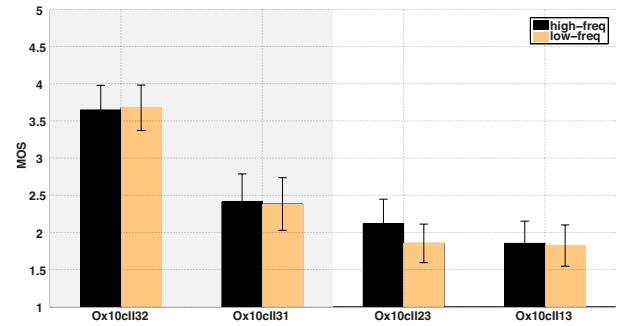
Fig. 6: Switching frequency (LabII): Switching frequency considering compression dimension has no measurable significant negative effect. For all used PVS the duration was the same .
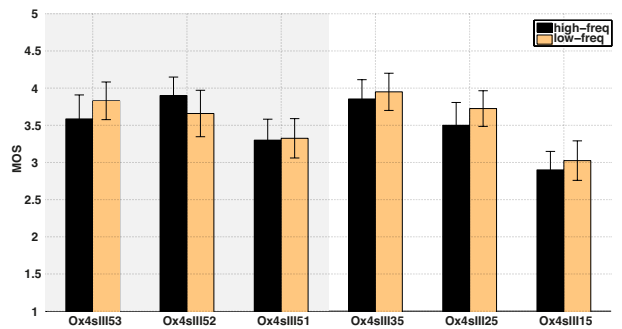

Fig. 7: Switching frequency (CSIII): Switching frequency considering spatial dimension has no measurable significant negative effect. For all used PVS the duration was the same .

The three left bar pairs, marked with the grey background, were starting and ending on the high quality level, whereas the three right bar pairs started and ended at the low quality level. The statistical analysis of these scenarios revealed that the switching frequency has no significant effect on the resulting MOS. This can also be seen from the individual comparisons for each bar pairs presented in Table X where none of the pairs is statistically significantly different. On the other hand,
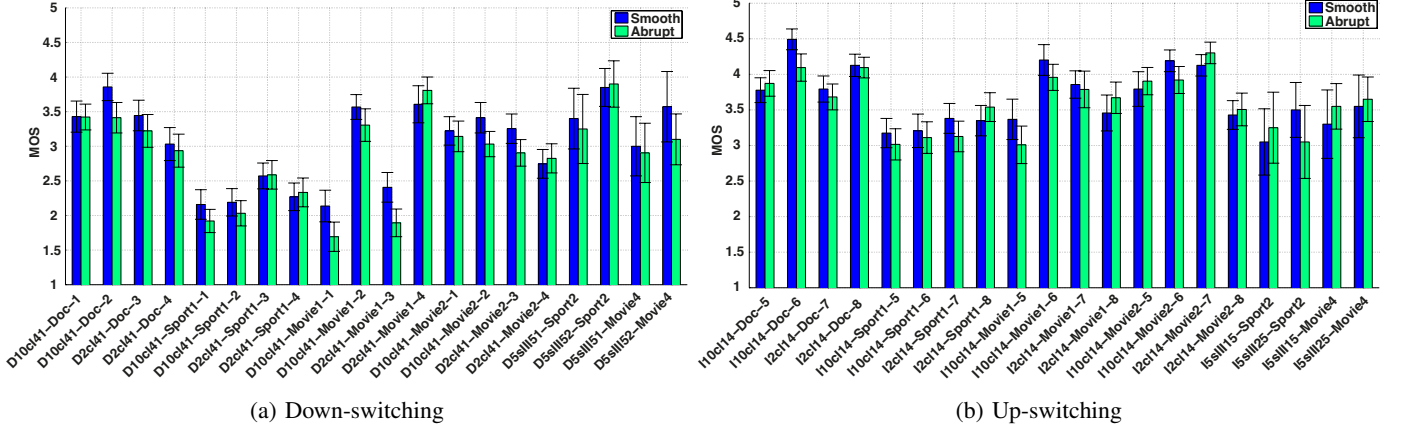
(a) Down-switching

(b) Up-switching

Fig. 8: Content type has significant effect on perception of smooth/abrupt down- and up-switching (cf. Table VIII). The PVSs shown in x-axis represent scenarios from Table III which are applied on different content in different studies (cf. Table V).

TABLE IX: Statistical test results about effect of switching frequency scenarios presented in Figure 6.

HRC	10cII32	10cII31	10cII23	10cII13
F	0.017	0.013	0.021	1.644
p	0.894	0.908	0.883	0.204

TABLE X: Statistical test results about effect of switching frequency scenarios presented in Figure 7.

HRC	4sIII53	4sIII52	4sIII51	4sIII35	4sIII25	4sIII15
F	1.439	1.484	0.017	0.290	1.362	0.481
p	0.233	0.226	0.896	0.591	0.246	0.489

the amplitude of the quality switch accounts for a significant difference with 'medium effect'.

Beyond the separate analysis of the results for each of the switching dimensions discussed above, the influence of switching frequency, switching amplitude and recency effect across all scenarios were also analyzed. The result underlined that switching frequency has no significant influence, but the influence of switching amplitude is significant with a 'small effect'. In addition, the starting and ending bitrate do have a 'large effect' (cf. Table VIII): sequences starting and ending at the high quality level yielding higher QoE scores compared to sequences starting and ending at the lower quality levels. For this result we want to note, that most probably the recency effect accounts largely for the ending bitrate rather then the influence of starting bitrate. This is backed by a two-way ANOVA of increasing and decreasing conditions. We found that the average bitrate of the test condition has only a 'small effect' ($F = 99.405$, $p < 10^{-4}$, $\eta_p^2 = 0.047$), but the switching direction (increase vs. decrease) had significant 'medium effect' ($F = 811.151$, $p < 10^{-4}$, $\eta_p^2 = 0.119$), which supports the impact of the target quality level of the sequence. These findings are in line with results from [16] but contradictory to results presented in [5] and [6]. We explain this with the fact that, [5] and [6] used considerably lower chunk lengths which led to flickering effects in the videos and very low QoE scores accordingly. For the studies presented in this paper, chunk lengths were 10 sec (LabII) and

4 sec (CSIII) which are in the typical range of current HAS solutions [22]. For such externally valid chunk lengths our results show that higher frequent switching does not lead to lower perceived video quality. However, other factors such as switching amplitude and recency effects do have a significant and larger influence on the resulting quality perception.

RQ3: What is the impact of content type on the QoE of different switching strategies?

Figure 8 show the comparison between smooth and abrupt down- and up-switching in different content. It can be observed that the perceptual effect of different switching scenarios varies in different content. This finding was confirmed by statistical analysis of the results, which showed a 'large effect' of the content on perception of both up- and down-switching (cf. Table VIII). In addition, the significant impact of content type on perception of abrupt and smooth switching were observed with 'small effect' for down-switching and 'very small effect' for up-switching (cf. Table VIII).

To study the influence of objective characteristics of the content on QoE of the adaptation, content were classified by a combination of their amount of spatial and temporal complexities (SI and TI) as formulated in [24]. Subsequently, four content classes resulted: low spatial-low temporal (LS-LT), low spatial-high temporal (LS-HT), high spatial-low temporal (HS-LT), and high spatial-high temporal activity (HS-HT). The result of ANOVA showed that spatial and temporal characteristics of the content have significant 'medium effect' on perception of adaption. Specifically, perception of the adaption in the HS-LT content class was significantly lower compared to the other content classes. This finding confirms the previous result presented in [33]. On the other hand, influence of the spatial activity individually (difference of high spatial vs. low spatial complexity) was significant with 'low effect', while in the case of temporal activity (difference of high temporal vs. low temporal complexity) the influence was significant with 'medium effect'.

In regard to the impact of content genre, it is worth noting that by looking at the results over the whole dataset (Figure 8) it is hard to derive a conclusion in regard to the effect of this factor

on perception of abrupt/smooth down/up-switching scenarios. However, considering the results of the content used in each study (cf. Table V), we can observe that the perception of adaptation in Sport video is worse than other content types, specially in LabI.

RQ4:Is it better to switch the quality or stay at a constant low quality level?

An important question for service providers is whether quality adaptation does yield better quality at all- and if yes, when (at which quality level) adaptation should be performed in order to achieve a true QoE improvement. To this end, Figure 9 and Figure 10 compare the existing MOS for the test sequences representing the adaptation representations (with the constant quality) and the MOS for the corresponded up-switching PVSs in LabI and LabII.

Considering Figure 10, adaptation in order to increase the quality (IS2cI14, IS10cI14, IA2cI14 and IA10cI14) yields significantly better MOS compared to the constant low video quality like 2cI and 1cI (Adaptation vs. 2cI: $F = 118.09$, $p < 10^{-4}$, $\eta_p^2 = 0.032$; Adaptation vs. 1cI: $F = 983$, $p < 10^{-4}$, $\eta_p^2 = 0.221$). Similarly, according to Figure 10, quality adaptation yields a clear QoE gain only compared to the lowest video quality level (1cII), as quality adaptation from the lowest to the highest video quality level (IS10cII13) results in significantly better quality ($F = 138.01$, $p < 10^{-4}$, $\eta_p^2 = 0.676$). Hence, if video quality is at the lowest level (based on the current study, equal or lower than 2cI in LabI, and equal or lower than 1cII in LabII), adaptation always improves the perceptual quality as the quality improvement
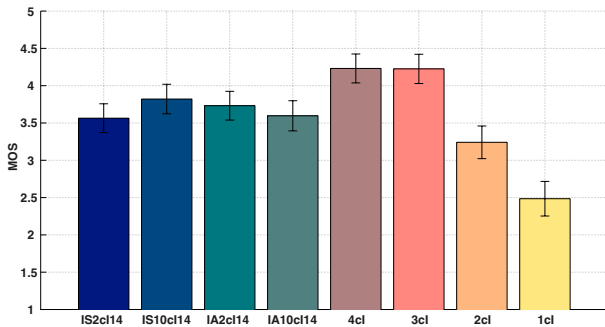

Fig. 9: Increasing vs. constant quality (LabI). Adaptation significantly outperforms (in terms of QoE) the lowest quality levels (1cI and 2cII).
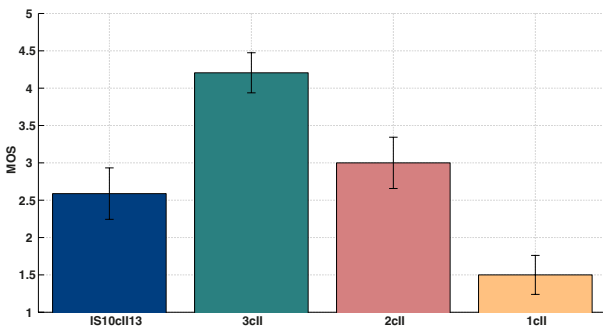

Fig. 10: Increasing vs. constant quality (LabII). Adaptation significantly outperforms (in terms of QoE) the lowest quality level (1cII).

might compensate (positively) the annoyance effect of the quality change.

V. DISCUSSION

In this paper, we investigated the influence of HAS related parameters on the video QoE by analyzing a large dataset obtained from four laboratory and one crowdsourcing experiments. Our statistical analysis which was performed on both individual studies and over the whole dataset demonstrated a strong symmetry between the studies i.e., the influence of identical study factors was perceived similarly across different experiments. Within this section we compare our results with results from related work and discuss commonalities and differences.

One of our most important observations was about the the effect of smooth vs. abrupt switching. While no significant difference between these scenarios was found for quality up-switching, the abrupt down-switching of video quality was perceived significantly worse than smooth down-switching (RQ1). In other words, the observers prefer to experience the higher quality as soon as possible, while in the case of quality decreasing, smoother transition between quality levels are more favorable. This indicates a clear interaction effect between switching direction (up/down) and switching strategy (smooth/abrupt). However, this finding also poses a practical challenge to improve adaptation algorithms, which accordingly should avoid an abrupt drop of the quality if the bandwidth decreases. decrease. Therefore, the process of down-switching the quality has to be initiated early enough so that a smooth transition would be possible. This could require that lower quality representations have to be additionally downloaded in order to prepare for presenting the upcoming chunk in an intermediate quality level even though a higher quality chunk is available in the buffer. This would avoid the abrupt switching to the target (lower) quality level. In order to not discarding such otherwise redundantly downloaded segments, SVC codecs could be used which can enhance an already downloaded representation to higher quality levels [34], [35]. Taking the aforementioned settings for quality oscillation scenarios into account, no significant effect of switching frequency on perceptual quality was observed (RQ2). This finding contradicts the statement that video quality switching is a degradation itself [4], but on the other hand it confirms the result of [16]. Thus, adaptation intervals do not have to be considered in the first place when seeking an appropriate chunk length for a HAS system. Nevertheless, the trade-off between the reaction time and data volume still remains valid. This means that on the one hand the chunk length should be short enough to be able to adapt fast to changing network conditions, on the other hand, a larger chunk length allows for higher coding efficiency and lower overhead. Moreover, by studying the effect of starting and ending bitrate of the quality oscillation scenarios, the significant influence of the recency effect with a large effect was revealed. This finding is in contrast to the results presented in [16] where no effect of the ending quality was reported. However, the reason of different observations could be related to the different parameters used

to produce the quality oscillations in the two studies (e.g. quality profiles, time staying on the last quality, etc.).

Our analysis of the results with regard to the content influence (RQ3) shows that content has a significant impact on subjective perception of video quality adaptation. Depending on video characteristics as well as the user attention and focus a video attracts, a quality change is perceived differently. Moreover, it has to be taken into account that watching content triggers different psychological processes (e.g., understanding, liking, commitment) that interact with perception processes [19]. All these points are plausible explanations for different perceived quality of identical (abrupt/smooth) switching patterns, depending on the actual content of the clip In this respect, content-specific properties of the video offer several possibilities which a HAS system can exploit for optimization. For example, the adaptation intervals or quality dimensions (e.i. resolution, frame rate, image quality) can be selected in such ways (e.g. video cuts, fast motion scenes, region of interest) that quality switches and degradation are obfuscated and QoE is improved. Some studies have already had some attempts in this regard. As an example of studies which consider the content for designing the adaptation scheme is [36]. According to this adaptation scheme, regions within each frame of the streamed content are adapted differently according to the user interest on them. Other studies on content-based adaptive video schemes are [37], [38].

On the other hand, we observed that taking the objective characteristics of the video into account, specifically spatial and temporal information, can be beneficial to improve the adaptation viewing experience. In this sense, we identified that both SI and TI characteristics as well as their combination exert significant influence on perception of quality adapted video sequences, although the temporal aspect might have a larger effect than spatial one. Thus, the influence of these factors on HAS QoE is measurable. However, deriving a model that comprehensively describes these influences remains an open research question.

Concerning the question about quality switching or keeping the quality constant (RQ4), our results showed that increasing the quality is beneficial in terms of QoE. The reason behind this observation is not the mere up-switching the quality per se. By considering the scenarios featuring the same average bitrate (e.g. ISxcI14 and IAxc14), there is no significant difference between their perceptual quality. However, comparing the scenarios that start on a certain quality level and increase later on with those that stay constant on the initial quality level (e.g. 1cI), up-switching improves the QoE. This finding is in line with the previous studies [13], [16] that showed similar results. The consequence is that HAS systems should be able to switch up to a higher quality as soon as possible.

Summarizing the results of our statistical analysis showed that the quality difference between initial and target quality level (switching amplitude) for both up- and down-switching should be kept as small as possible (RQ1). With respect to switching frequency in the quality oscillation scenarios (RQ2), contradictory to previous studies, no significant perceptual effect was observed. However, the recency effect caused by the target quality level of the sequence was significant.

Considering content characteristics, we showed an influence of objective content characteristics on QoE perception of different switching parameters (RQ3). Our results also indicate interaction effects as well: while it is advisable to perform smooth down-switching, we found no gain in doing so in up-switching (RQ1). In addition, while abrupt down-switches should be avoided, switching up to a higher video quality level as soon as possible is beneficial (RQ4).

Finally, it is worth reminding that the presented results are based on the settings considered in the current dataset. In order to find more robust conclusion regarding the switching frequency and recency effect (RQ2) in addition to the impact of content genre (RQ3) further investigation is recommended.

VI. CONCLUSION

In this article we investigated the QoE influence of different video quality adaptation-related factors. Our research was motivated by a lack of evidence for certain HAS typical impairments and contradicting results in related work. To this end, we investigated the perceptual impact of different adaptation-related parameters by means of totally five subjective quality experiments conducted in different environments and settings (lab, crowdsourcing). The test scenarios included numerous quality profiles for increasing and decreasing the video quality with different parametrisation and also quality oscillations. The stimuli were prepared by applying different chunk lengths, video quality representations and switching dimensions (compression and spatial) on a large set of video contents featuring different genres and different spatial and temporal characteristics, respectively.

Our thorough statistical analysis demonstrated that not all assumptions and claims in related work on HAS QoE are robust and that they do not hold true in several cases. Specifically, the non-significant impact of switching frequency on perception of quality oscillation indicates that quality switching does not represent a degradation per se. In addition, the negative QoE influence of abrupt vs. smooth switching is not omnipresent. It is rather connected to the switching direction by an interaction effect. For down-switching smooth-switching performs better, but for switching up to higher video quality abrupt-switching as soon as possible is beneficial. Another result showed the influence of target quality levels on perception of the quality oscillation scenarios. This can be explained by the well-known recency effect. Furthermore, we were also able to confirm the influence of chunk length, switching amplitude and content characteristics on the perceived video quality of HAS.

This article also shows that subjective data gathered in different lab contexts provides comparable results. Hence, such data pooling can be effectively used for comparing the perceptual quality of a large number of adaptation scenarios. The presented results are derived from a diverse range of experimental settings and setups (including PC/laptop and TV display) hence can be used for the creation of holistic models for improving the QoE of adaptive streaming services in the future Internet. Future work will focus on influence of content classes, the interaction effect between quality switching direction and switching strategy, in addition to evaluating the

switching behaviors using different video codec (e.g. HEVC) towards obtaining codec independent results. Gaining deeper insights into these factors will path the way for the future HAS QoE models and respective optimized switching strategies that can be successfully deployed in future Internet services.

Acknowledgment

References

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

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

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

[4] B. Lewcio *et al.*, "Video Quality in Next Generation Mobile Networks – Perception of Time-varying Transmission," in *Proc. of the IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR)*, Naples, FL, USA, 2011.

[5] 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.

[6] P. Ni, R. Eg, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Flicker effects in adaptive video streaming to handheld devices," in *Proc. of the 19th ACM international conference on Multimedia*. ACM, 2011, pp. 463–472.

[7] S. Tavakoli, J. Gutiérrez, and N. García, "Subjective Quality Study of Adaptive Streaming of Monoscopic and Stereoscopic Video," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 684–692, April 2014.

[8] R. Mok, X. Luo, E. Chan, and R. Chang, "QDASH: A QoE-aware DASH System," in *Proceeding of 3rd Multimedia Systems Conference*, Feb 2012, pp. 11–22.

[9] N. Staelens, J. De Meulenaere, M. Claeys, G. Van Wallendael, W. Van den Broeck, J. De Cock, R. Van de Walle, P. Demeester, and F. De Turck, "Subjective quality assessment of longer duration video sequences delivered over http adaptive streaming to tablet devices," *Broadcasting, IEEE Transactions on*, vol. 60, no. 4, pp. 707–714, Dec 2014.

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

[11] M. Zink, O. Künzel, J. Schmitt, and R. Steinmetz, "Subjective Impression of Variations in Layer Encoded Videos," *Proc. 11th International Conference on Quality of Service*, pp. 137–154, 2003.

[12] D. C. Robinson, Y. Jutras, and V. Craciun, "Subjective Video Quality Assessment of HTTP Adaptive Streaming Technologies," *Bell Labs Tech. Journal*, vol. 16, 2012.

[13] A. K. Moorthy, L. Choi, A. C. Bovik, and G. de Veciana, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, 2012.

[14] S. Tavakoli, K. Brunnström, K. Wang, B. Andrén, M. Shahid, and N. García, "Subjective quality assessment of an adaptive video streaming model," *Proc. IS&T/SPIE Int. Conf. on IQSP XI*, vol. 9016, Feb. 2014.

[15] C. Chen, L. Choi, G. de Veciana, C. Caramanis, R. Heath, and A. Bovik, "Modeling the Time Varying Subjective Quality of HTTP Video Streams with Rate Adaptations," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2206–2221, May 2014.

[16] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner, "Assessing Effect Sizes of Influence Factors Towards a QoE Model for HTTP Adaptive Streaming," in *6th International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, Sep. 2014.

[17] P. Kortum and M. Sullivan, "The effect of content desirability on subjective video quality ratings," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 1, pp. 105–118, 2010.

[18] L. Janowski and P. Romaniak, "Qoe as a function of frame rate and resolution changes," in *Proc. of the 3rd International Conference on Future Multimedia Networking (FMN)*. Berlin: Springer-Verlag, 2010, pp. 34–45.

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

[20] S. Tavakoli, K. Brunnström, J. Gutiérrez, and N. García, "Quality of experience of adaptive video streaming: Investigation in service parameters and subjective quality assessment methodology," *Signal Processing: Image Communication, Special Issue on Recent Advances in Vision Modelling for Image and Video Processing*, vol. 39-B, pp. 432–443, Nov. 2015.

[21] M.-N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnström, and A. Raake, "Quality of Experience and HTTP Adaptive Streaming: a Review of Subjective Studies," in *Proc. 6th International Workshop on Quality of Multimedia Experience (QoMEX)*, Sep. 2014, pp. 141–146.

[22] S. Lederer, C. Müller, and C. Timmerer, "Dynamic adaptive streaming over http dataset," in *Proc. of the 3rd Multimedia Systems Conference*. ACM, 2012, pp. 89–94.

[23] 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.

[24] International Telecommunication Union, "Subjective video quality assessment methods for multimedia applications," *ITU-T Recommendation P.910*, April 2008.

[25] J. Gutierrez, P. Perez, F. Jaureguizar, J. Cabrera, and N. Garcia, "Subjective Assessment of the Impact of Transmission Errors in 3DTV Compared to HDTV," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, May 2011, pp. 1–4.

[26] K. Brunnström, R. Cousseau, J. Jonsson, Y. Koudota, V. Bagazov, and M. Barkowsky, "VQEGplayer For Performing Subjective Video Quality Experiments," Video Quality Expert Group, Available: http://vqegjeg.intec.ugent.be/wiki/index.php/VQEGplayer-main, 2015.

[27] International Telecommunication Union, "Methodology for the Subjective Assessment of the Quality of Television Pictures," *ITU-R Recommendation BT.500*, 2012.

[28] VQEG, "Report on the Validation of Video Quality Models for High Definition Video Content," Video Quality Expert Group, Available: www.vqeg.org, June 2010.

[29] S. Egger, B. Gardlo, M. Seufert, and R. Schatz, "The impact of Adaptation Strategies on Perceived Quality of HTTP Adaptive Streaming." in *1st Workshop on Design, Quality and Deployment of Adaptive Video Streaming (VideoNext)*, Dec. 2014.

[30] T. Hoßfeld and C. Keimel, "Crowdsourcing in QoE Evaluation," in *Quality of Experience: Advanced Concepts, Applications and Methods*, A. R. Sebastian Mller, Ed. Springer: T-Labs Series in Telecommunication Services, ISBN 978-3-319-02680-0,, Mar. 2014.

[31] B. Gardlo, S. Egger, and M. Seufert, "Crowdsourcing 2.0: Enhancing Execution Speed and Reliability of Web-based QoE Testing," in *Proc. IEEE ICC, Sydney, Australia*, Jun. 2014.

[32] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2013.

[33] S. Tavakoli, M. Shahid, K. Brunnström, and N. García, "Effect of Content Characteristics on Quality of Experience of Adaptive Streaming," in *6th International Workshop on Quality of Multimedia Experience (QoMEX)*, Sep. 2014, pp. 63–64.

[34] J. Famaey, S. Latre, N. Bouten, W. Van de Meerssche, B. De Vleeschauwer, W. Van Leekwijck, and F. De Turck, "On the merits of SVC-based HTTP Adaptive Streaming," in *IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Ghent, Belgium, 2013.

[35] C. Müller *et al.*, "Using Scalable Video Coding for Dynamic Adaptive Streaming over HTTP in Mobile Environments," in *Proc. of the 20th*

European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania, 2012.

[36] B. Ciubotaru, G. Ghinea, and G.-M. Muntean, "Subjective Assessment of Region of Interest-Aware Adaptive Multimedia Streaming Quality," *Broadcasting, IEEE Transactions on*, vol. 60, no. 1, pp. 50–60, March 2014.

[37] S. Hu, L. Sun, C. Gui, E. Jammeh, and I.-H. Mkwawa, "Content-Aware Adaptation Scheme for QoE Optimized DASH Applications," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, Dec 2014, pp. 1336–1341.

[38] Y. J.-Q. Hu Sheng-Hong, "Content-Based Adaptive Transmission for Soccer Video," *Advanced Science Letters*, vol. 10, no. 1, pp. 478–485(8), 2012.

Samira Tavakoli received the Master degree in Telecommunication Engineering from the Blekinge Tekniska Högoskola (BTH), Sweden, in 2010. In 2015, she finished her Ph.D. thesis on "Subjective QoE Analysis of HTTP Adaptive Streaming Applications" in the Universidad Politécnica de Madrid (UPM), Spain. Since 2010, she is a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) at the UPM. From 2012, she has been working with Acreo Swedish ICT AB in the area of subjective multimedia quality studies. Her interests include research on Quality of Experience in multimedia services, evaluation of user-oriented techniques in the ICT domain and advances in laboratory methodologies.

Sebastian Egger, is Scientist at AIT, working on Technology Experience and QoE evaluation in the domains of human-to-human mediated interaction, interactive data services and online video. He holds a Dipl.-Ing. and doctoral degree in electrical engineering from Graz University of Technology as well as a Mag.phil. in Sociology from University of Graz. Since 2010 he participates in standardization activities of the ETSI STQ and ITU-T Study Group 12 on Performance, QoS and QoE where he successfully acted as an editor for two new standards on Web QoE, namely ITU-T G.1031 and ITU-T P.1501. He is author of numerous conference and journal papers and acts as reviewer and TPC member for international conferences and journals such as IEEE ICME and IEEE Transactions on Image Processing. On a European level he has been actively involved in COST 298, COST IC1003 QUALINET and the CELTIC project QuEEN. His main research interests in QoE are evaluation methodologies, quality evaluation for speech, video and interactive video services as well as behavioural aspects of transmission quality changes.

Michael Seufert, studied Computer Science, Mathematics, and Education at the University of Würzburg, Germany. In 2011, he received his Diploma degree in Computer Science, and additionally passed the state examinations for teaching Mathematics and Computer Science in secondary schools. From 2012 to 2013, he was with FTW Telecommunication Research Center, Vienna, Austria, working in the area of user-centered interaction and communication economics. He is currently a Researcher at the Chair of Communication Networks, University of Würzburg, working toward the Ph.D. degree. His research mainly focuses on QoE of Internet applications, social networks, performance modeling and analysis, and traffic management solutions.

Raimund Schatz, has recently joined the Austrian Institute of Technology (AIT), business unit Technology Experience. Until end of 2015 he was Key Researcher at the Telecommunications Research Center Vienna (FTW) and Manager of the User-centered Interaction, Services, and Systems Quality Department. He holds an Msc. in Telematics (TU-Graz), a PhD in Informatics (TU-Vienna), an MBA and an MSc. in International and Finance Management (both from Open University, UK). From 2009 to 2015, Dr. Schatz has lead FTWs research projects on Quality-of-Experience assessment and monitoring of broadband services in wireless and wireline networks, conducted together with a number of industry partners in the telecom industry. Furthermore, he is or has been actively involved in various QoE-related EU projects and networking activities, including Optiband (FP7), CELTIC QuEEN and COST Action IC1003 (Qualinet) and IC1304 (ACROSS) as well as the organization of various QoE-related workshops and events (e.g. Special Session Chair for QoMEX 2013; Organizer of PQS 2013, general chair of QoENAM 2014, QoE-FI 2015 and QoE-FI 2016). Being an active member of ACM and IEEE, he is author of more than 100 publications in the areas of Quality of Experience, HCI and Pervasive Computing.

Kjell Brunnström, Ph.D., is a Senior Scientist at Acreo Swedish ICT AB and Adjunct Professor at Mid Sweden University. He is an expert in image processing, computer vision, image and video quality assessment having worked in the area for more than 25 years, including work in Sweden, Japan and UK. He has written a number of articles in international peer-reviewed scientific journals and conference papers, as well as having reviewed a number of scientific articles for international peer-reviewed journals. He has supervised Ph.D. and M.Sc students. Currently, he is leading standardization activities for video quality measurements as Co-chair of the Video Quality Experts Group (VQEG). His research interests are in Quality of Experience for visual media in particular video quality assessment both for 2D and 3D, as well as display quality related to the TCO requirements.

Narciso García received the Ingeniero de Telecomunicación degree (five year engineering program) in 1976 (Spanish National Graduation Award) and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) in 1983 (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. Since 1977 he has been a member of the faculty of the UPM, where is currently a Professor of Signal Theory and Communications. He leads the Grupo de Tratamiento de Imágenes of the UPM. He has been actively involved in Spanish and European research projects, serving also as evaluator, reviewer, auditor, and observer of several research and development programs of the European Union. He was a co-writer of the EBU proposal, base of the ITU standard for digital transmission of TV at 34-45 Mb/s (ITU-T J.81). He was Area Coordinator of the Spanish Evaluation Agency (ANEP) from 1990 to 1992 and General Coordinator of the Spanish Commission for the Evaluation of the Research Activity (CNEAI) from 2011 to 2014. He was awarded the Junior and Senior Research Awards of the Universidad Politécnica de Madrid in 1987 and 1994, respectively. His professional and research interests are in the areas of digital image and video compression and of computer vision.