User Experience Modeling for DASH Video

Yao Liu, Sujit Dey
University of California, San Diego
{yal019, dey}@ece.ucsd.edu

Don Gillies, Faith Ulupinar, Michael Luby
Qualcomm Technologies, Inc.
{dgillies, fulupina, luby}@qti.qualcomm.com

***Abstract*— Ever since video compression techniques have been introduced, measurement of perceived video quality has been a non-trivial task. Recently, a new class of video transport techniques has been introduced for transmission of video over varying channels such as wireless network. These transport techniques, called adaptive streaming, vary the bit rate and quality of the transmitted video to match the available channel bandwidth. DASH, Dynamic Adaptive Streaming over HTTP, is a new worldwide standard for adaptive streaming of video, audio and other media such as closed captioning. The adaptive streaming techniques introduce an additional level of complexity for measuring perceived video quality, as it varies the video bit rate and quality. In this paper, we study the perceived video quality using DASH. We investigate three factors which impact user perceived video quality: *initial delay*, *stall* (frame freezing), and *bit rate* (frame quality) fluctuation. Moreover, for each factor, we explore multiple dimensions that can have different effects on perceived quality. For example, in the case of the factor *stall*, while most previous research have studied how stall duration correlates with user perceived quality, we also consider when the stalls happen and how the stalls are distributed, since we believe they may also impact user experience. We design and conduct extensive subjective tests to study the impairments of the different dimensions of the three factors on user perceived video quality. We will describe the methodology to design the subjective tests, and present the results of the subjective tests. Based on the subjective tests, we derive impairment functions which can quantitatively measure the impairment of each factor on the user experience of any DASH video, and also provide validation results.**

I. INTRODUCTION

The advent of video compression techniques together with the wide adoption of mobile devices such as smart-phones and tablets, and the fast growth of mobile networks, have brought videos closer to everyone's life. Video streaming applications are growing exponentially and becoming more and more popular. This has increased the desire for video quality assessment tools which can be conveniently applied in the context of video streaming. Streaming service providers can use these tools to monitor and control the end user's Quality of Experience (QoE).

Recently, a new class of video transport techniques has been introduced for transmission of video over varying channels such as wireless network. These transport techniques, called adaptive streaming, vary the bit rate and quality of the transmitted video to match the available channel bandwidth and alleviate the problems caused by network congestion, such as large latency and high packet loss rate. DASH, Dynamic Adaptive Streaming over HTTP, is a new worldwide standard for adaptive streaming [1], which enables delivering media content from conventional HTTP web servers. DASH works by splitting the media content into a sequence of small segments, encoding each segment into several versions with different bit rates and quality, and streaming the segments according to the requests from mobile client. On the mobile device side, the DASH client will keep monitoring the network and dynamically select the suitable version for the next segment that need to be downloaded, depending on the current network conditions. For good network conditions, the mobile client will request the server to deliver a high quality (high bit rate) version. When the network condition becomes bad, the mobile client will detect the network change and request a low quality (low bit rate) version.

On the DASH server side, each media segment is made available at a variety of bit rates. Each bit rate will be associated with a set of other encoding factors such as frame rate and resolution. Different streaming service providers might use different encoding options for a given bit rate. .As an example, Table I shows the bit rate options and the associated frame rates and resolutions that was used for streaming Vancouver Olympics [3] video using DASH. In this paper, we use the term *level* to represent a bit rate and the associated frame rate and resolution. As shown in Table I, the video segments are encoded using any of the 8 levels; each of them has a specific bit rate, frame rate, and resolution.

DASH has introduced an additional level of complexity for measuring video quality, since it varies the video quality during streaming. In this paper, we derive a set of impairment functions as a tool to quantitatively measure the user experience. We first identify three factors that will impact the user experience: *initial delay*, *stall* and *level variation*. We show that each of these factors have multiple dimensions which may impact user experience differently. We design and conduct subjective experiments by which viewers evaluate the effect on viewing experience when each of the three factors is varied. Based on the evaluations given by the participants of the subjective experiments, we derive impairment functions for each of the factors, which can quantify the impairment caused by this factor. For a given video resulting from streaming using DASH, the resulting impairment functions

Level	Bit Rate (kbps)	Resolution	Frame Rate
1	400	312 x176	15
2	600	400 x 224	15
3	900	512 x 288	15
4	950	544 x 304	15
5	1250	640 x 360	25
6	1600	736 x 416	25
7	1950	848 x 480	25
8	3450	1280x720	30

TABLE I. ENCODING STETTINGS FOR STREAMING VANCOUVER OLYMPICS

978-1-4799-2172-0/13/$31.00 ©2013 IEEE

can be used to measure the impact on user experience caused by each factor. Note the impairment functions such developed are not reference based measures, and no access is needed for the original video source. Hence, besides offline measurement of user experience, the impairment functions can be conveniently incorporated into DASH clients on mobile devices to measure the impairments during a live video session.

The impairment of the three factors on user experience may vary depending on the type of video content and watching environment. For instance, the impairment due to initial delay when watch a 2-hour movie may be quite different (and less) than how it impairs user experience when watch a 1-minute short video clip. Similarly, the impairments caused by a factor can be very different when watching on a large TV screen versus watching on a smaller phone screen. In this paper, we will focus on short video clips (such as Youtube videos that are a few minutes long) on mobile devices with up to medium size displays, such as smart phones and tablets.

Numerous video quality assessment methods have been proposed over the past years. Most of them [4][5][6] focus on measuring the video spatial quality (visual quality of video frame) and ignore the temporal artifacts such as stalls. In [7][8], models have been proposed to study the video temporal quality, but they don't include the variation of bit rate (visual quality) during the streaming session, and are therefore not suitable for DASH video. In [9][10], the authors have studied the impact of bit rate variation on user experience. While they derive interesting observations about how variation frequency and amplitude will affect user experience, but they do not develop ways to quantitative measure the effects. Moreover, they do not consider temporal artifacts such as stall. To the best of our knowledge, this paper is the first study to develop a quantifiable measure of user experience for DASH video, considering both spatial and temporal quality.

The remainder of the paper is organized as following: in section II, we introduce the factors that will affect user experience of DASH video. In section III, we first explain the characterization experiment we conducted to study how DASH performs in various mobile network conditions, and then we explain how we use the characterization experiment as a guideline to generate the test videos for subjective experiments. Section IV describes the subjective experiments. In section V, we analyze the subjective experiment results, and derive impairment functions for the different factors which affect user experience. Section VI concludes this paper and points out future work, specifically in developing and validating an overall user experience model which incorporates the impairment of all the factors. We might present the overall model in the final version if this paper is accepted.

II. FACTORS AFFECTING USER EXPERIENCE FOR DASH VIDEO

The first step to study and model user perceived video quality is to identify the impairment factors which impact user experience. In this section, we propose and explain three impairment factors that will affect the user experience for DASH video.

During a DASH video watching session, video will be transmitted over wireless network, which is characterized by quickly fluctuating and unpredictable bandwidth. In this streaming process, there are mainly three kinds of events which may affect the user perceived video quality: 1) there is an initial delay before the first frame of the video can be displayed, due to the need for the video client to buffer a certain amount of video data; the initial delay will be affected by the bandwidth experienced during initial buffering; 2) during a video session, it is possible that the bit rate adaptation cannot keep up with the network bandwidth fluctuation, leading to buffer underflow and stalling.; 3) during a video session, the video quality might keep switching, reducing the video quality will cause impairment to user experience, and continuous video quality switches will also harm user experience.

As shown in Figure 1, we investigate three objective factors: *initial delay*, *stall* (frame freezing) and *level variation*. The user experience for DASH video mainly depends on two subjective factors: temporal quality and spatial quality. The *initial delay* and *stall* will determine the temporal quality of the perceived video, and the *level variation* will determine the spatial quality of video.

Unlike *initial delay*, the factors *stall* and *level variation* are more complex and have multiple dimensions associated with them. For the stall *factor*, the total stall duration (in seconds) is crucial. Most of the previous research only study how stall duration correlates with user perceived quality. However, we think the number of stalls is also an important dimension. For example, consider total stall duration of 5 seconds: the effect on user experience may be different if there is a single stall of 5 seconds duration, versus five 1-second stalls. Similarly when the stalls occur (more at the beginning, versus other distributions) may also impact user experience differently. Hence, besides the conventional factor of stall duration, we would like to also consider the number and timing of the stalls as three dimensions of the factor *stall*.

Similarly we propose three dimensions for factor *level variation*: 1) average level, which indicates the average quality of the entire video session; 2) number of switches, which indicates the frequency of quality switch; 3) average switch magnitude, which indicates the average amplitude of quality change. For instance, for a level

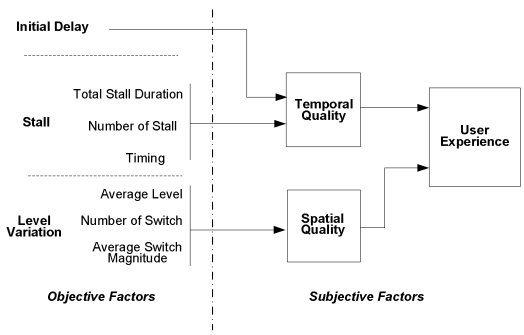
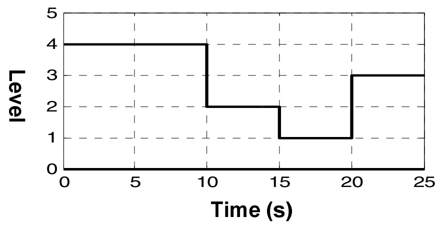


Figure 1: Factors affecting user experience of DASH video.



Average Level = (4+4+2+1+3)/5= 2.8

Number of Switch = 3

Average Switch Magnitude = [(4-2)+(2-1)+(3-1)]/3= 1.67

Figure 2: Level variation pattern.

variation pattern as shown in Figure 2, the average level is 2.8, number of switch is 3, and the average switch magnitude equals 1.67. Noted that in Figure 1 we haven't differentiate positive level switch (when the bit rate increases) and negative level switch (when the bit rate decreases). But as explained in section V, we will decide whether to treat positive switch and negative switch differently based on the results we get from subjective tests.

III. TEST VIDEO GENERATION

In order to derive functions to quantitatively measure the impairment of the 3 factors proposed in section II, we need to conduct extensive subjective tests, where each participant watches DASH video while one of the three factors varies. However, due to the multi-dimensional nature of the factors *stall* and *level variation*, there may be numerous cases we need to cover in the test videos. On the other hand, we need to constraint the number of test videos a subject can watch before loss of focus and fatigue can affect the quality of the testing. Motivated by this conflict, we need to design the test videos in an efficient way such that they cover a wide and representative range of the 3 factors, and we are able to derive impairment functions from limited number of test videos. In this section, we describe how we generate the test videos for the subjective tests.

A. DASH Video Streaming Characterization Experiments

In order to generate meaningful and representative test videos, we first conduct a set of DASH video streaming experiments to characterize how DASH will perform under real mobile network conditions. From the streaming experiments, we can understand what will be the possible range and distribution for the 3 factors under various network conditions. This range and distribution information will be used as a guideline to generate the test videos.

Figure 3 shows the testbed for the DASH characterization experiments. DASH videos are pre-encoded and stored at the media server. The media server and the mobile devices are connected through a network emulator, which can be used to control network parameters, such as bandwidth, latency and packet loss rate. On the network emulator, we can apply different mobile network bandwidth traces. At the mobile device side, a DASH player is running to display the received video and make decisions about the video level switch.

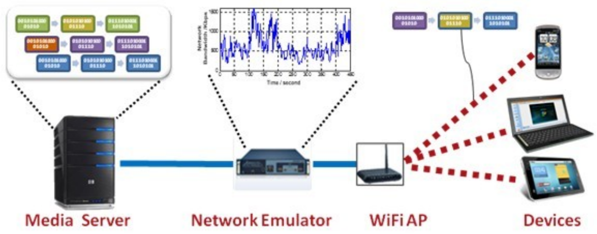


Figure 3: Testbed of DASH video streaming characterization experiment.

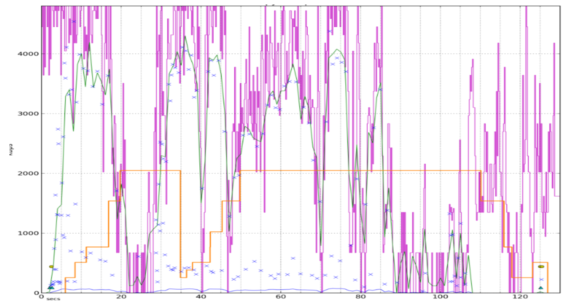


Figure 4: Results for characterization experiments: (1) purple curve: network bandwidth; (2) green curve: segments download bit rate; (3) yellow curve: video bit rate.

After each video streaming session, a log file is generated on the mobile device, including information about the 3 factors for this streaming session. For instance, this log file will tell during the streaming session, what the level variation pattern is, how many stalls occur and when they occur.

This testbed offers the flexibility for us to stream under different network conditions, and record the values of the 3 factors. In the characterization experiments, we stream a DASH video to an Android tablet, under 20 different mobile network conditions. This selected DASH video is 2-minute long and has medium amount of motion. It is splitted into 2-second segments and pre-encoded into 7 levels. The 20 mobile network traces are very wide-ranging and representative, captured with different mobile operator networks at different geographical locations, and include stationary as well as different mobility scenarios (like pedestrian, car, train, etc.). The bandwidth of the network traces vary between 4Mbps and 150kbps. Among the 20 network traces, the average bandwidth of each trace varies between 750kbps ~ 1850 kbps.

Figure 4 shows a representative result of the characterization experiments. The purple curve represents available mobile network bandwidth, the green curve shows the video segments downloading rate, and the

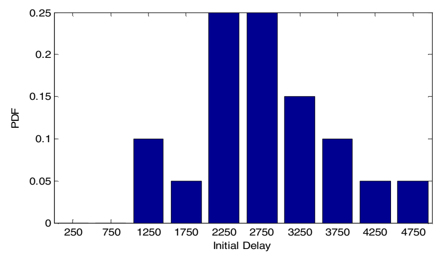


Figure 5: Distribution of initial delay (ms) among 20 streaming sessions.

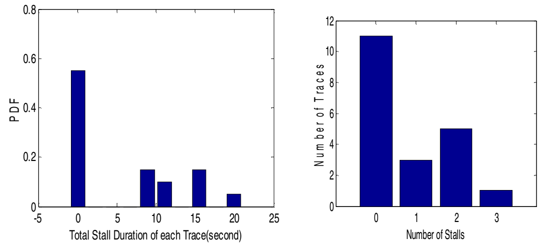


Figure 6: (a) left, distribution of total stall duration; (b) right, distribution of number of stalls.

yellow curve shows the actual adaptive video bit rate. We can see that the DASH adaptive bit rate (yellow curve) will switch up and down between several discrete steps due to the fast fluctuation of mobile network bandwidth (purple curve).

Figure 5 shows the distribution of initial delay among the 20 streaming sessions. The 20 initial delay values are between 1280ms and 4890ms. Figure 6(a) and 6(b) show the distribution of total stall duration and stall number among the 20 streaming sessions (each of them is 2-minute long). We can see that in 55% of the streaming sessions there is no stall happening. In the rest sessions, the stall number is less than 3. And the stall duration of a video session can be as long as 20 seconds. Figure 7(a), 7(b) and 7(c) show the distribution of the average level, number of switches, and average switch magnitude respectively. We can see that during the 2-minute streaming session, the number of level switches can vary from 6 to 21. The average switch magnitude is between 1 and 1.5, which indicates that the current DASH technique mainly utilizes switches with small magnitude, trying to avoid impairment caused by large quality change.

B. Generated Test Cases

After presenting the ranges and distribution of the 3 factors, in this subsection we will use them as a guideline to generate the test videos for subjective tests. However, we may also include test videos whose characteristics are outside of what was observed in the DASH characterization tests, to ensure we are covering more extreme cases. For instance, although the initial delay value we obtain from real experiments are all less than 5 seconds (Figure 5), we will also have test video with very long initial delay, like 15-second initial delay.

We design 41 test videos for subjective tests. Each of them is 1 minute long. In each test video, we only vary one factor and keep the other two factors at their best values. For instance, we have 5 test videos for deriving

Video ID	1	2	3	4	5
Initial Delay Value (s)	2	4	6	10	15

TABLE II. TEST VIDEOS FOR INITIAL DELAY

Video ID	6	7	8	9	10	11	12	13	14	15	16
Stall Duration	4 sec			8 sec				12 sec			4 sec
Stall Number	1	2	4	1	2	3	8	1	3	12	1

TABLE III. TEST VIDEOS FOR STALL

Figure 8: Snapshot of test videos for 3 factors (a) left, for initial delay; (b) middle, for stall; (c) right, for level variation.

the impairment function for initial delay. In these 5 test videos, there is only initial delay impairment; no stall occurs and the video level remains at the highest value. When people watch these 5 videos and give evaluations, they are only evaluating the impairment caused by initial delay. By generating test videos in this manner, we can separate the 3 factors, and be able to derive impairment function for each of them.

In order to keep people interested while watching the 41 test videos, we select 3 different videos. They are interesting open source videos with medium amount of motion. Figure 8 shows the snapshot of the video contents we use.

We use video #1~ #5 for deriving the impairment function of initial delay. As shown in Table II, we investigate the initial delay between 2 to 15 seconds.

Videos #6~ #16 are used to investigate impairment due to stall. As shown in Table III, the stall duration values we investigated are [4, 8, 12] seconds. Since we observed stall duration between 0 and 20 seconds for the 2-minute video sessions in the DASH characterization tests (Figure 6(a)), we assume that considering stall durations between 4 and 12 seconds will be reasonable for the subjective stets to be conducted with videos of 1-minute duration. Similarly, in video 6~16, we consider the number of stalls of 1~3, which corresponds to the result shown in Figure

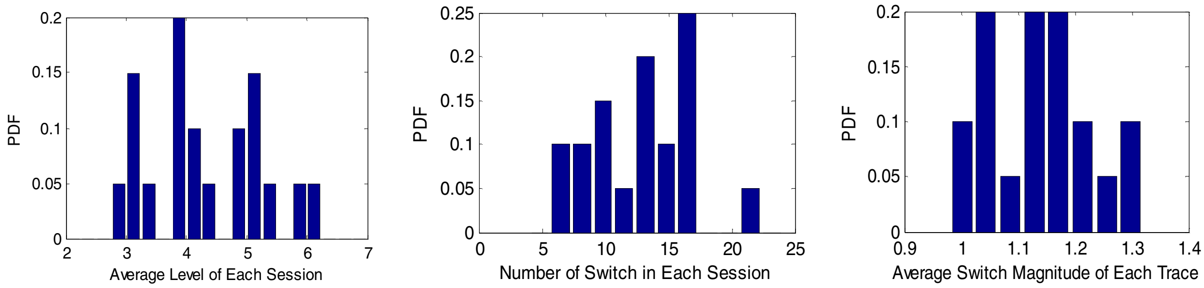


Figure 7: Distribution of level variation: (a) left, average level; (b) middle, number of switch; (c) right, average switch magnitude.

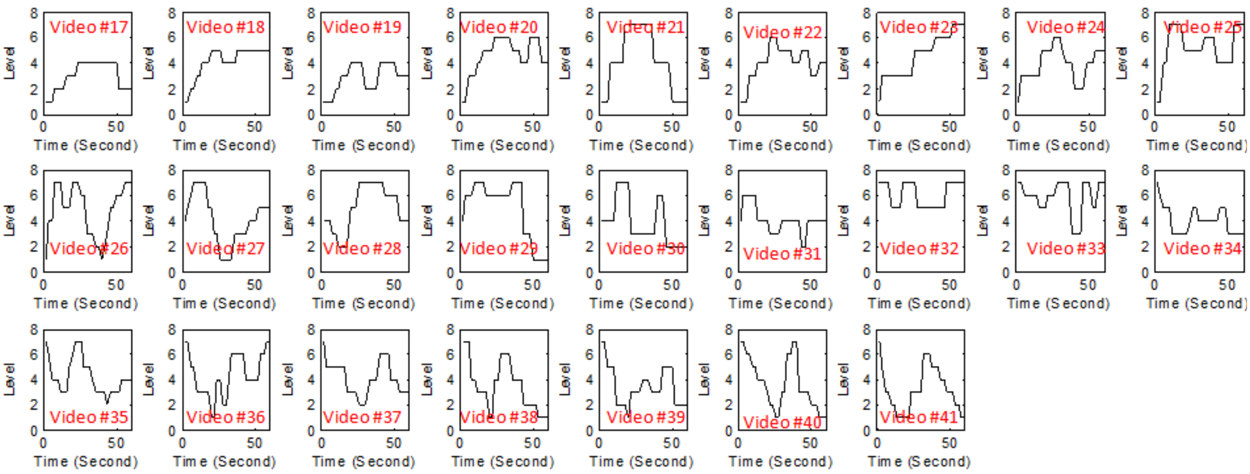


Figure 9: Test videos for level variation factor.

6(b). Moreover, we want to study the extreme cases where there are a lot of very short stalls and the stall number is bigger than 3. Video #8, #12 and #15 are videos with lots of 1-second stalls, and we want to understand how people feel with these frequent short stalls. Lastly, video #16 has the same stall number and stall duration as video #6, but in video #6, the stall happens in the middle part of video while in video #16 the stall happens at the beginning (3^{rd} second) of the video. By comparing the experiment result of video #6 and #16, we can understand whether early stall will be more annoying than stalls happening at later part of video.

Videos #17~#41 are designed for deriving impairment function of level variation factor. Figure 9 shows the level variation pattern of these 25 test videos. These 25 level variation patterns are designed to guarantee that: 1) the experiment results (range and distribution) shown in Figure 7 are met; 2) include plenty of different switch frequency and magnitude, both increasing switch and decreasing switch; 3) include different video starting levels and ending levels.

After generating these 41 test videos, in the next section, we will describe the subjective tests for deriving the impairment functions.

IV. Subjective Experiments

The subjective quality assessment experiments follow ITU-T Recommendations [11]. Each test video is presented one at a time, and each subject gives individual evaluation about the perceived video quality with a 100 point quality scale, as shown in Table IV. As the subjects are evaluating the perceived video quality, denoted as R, the corresponding impairment will be 100-R. The experiment is conducted in a lab environment with good light condition. A Qualcomm MSM8960 tablet with 1280x768 display resolution is used to watch the test videos.

TABLE IV. Rating Creteria for Video Quality

Quality Evaluation	Description
100	Excellent experience, no impairment at all
80-100	Minor impairment, will not quit
60-80	Noticeable impairment, might quit
40-60	Clearly impairment, usually quit
0-40	Annoying experience, definitely quit.

30 subjects from UCSD, with age ranges from 18 to 28, were selected for the study, satisfying the requirement of number of viewers specified by ITU-T Recommendations [11]. To ensure their evaluations are not biased, the selection of the subjects is done so that they don't have prior knowledge or watching experience of DASH video. Each subject is first presented with a training sequence which is different from the test videos to help him/her get familiar with the experiment environment and adjust his/her comfortable viewing distance and angle.

The experiment, which contains 41 one-minute test videos, will be divided into two sessions, with a 10 minutes break. This ensures that the time period of continuous video watching will not exceed half an hour, such that viewers will not experience discomfort.

The evaluations for each test video *i* are averaged over all subjects to obtain an average video quality value, denoted by R_i. Correspondingly, the average impairment value of video *i* will be 100- R_i. In the next subsection, we will use these average impairment values to derive impairment functions for all the 3 factors.

V. Derivation of Impairment Functions

In this section, we analyze the results from the subjective experiments, and use them to derive impairment functions for the 3 factors. Noted that the functions we derive in this section can only be applied on short video, since our subject test is based on short 1-minute video.

A. Impairment Function for Initial Delay

In test videos #1~ #5, we add different length of initial delay in the beginning of video. The relation between the initial delay value and the average subjective impairment values is shown in Figure 10. We can see that the average

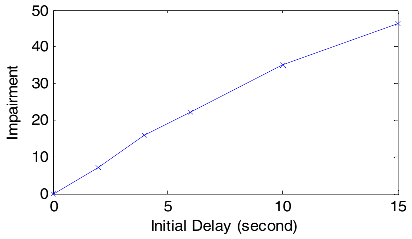


Figure 10 Relationship between impairment and initial delay.

Coefficient	α	a	b	c	k	B_1	B_2
Value	3.2	3.8	4.2	2.6	0.02	75.6	48.2

TABLE V. VALUES OF COEFFICENTS

subjective impairment value over the 30 subjects is almost linear with the initial delay values. Therefore, the impairment function for initial delay can be formulated as the following linear equation:

$$I_{ID} = \min\{\alpha * L_{ID}, 100\} \qquad (1),$$

where I_{ID} stands for the impairment due to initial delay, L_{ID} is the length of initial delay (in seconds). The coefficient α is computed by linear regression and is listed in Table V. We have also used a *min* function to clip the value of I_{ID} to force it to be within the range [0,100].

B. Impairment Function for Stall

As shown in Table III, we use test video #6 ~ #16 to investigate viewers' feedback about the 3 dimensions of stall: firstly, comparing video #6 and #16, we can understand whether stall happening in the beginning will cause more impairment than later stall; secondly, video #6 ~#15 let us know viewers' feeling with different stall duration and stall number.

Figure 11 shows viewers' preference about early stall and late stall (comparing video #6 and #16). Figure 11(a) shows that the average impairment evaluations people give for video #16 (early stall) and video #6 (late stall) is almost the same. Figure 11(b) shows users' answers when we ask them explicitly whether they think a early stall is more annoying than late stall after video #16 is shown to them. We can see from Figure 11(b) that 13 people think early stall is more annoying than late stall, 11 people think they are equally annoying, and 6 people think late stall causes more impairment. From Figure 11(a) and 11(b), we conclude that the early stall will cause a slightly bigger impairment but the difference is not obvious. Therefore we don't include early stall as a factor in deriving the impairment function for stall.

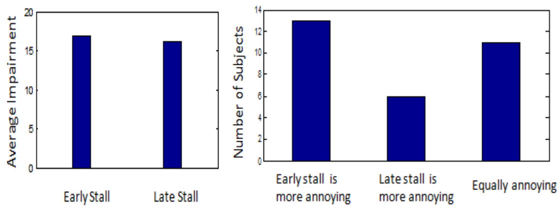


Figure 11: (a) left, average impairment values; (b) right, viewers' preference.

Table VI shows the average impairment values for video #6 ~ #15, where stall duration and stall number vary. From these results, we have the following observations:

a) When stall number is fixed, the impairment value will increase monotonically with stall duration.

b) When stall duration is fixed, the impairment value will not increase monotonically with stall number. However, we observe that the impairment value is highest with the highest stall number, which indicates that frequent stalls will cause big impairment on user experience.

Observation (b) tells us that we cannot use a linear equation to model the relationship between impairment due to stall with stall number and duration. Therefore, we propose to use a non-linear equation (2) as the impairment function for stall:

$$I_{ST} = a * D_{ST} + b * N_{ST} - c * \sqrt{D_{ST} * N_{ST}} \qquad (2)$$

In equation (2), I_{ST} stands for the impairment due to stall, D_{ST} indicates the total duration of stall, and N_{ST} stands for the number of stall. The cross-term in equation (2) is used to compensate the simultaneous effects of stall duration and stall number, and try to match the phenomenon explained in observation (b). The values of coefficients *a*, *b* and *c* in equation (2) are derived by linear regression and are listed in Table V.

C. Impairment Function for Level Variation

Level variation is the most complex factor to study, since it is difficult to characterize and abstract the complex patterns of level variations during a video session. As introduced in section II, there are 3 dimensions for level variation factor: average level, number of switches, and average switch magnitude. We need to derive an impairment function which can cover and reflect all the 3 dimensions.

Table VII shows the average evaluation of the impairment for test videos #17 ~ #41(shown in Figure 9). From the results we have the following observations:

a) All the 3 dimensions of the factor level variation will together affect user experience in a complex manner. For instance, comparing video #18 with video #21, both of them have average level of 4.1, but the impairment of video #21 is much bigger than that of #18. The same average level may lead to totally different user experiences, depending on the level fluctuation pattern. Therefore it may be difficult to use the method used for deriving I_{ID} and I_{ST} to derive the impairment due to level variations.

b) The annoyance of staying at a low level (low quality) will grow exponentially with the duration that the

Video ID	6	7	8	9	10	11	12	13	14	15
Stall Duration	4 sec			8 sec				12 sec		
Stall Number	1	2	4	1	2	3	8	1	3	12
Impairment Value	16.5	21.8	31.3	31.1	27.3	33.3	47.5	40.8	37.5	58.5

TABLE VI. SUBJECTIVE EXPERIMENT RESULTS FOR DIFFERENT STALL DURATION AND STALL NUMBER

Test ID	17	18	19	20	21	22	23	24	25
Impairment	33.5	24.4	32.9	21.3	44	28	13.6	19.6	12.2
Test ID	26	27	28	29	30	31	32	33	34
Impairment	29	34.3	22.7	41.5	24.8	23.8	13	10.2	19.3
Test ID	35	36	37	38	39	40	41		
Impairment	18.5	36.5	25.2	40.2	30.2	44.6	47.5		

TABLE VII. Subjective Experiment Results for Level Variation Tests

low level is maintained. Comparing video #26 with #29, both of them have average level of 4.9 and similar amount of level switch magnitude, but video #26 has much smaller impairment than #29. This is because in video #26, when the level drops to the lowest value (level 1), it only lasts for about 2 seconds and then jumps up; in video #29, level stays at 1 for more than 10 seconds. If the low level (bad quality) just lasts for a short period of time, the viewer might not complain much. But if a low level is maintained for a long time (such as more than 10 seconds), people will feel great annoyance.

c) The impact of decreasing level switch is much larger than that of increasing switch. Comparing video #18 with video #37, they both have average level of 4.1, but video #37 has much more impairment than video #18. This is because the level switches in video #18 are mostly increasing switches, while the switches in video #37 are mostly decreasing switches. Therefore, we cannot treat increasing switch and decreasing switch equally when we derive the impairment function.

Based on the results and observations, we next discuss how to derive an impairment function for the factor *level variations*. Firstly, we need to point out that we cannot use "level" directly in the impairment function. Different streaming service providers will have different encoding settings for each level. For the same level, different service providers will specify different frame rates and resolutions associated with it. If we derive an impairment function based on level value, then this impairment function cannot be applied generally.

Therefore, we propose to use VQM [2] instead of level in impairment function. VQM is a widely accepted objective video quality metric, which has been proven to have good correlation with human perception. But VQM cannot be applied directly to DASH video because it doesn't consider the impairment due to level variation. The VQM value is a number between 0 and 1. A lower VQM value indicates a better video quality. In this paper, we use VQM_i to indicate the amount of annoyance of video segment i. A lower VQM_i value means better quality and less annoyance.

The need to use VQM will not cause too much additional effort for content providers. On the DASH media server, the video sources are split into fixed-length segments and encoded into different levels. For each video segment i encoded at level j, we can obtain its VQM value, VQM_{ij}. The process of obtaining VQM value for each segment at each layer can be conducted offline on the media server, and it only needs to be carried out once. Once this process is done, the VQM values can be utilized to measure experienced impairments for all the future users.

Figure 12 shows an example of the VQM values for different levels for the encoding settings we used in our study. We can see that increasing the bit rate will cause a sharp decrease in VQM when bit rate is low. When bit rate becomes higher, further increasing bit rate will not lead to significant decrease in VQM.

Next we will derive an impairment function using metric VQM. Basically, the impairment caused by level variation during a DASH video session consists of 2 parts: 1) the impairment caused by low level (bad video spatial quality); 2) the impairment caused by level fluctuations.

In order to derive impairment function, we first define the following terms: assuming in a video session, totally N video segments are being transmitted. All the video segments have the same duration T. Depending on the DASH implementation, value of T can be 2 seconds, 5 seconds or 10 seconds, etc. For each segment i, we define a term D_i, which indicates how many continuous segments before segment i have the same level with segment i. D_i is an integer that will accumulate if level remains at the same value. For example, if the first 3 segments of a video are all encoded in level 5, and the 4th segment is encoded in level 2, then D_1, D_2, D_3, D_4 will equal 0,1,2,0, respectively.

Then the first part of impairment (caused by low level itself) can be modeled as:

$$P_1 = \frac{1}{N}\sum_{i=1}^{N} VQM_i * e^{k*T*D_i} \qquad (3)$$

As shown in equation (3), the P_1 value (impairment due to low level) is a weighted average of VQM values of each video segment. The exponential term in equation (3) is added to comply with our observation (b) that the annoyance caused by a low level will grows exponentially with the duration that low level is maintained. We use value D_i to indicate how long the level of segment i has been maintained, and multiply VQM_i with the exponential term to obtain the real annoyance of segment i. The coefficient k in equation (3) is used to control how fast the annoyance grows with time. The value of k is determined heuristically and listed in Table V.

The second part of the impairment caused by level fluctuations can be modeled as:

$$P_2 = \frac{1}{N}\sum_{i=1}^{N-1} \left|VQM_i - VQM_{i+1}\right|^2 * sign(VQM_{i+1} - VQM_i) \qquad (4),$$

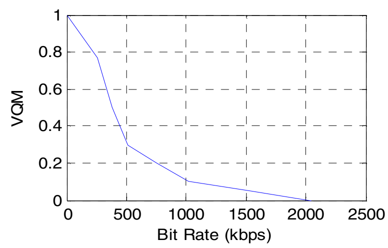


Figure 12 Relationship between VQM value and bit rate

where $sign(x) = \begin{cases} 1, & x > 0 \\ 0, & otherwise \end{cases}$ (5)

The value of P_2 is the average of the square of VQM differences between adjacent segments. According to our observation (c), the impairment caused by increasing switch is much smaller than that caused by decreasing switch. . Therefore in equation (4) we use sign function (equation (5)) to only include decreasing level switches and exclude increasing level switches.

Finally, the impairment due to *level variation* denoted as I_{LV}, is modeled as a weighted sum of P_1 and P_2:

$$I_{LV} = B_1 * P_1 + B_2 * P_2 \quad (6),$$

where B_1 and B_2 are coefficients which need to be derived later. Note that the proposed impairment function I_{LV} covers all the 3 dimensions of level switch: 1) average level, covered by P_1; 2) number of switch, covered by P_2; 3) average magnitude of switch, covered by P_2.

We have conducted a two-fold cross validation for the impairment function $I_{LV.}$ With the 25 test videos for level switch, we randomly choose 15 videos for developing the impairment function I_{LV}, and use the other 10 videos for validating the derived I_{LV}. Then we shuffle the 25 test videos, choose another set of 15 videos for developing impairment function, and use the rest for validation.

Figures 13(a) (b) show the results of the two different validations tests, specifically the relation between the subjective impairment values given by viewers with the objective impairment values computed by I_{LV}. The two validation tests achieve high correlation values of 0.88 and 0.84. We will pick the impairment function derived in the first round of validation as our final selected impairment function I_{LV}, as the first round validation leads to higher correlation. The corresponding coefficient values, B_1 and B_2 are derived using linear regression technique and are listed in Table V.

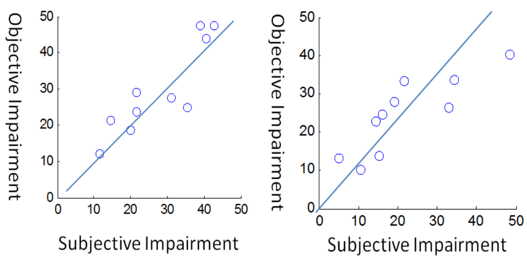


Figure 13: Relationship between subjective and objective impairments: (a) left, first round of validation for impairment function I_{LV}; (b) right, second round of validation for impairment function I_{LV}

VI. CONCLUSIONS

In this paper, we have proposed and derived a set of impairment functions which can quantitatively measure the user experience of DASH video, by taking into account both spatial and temporal artifacts. We investigate 3 factors which will impact user experience: *initial delay*, *stall* and *level variation*. Extensive subjective experiments are designed and conducted to derive the impairment functions. The proposed impairment functions don't need to access the uncompressed video source and can be conveniently incorporated into DASH client software to quantify user experience in real time. We believe they can serve as useful and light-weight tools for DASH service provider to monitor and control their quality of service.

However, in real DASH video applications, it is likely that more than one of the 3 factors will cause artifacts in one streaming session. Hence in the future, we will develop an overall user experience model which combines the three impairment functions together to measure the combined impairment of the three factors. To validate the combined user experience model, we will design and conduct another round of subjective experiments, in which subjects will evaluate video quality when they experience combined artifacts in a video session.

In this paper, we focused on short video clips with medium motion. The observations and impairment functions derived may not be applicable for long form videos such as movies, and for high motion videos such as sports videos. In the future, we intend to study how the three factors impact user experience for long form video and videos of high or low motion, and if the results are different, develop appropriate impairment functions.

VII. ACKNOWLEDGEMENT

The research reported in this paper was supported by a Qualcomm Fellow-Mentor-Advisor (FMA) Fellowship award.

REFERENCES

[1] ISO/IEC JTC 1/SC 29/WG 11 (MPEG), "Dynamic adaptive streaming over HTTP", w11578, CD 23001-6, Guangzhou, 2010.

[2] M.Pinson and S.Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. On Broadcasting,* vol. 50, no.3, pp.312-322, Sept. 2004.

[3] Jan Ozer, "Adaptive Streaming in the Field", in Streaming Media Magazine, 2011.

[4] S. Hemami and A. Reibman, "No-reference image and video quality estimation:Applications and human-motivated design," Signal Process.:Image Commun., vol. 25, no. 7, pp. 469–481, Aug. 2010.

[5] M. Ries, O. Nemethova and M. Rupp, "Video quality estimation for mobile H.264/AVC video streaming", Journal of Communications, Vol. 3, No.1, Jan. 2008, pp.41-50.

[6] Y.-C. Lin, D. Varodayan, and B. Girod, "Video quality monitoring for mobile multicast peers using distributed source coding," in Proc. 5th International Mobile Multimedia Communications Conference, London, 2009.

[7] R. Mok, E.Chan, "Measuring the Quality of Experience of HTTP Video Streaming", in *Proceeding of 2011 IEEE International Symposium on Integraed Network Magnagement,* Dublin, 2011.

[8] K. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of Experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC," Proceedings of Consumer Communication and Networking Conference (CCNC 2012)(Las Vegas, USA).

[9] P. Ni, R. Eg, A. Eichhorn, "Flicker Effects in Adaptive Video Streaming to Handheld Devices", in *Proceedings of the ACM International Multimedia Conference (ACM MM*), 2011

[10] P.Ni, R.Eg, A. Eichhorn, "Spatial Flicker Effect in Video Scaling", in *2011 International Workshop on Quality of Multimedia Experience*, Mechelen, 2011.

[11] BT-500-11: Methdology for Subjective Assessment of the Quality of Televison Picture, International Telecommunication Union.