Deriving and Validating User Experience Model for DASH Video Streaming

Yao Liu, Sujit Dey, *Fellow, IEEE*, Fatih Ulupinar, Michael Luby, *Fellow, IEEE*, and Yinian Mao, *Member, IEEE*

***Abstract*—Ever since video compression and streaming techniques have been introduced, measurement of perceived video quality has been a non-trivial task. Dynamic adaptive streaming (DASH) over hypertext transfer protocol, is a new worldwide standard for adaptive streaming of video. DASH has introduced 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: 1) initial delay; 2) stall (frame freezing); and 3) bit rate (frame quality) fluctuation. 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 experience, we also consider how the stalls are distributed together with the amount of motion in the video content, since we believe they may also impact user perceived quality. We conduct extensive subjective tests in which a group of subjects provide subjective evaluation while watching DASH videos with one or more artifacts occurring. Based on the subjective tests, we first derive impairment functions which can quantitatively measure the impairment of each factor, and then combine these impairment functions together to formulate an overall user experience model for any DASH video. We validate with high accuracy the user experience model, and demonstrate its applicability to long videos.**

***Index Terms*—Multimedia communication, quality of service, streaming media, videos**

I. Introduction

THE WIDE adoption of more capable mobile devices such as smart-phones and tablets, together with the deployment of higher capacity mobile networks and more efficient video compression techniques, are making mobile video consumption very popular. According to Cisco's mobile traffic forecast [1], mobile video consumption will increase 14-fold between 2013 and 2018, accounting for 69 percent of total mobile data traffic by the end of 2018. However, the success of mobile video streaming will largely depend on meeting user experience expectations. Therefore, it is highly desirable for video streaming service providers to be able to define, measure and, if possible, ensure mobile video streaming user experience.

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 international standard for adaptive streaming [2], 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 streaming client. On the client 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.

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 were used for streaming the Vancouver Olympics [4] videos 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.

It is well known that DASH video streaming is based on HTTP (Hypertext Transfer Protocol) and TCP (Transmission Control Protocol) which assure reliable video packets delivery and retransmission of lost packets. Using TCP retransmission and buffering mechanism can avoid audiovisual distortions caused by network artifacts such as jitter or packet loss. Instead, these network artifacts would lead to rebuffering interruptions and additional initial delay, which would deform the video's temporal structure and impact user experience. Furthermore, unlike regular TCP-based video streaming, DASH has introduced an additional level of difficulty for measuring video quality, since it varies the video quality during streaming. Although the video quality adaptation scheme can mitigate the temporal impairments such as rebuffering,

Manuscript received August 2, 2014; revised May 1, 2015; accepted May 12, 2015. Date of publication August 25, 2015; date of current version December 5, 2015. This work was supported by the FMA Fellowship from Qualcomm.

Y. Liu and S. Dey are with the Department of Electrical and Computer Engineering, University of California at San Diego, San Diego, CA 92093 USA (e-mail: yal019@ece.ucsd.edu; dey@ece.ucsd.edu).

F. Ulupinar, M. Luby, and Y. Mao are with Qualcomm Technologies Inc., San Diego, CA 92121 USA (e-mail: fulupina@qti.qualcomm.com; luby@qti.qualcomm.com; yinianm@qti.qualcomm.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2015.2460611

TABLE I
ENCODING SETTINGS FOR STREAMING VANCOUVER OLYMPICS

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

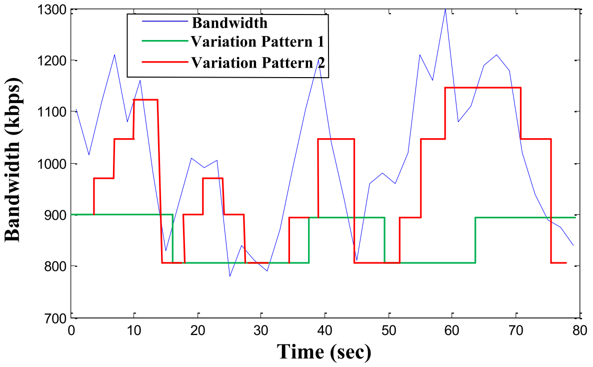


Fig. 1. Mobile bandwidth trace and DASH video quality variation patterns.

the quality variation during streaming may also impact the user experience of the viewers.

Quantifying the impairment caused by quality variation is non-trial but highly desirable. For example, Fig. 1 shows a mobile bandwidth trace and two associated video bitrate adaptation patterns produced by using different DASH algorithms. Pattern 1 is more conservative but provides a more stable overall quality. Pattern 2 is more aggressive, and it tries to increase the video quality whenever the available network bandwidth increases. It is difficult to tell which one is more preferable from a user experience perspective. The answer to this question may be helpful for video service providers to optimize their DASH quality adaptation algorithm.

Therefore, the aim of this paper is to derive a model to quantitatively measure the user experience of a video streamed using DASH, considering both temporal artifacts (like rebuffering) and spatial artifacts (like video quality variation). 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 one or more of the three factors are varied. Based on the evaluations given by the participants of the subjective experiments, we derive impairment functions for each of the factors, and then combine them together to form an overall user experience model. Note the proposed user experience model is a non-reference model, and no access is needed for the original video source. Hence, the proposed user experience model 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 video content, such as the amount of motion, or the duration of the video. For instance, the impact of stalls on user experience may be higher for a high motion video and less for a low motion video. Similarly, the impact of initial delay may be higher for a video with short duration and less for a video with long duration. Our proposed user experience model considers the amount of motion in the video content, and as we later demonstrate, can be applied to short to medium length videos covering most of online videos. Note that the impairment on user experience may also depend on other factors like how much a user likes the video or the type of mobile device (screen size/resolution) used. Our research and the proposed model do not consider these factors, which can be possibly investigated further in our future work.

Numerous video quality assessment methods have been proposed over the past years. Most of them [5]–[7], [14]–[17] focus on measuring the video spatial quality (visual quality of video frame) and ignore the temporal artifacts such as stalls. In [8], [9], [18], and [19], 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 [10] and [11], the authors have studied the impact of bit rate variation on user experience. While they derive interesting observations about how variation frequency and amplitude affect user experience, they do not develop ways to quantitatively 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 experiments we conducted to study how DASH performs in various mobile network conditions, and then we explain how we use the characterization experiments as a guideline to generate the test videos for subjective experiments. In Section IV, we describe the first round of subjective experiments, and derive impairment functions for the different factors based on experiment results. Section V describes a second round of subjective tests and the derivation of the overall user experience model. Section VI demonstrates application of the proposed model to long videos. Section VII concludes this paper and points out future work.

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.

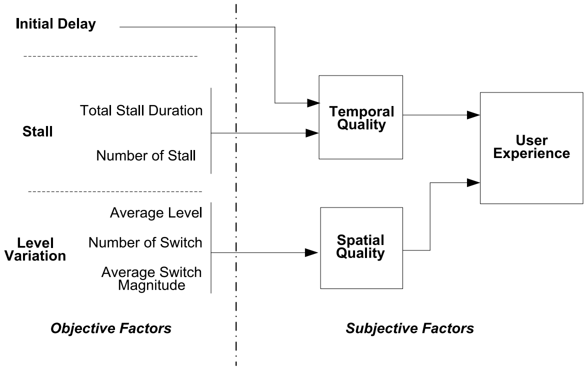


Fig. 2. Factors affecting user experience of DASH video.

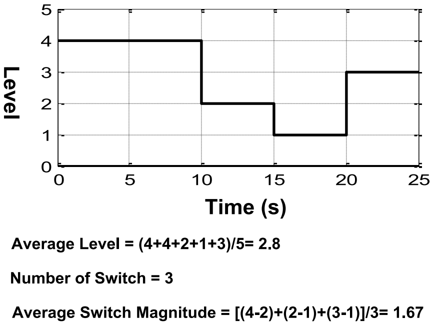


Fig. 3. Level variation pattern.

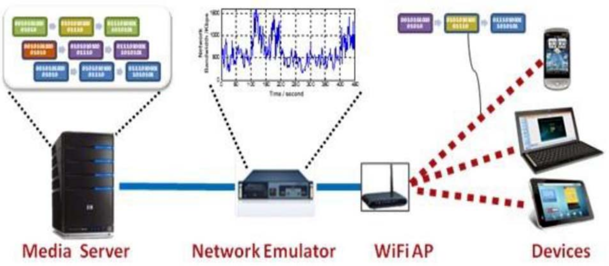


Fig. 4. Testbed of DASH video streaming characterization experiments.

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; 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 (rebuffering); 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 Fig. 2, we investigate three objective factors: *initial delay*, *stall* (rebuffering) 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 studied 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. Hence, besides the total stall duration, we would like to also consider the number of the stalls as a second dimension 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 variation pattern as shown in Fig. 3, the average level is 2.8, number of switch is 3, and the average switch magnitude equals 1.67. Noted that in Fig. 2 we haven't differentiated increasing level switch (when the bit rate increases) and decreasing level switch (when the bit rate decreases). But as explained in Section IV, we will decide whether to treat increasing switch and decreasing switch differently based on the results we obtain 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 tradeoff, we designed 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 a 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 performs 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.

Fig. 4 shows the testbed for the DASH characterization experiments. DASH videos are pre-encoded and stored at the

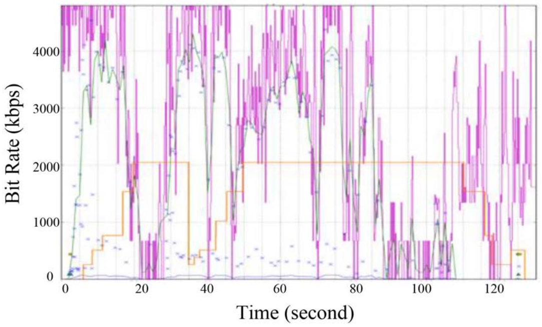


Fig. 5. Results for characterization experiments: (1) purple curve: network bandwidth; (2) green curve: segments download bit rate; (3) yellow curve: video bit rate.

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 displays the received video and makes video level switch decisions. 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 occurred and when they occurred.

This testbed offers the flexibility for us to stream under different network conditions, and records 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 split into 2-second segments and pre-encoded into 7 levels (with encoding bit rate of 256, 384, 512, 768, 1024, 1536, 2048 kbps, respectively). We use 20 mobile network traces that are wide-ranging and representative, captured with different mobile operator networks at different geographical locations, and include stationary as well as different mobility scenarios, such as pedestrian, car, train, etc. The bandwidth of the network traces varies between 4Mbps and 150kbps. Among the 20 network traces, the average bandwidth of each trace varies between 750kbps ~ 1850 kbps.

Fig. 5 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 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).

Fig. 6 shows the distribution of initial delay among the 20 streaming sessions, each using one of the 20 different network traces. The 20 initial delay values are between 1280ms and 4890ms. Fig. 7(a) and (b) show the distribution of total stall duration and stall number among the 20 streaming sessions (each of them is 2-minute long). We find that in 55% of the streaming sessions there is no stall happening. In the other sessions, the stall number is less than 3. And the stall duration of a video session can be as long as 20 seconds. Fig. 8(a)–(c) show the distribution of the average level, number of switches, and average switch magnitude respectively. We can see that during a 2-minute streaming session, the number of level switches can vary from 6 to 21. The average switch magnitude is between 1 and 1.3, which indicates that the current DASH technique mainly utilizes small magnitude switches to avoid impairment caused by large quality change.

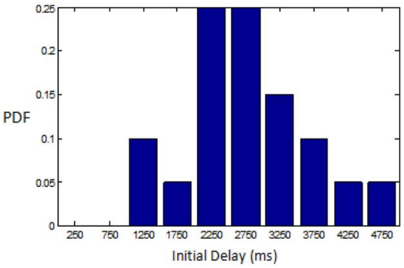


Fig. 6. Distribution of initial delay among 20 streaming sessions.

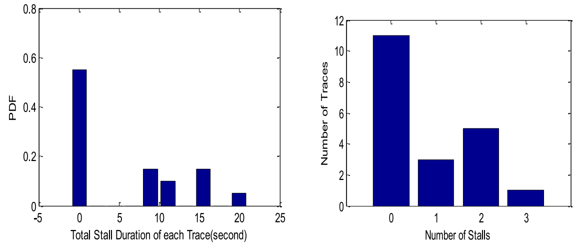


Fig. 7. (a) Left, distribution of total stall duration; (b) right, distribution of number of stalls.

B. Generated Test Cases for Round I Subjective Tests

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 round I subjective tests. We may also include test videos whose characteristics are outside of what was observed in the DASH characterization tests to cover more extreme cases. For instance, although the initial delay values we obtain from all real experiments are less than 5 seconds (Fig. 6), we will also have test video with very long initial delay, like 15-second initial delay.

We design 40 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 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. Fig. 9 shows a snapshot of the video contents we use, and Table II lists their descriptions. As can be seen, the 6 videos contents are selected to cover different

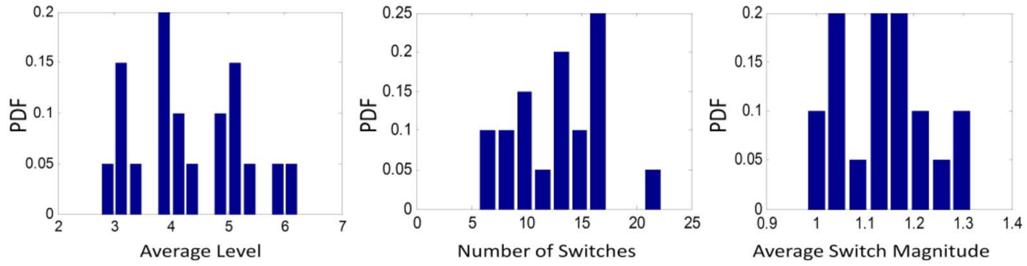


Fig. 8. Distribution of level variation: (a) left, average level; (b) middle, number of switches; (c) right, average switch magnitude.

Fig. 9. Snapshot of test videos.

TABLE II
VIDEO CONTENT DESCRIPTION

Video Name	Description
News	A woman reading news, low motion
Bunny	Cartoon about animals, medium motion
Sintel	Animated movie, medium motion
Steel	Human action movie , medium motion
Soccer	A soccer match, high motion
Surfing	Surfing, high motion, with multiple scene changes

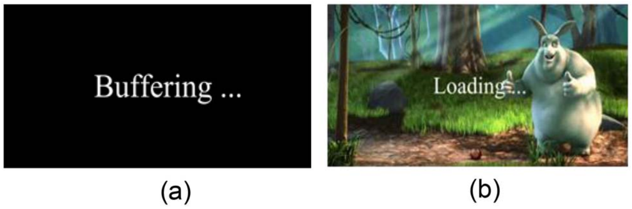


Fig. 10. Special frames used to simulate initial delay and stall: (a) left, 'buffering' frame for simulating initial delay; (b) right, 'loading' frame for simulating stall.

TABLE III
TEST CASES FOR INITIAL DELAY

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

TABLE IV
TEST CASES FOR STALL

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

video genres (news, animation, movie, sports) and different motion characteristics.

In order to simulate one of the three impairments in each test video, in this study we use ffmpeg [24] software as the tool for video encoding and processing. The version of ffmpeg we used is 0.8.15 and the selected video codec is H.264/AVC. To simulate the initial delay, we insert some identical 'buffering' frames (shown in Fig. 10(a)) in front of the raw video frames and then encode all the frames. The duration of the initial delay is controlled by the number of 'buffering' frames inserted. Similarly, the stall impairment is simulated by inserting some identical 'loading' frames (shown in Fig. 10(b)) in the middle of raw video frames. As shown in Fig. 10(b), the inserted 'loading' frame is the last raw video frame with a watermark of word 'Loading'. Moreover, the level variation impairment is simulated by encoding different groups of raw video frames with different encoding parameters and concatenating all the encoded video streams together.

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

Videos #6∼#15 are used to investigate the impairment due to stall. As shown in Table IV, the total 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 [Fig. 7(a)], we assume that considering stall durations between 4 and 12 seconds will be reasonable for the subjective tests conducted with videos of 1-minute duration. Similarly, in video 6∼15, we consider the number of stalls of 1∼3, which corresponds to the result shown in Fig. 7(b). We also 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.

Videos #16∼#40 are designed for deriving impairment function of level variation factor. Fig. 11 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 Fig. 8 are met;

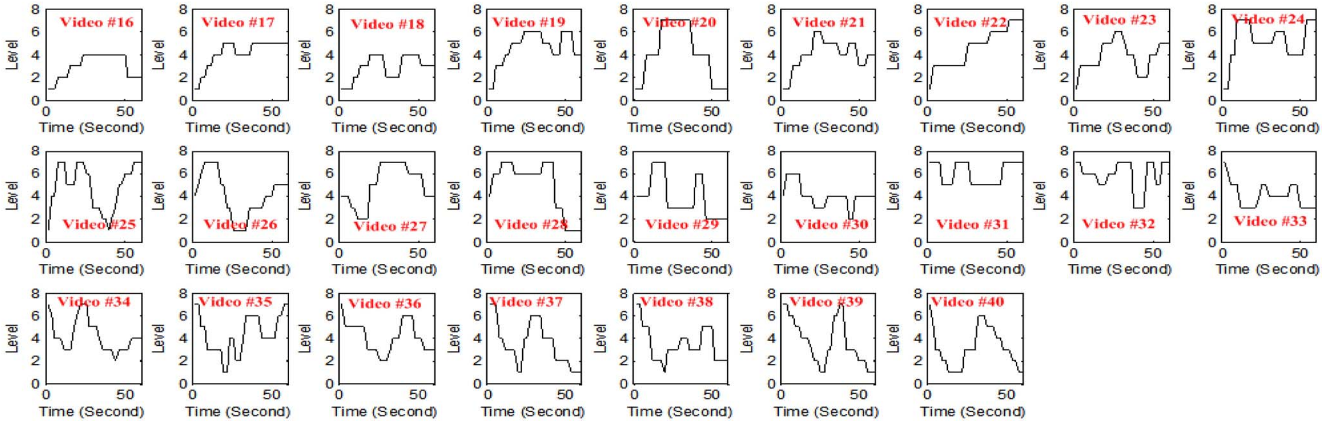


Fig. 11. Test videos for level variation factor.

TABLE V
RATING CRITERIA 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.

2) include plenty of different switch frequencies and magnitudes, both increasing switches and decreasing switches; 3) include different video starting levels and ending levels.

IV. DERIVATION OF IMPAIRMENT FUNCTIONS

After generating these 40 test cases, in this section, we will describe the first round of subjective tests, and then derive the impairment functions for the 3 factors according to the test results.

A. Round I Subjective Experiments

The subjective quality assessment experiments follow ITU-T Recommendations [12]. 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 V. 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.

30 subjects from University of California, San Diego (UCSD), with age ranging from 18 to 28, were selected for the study, satisfying the requirement of number of viewers specified by ITU-T Recommendations [12]. 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 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.

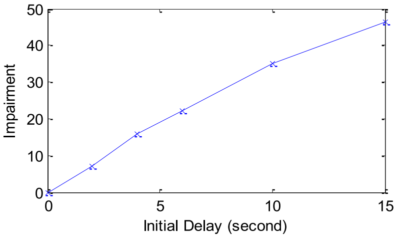


Fig. 12. Relationship between impairment and initial delay.

TABLE VI
VALUES OF COEFFICIENTS

α	a	b	c	d	k	B_1	B_2
3.2	3.35	3.98	2.50	1800	0.02	73.6	1608

B. Impairment Function for Initial Delay

In test videos #1∼#5, we add different length of initial delay in the beginning of the video. The relation between the initial delay value and the average subjective impairment values is shown in Fig. 12. We can see that the average subjective impairment of the 30 subjects is almost linear with the initial delay. Therefore, the impairment function for initial delay can be formulated as the following linear equation:

$$I_{ID} = \min\{\alpha^* L_{ID}, 100\} \tag{1}$$

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 VI. We have also used a *min* function to limit the impairment when it reaches its maximum.

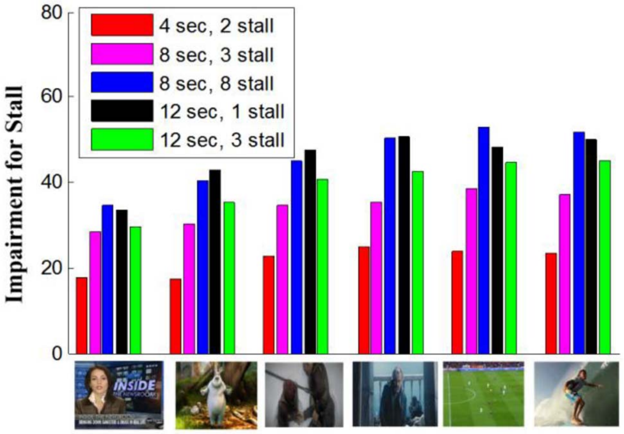


Fig. 13. Subjective stall impairment results for different video contents, stall duration and stall number.

C. Impairment Function for Stall

We have prepared test videos with different combinations of stall duration and stall number, for different video content. Fig. 13 shows an example of the stall impairment results of 5 different stall distributions on 6 different videos (as introduced in Fig. 9 and Table II). We find that besides stall number and stall duration, video content, more specifically the amount of motion in the video, also plays a crucial role in determining impairment value. We can see that for the same stall duration and stall number, high motion video has bigger impairment than low motion video. This may be due to the higher expectation/requirement of fluidness for high motion video content such as sports video.

Therefore in order to model the stall impairment, we first characterize the amount of motion in video. Then we will develop a function to model stall impairment.

1) Video Motion Characterization: We propose to use the average magnitude of motion vectors to characterize the amount of motion of a certain video content. In any regular video application, motion vector can be directly extracted from the encoded video bitstream without further computation. Furthermore, from the motion vectors in x and y direction, we compute the Motion Vector Magnitude (MVM) for each 16x16 macroblock, $\mathrm{MB_{ij}}$, which we define as:

$$MVM_{ij} = \sqrt{\left(\frac{m_{ij,x}}{N_x}\right)^2 + \left(\frac{m_{ij,y}}{N_y}\right)^2}, \qquad (2)$$

where N_x, N_y are the number of 16x16 MBs in the horizontal and vertical directions; and $m_{ij,x}$, $m_{ij,y}$ are the projection of motion vector on x and y directions for $\mathrm{MB_{ij}}$. In equation (2) we have normalized the MVM by the width and height of video frame to get rid of the influence of video spatial resolution on the motion characteristic of video content.

We then take the average of all the MBs to obtain the average MVM value of a video frame, denoted as $\overline{MVM}$:

$$\overline{MVM} = \frac{1}{N_x * N_y} \sum_{i=1}^{Nx} \sum_{j=1}^{Ny} MVM_{ij}$$

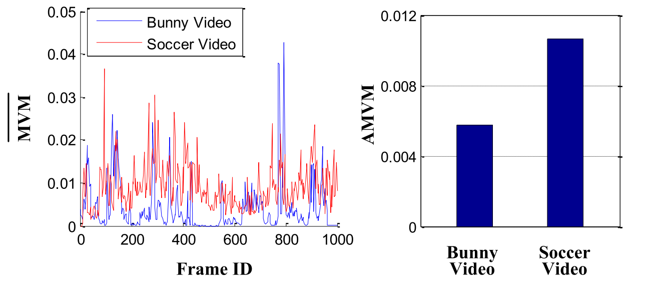


Fig. 14. (a) Left, $\overline{MVM}$ value of each frame; (b) right, AMVM for the whole video.

TABLE VII
SUBJECTIVE EXPERIMENT RESULTS FOR DIFFERENT STALL DURATION AND STALL NUMBER, FOR VIDEO *Bunny*

Total 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

Let us use $\overline{MVM_k}$to denote the average MVM value for the k-th frame, then the Average Motion Vector Magnitude (*AMVM*) of the whole video can be computed as:

$$AMVM = \frac{1}{M} \sum_{k=1}^{M} \overline{MVM_k},$$

where M is the number of frames in a video.

We will use AMVM as the metric to characterize the amount of motion of a video. As an example, Fig. 14(a) shows the motion vector magnitude of video *Bunny* and *Soccer* (as described in Table II) frame by frame. Fig. 14(b) shows the average motion vector magnitude (AMVM) of the two videos. We can see that using AMVM we can clearly differentiate high motion video and medium motion video. Hence, AMVM is an effective and easy-to-obtain metric to characterize the amount of motion in video.

2) Stall Impairment Function Derivation: After being able to quantify the amount of motion in a video, we then derive a stall impairment function, I_{ST}, based on the test results under different combinations of stall number, stall duration and AMVM.

First, we investigate for a given AMVM value, how stall number and stall duration affect the impairment due to stall (I_{ST}). Table VII shows the average impairment values for a certain video (we choose video *Bunny* as an example, it is a cartoon video with medium motion), where stall duration and stall number vary but AMVM is fixed. From the results listed in Table VII, we make the following observations:

Observation (a): When stall number is fixed, the impairment value increases monotonically with stall duration.

Observation (b): When stall duration is fixed, the impairment value does not increase monotonically with stall number. We also observe that the impairment value is highest with the highest stall number, which indicates that frequent stalls will cause high impairment on user experience.

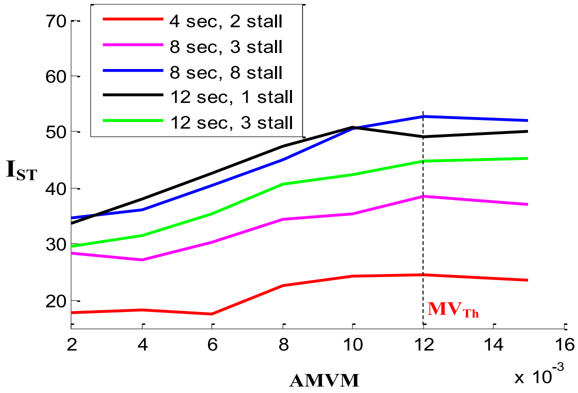


Fig. 15. Relationship between I_{ST} and video motion under different stall distributions.

Secondly, we consider how motion information (AMVM) will affect I_{ST}. Fig. 13 shows the stall impairment values with different video contents (from low motion news video to fast moving sports video). We can see that for a given video content, the observations (a) (b) still hold. Moreover, Fig. 15 shows the relation between stall impairment and the AMVM value. From Figs. 13 and 15, we have the following observation:

Observation (c): *For the same stall duration and stall number, the impairment due to stalling will increase as the motion (AMVM) increases. But after AMVM reaches a certain threshold (when the motion level is high enough), the impairment will not further increase.*

Observations (a), (b) and (c) tell us that we cannot use a linear equation to model the relationship between stall impairment with stall number, stall duration and AMVM. Therefore, we propose to use equation (3) as the impairment function for stall:

$$I_{ST} = \begin{cases} a * D_{ST} + b * N_{ST} - c * g(D_{ST}, N_{ST}) \\ \quad + d * AMVM \quad (if\ AMVM < MV_{Th}) \\ a * D_{ST} + b * N_{ST} - c * g(D_{ST}, N_{ST}) \\ \quad + d * MV_{Th} \quad (if\ AMVM >= MV_{Th}) \end{cases} \tag{3}$$

In equation (3), I_{ST} stands for the impairment due to stall, D_{ST} indicates the total duration of stall, N_{ST} stands for the number of stall. Function $g(D_{ST}, N_{ST})$ is used to compensate the simultaneous effects of stall duration and stall number and to match the phenomenon explained in observation (b). We use a piecewise function to ensure that once the AMVM exceeds threshold MV_{Th}, the stall impairment will not further increase. According to the results shown in Fig. 15, the threshold MV_{Th} is set to be 0.012.

In order to derive $g(D_{ST}, N_{ST})$ and the coefficients in equation (3), we start with randomly selecting 60% of the test results associated with stall impairment, and use them to train the model for I_{ST} (equation (3)). During the training, we use different types of formulas for $g(D_{ST}, N_{ST})$, including $k_1 * D_{ST} + k_2 * N_{ST}$, $D_{ST}^{k_1} * N_{ST}^{k_2}$ *and* $D_{ST}^{k_1} + N_{ST}^{k_2}$, and use non-linear regression to compute the coefficients in equation (3). Then we use the remaining 40% of the test results associated with stall impairment to validate the proposed I_{ST} function with all possible $g(D_{ST}, N_{ST})$ formulas. Finally we select the formula shown in equation (4), since it achieves highest correlation in the validation process.

$$g(D_{ST}, N_{ST}) = \sqrt{D_{ST} * N_{ST}} \tag{4}$$

The values of coefficients *a*, *b*, *c* and *d* in equation (3) are listed in Table VI.

TABLE VIII
SUBJECTIVE EXPERIMENT RESULTS FOR LEVEL VARIATION TESTS

TEST ID	16	17	18	19	20	21	22	23	24
IMPAIRMENT	33.5	24.4	32.9	21.3	44	28	13.6	19.6	12.2
TEST ID	25	26	27	28	29	30	31	32	33
IMPAIRMENT	29	34.3	22.7	41.5	24.8	23.8	13	10.2	19.3
TEST ID	34	35	36	37	38	39	40		
IMPAIRMENT	18.5	36.5	25.2	40.2	30.2	44.6	47.5		

D. Impairment Function for Level Variation

Level variation is the most complex factor to study, since it is difficult to characterize the complex patterns of level variations during a video session. As introduced in Section II, there are 3 dimensions for the 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 3 dimensions.

Table VIII shows the average evaluation of the impairment for test videos #16 ~ #40 (shown in Fig. 11). From the results we have the following observations:

Observation (d): All 3 dimensions of level variation factor will together affect user experience in a complex manner. For instance, comparing video #17 with video #20, both of them have an average level of 4.1, but the impairment of video #20 is significantly larger than that of #17. The same average level may lead to a completely different user experience, depending on the level fluctuation pattern. Therefore it may be difficult to reuse the method used for deriving I_{ID} and I_{ST} to also derive the impairment due to level variations.

Observation (e): The annoyance of staying at a low level (low quality) will grow exponentially with the duration that the low level is maintained. Comparing video #25 with #28, both of them have average level of 4.9 and similar amount of level switch magnitude, but video #25 has much smaller impairment than #28. This is because in video #25, when the level drops to the lowest value (level 1), it only lasts for about 2 seconds and then jumps up; in video #28, 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 as much. But if a low level is maintained for a long time (such as more than 10 seconds), people will feel great annoyance.

Observation (f): The impact of decreasing level switch is much larger than that of increasing switch. Comparing video #17 with video #36, they both have an average level of 4.1, but video #36 has much more impairment than video #17. This is because the level switches in video #17 are

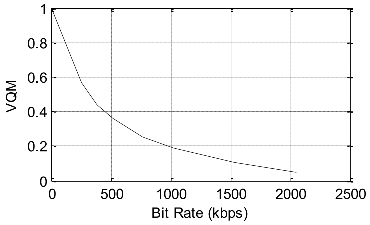


Fig. 16. Relationship between VQM value and bit rate.

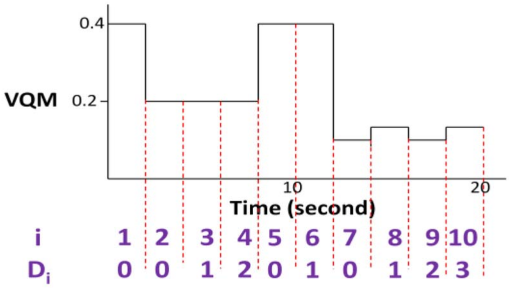


Fig. 17. D_i values and VQM values for a 20-second DASH video.

mostly increasing switches, while the switches in video #36 are mostly decreasing switches. Therefore, we cannot treat increasing switches and decreasing switches 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 the level value, then this impairment function cannot be applied generally.

Therefore, we propose to use VQM [3] 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.

Fig. 16 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 the 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, the 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 the number of consecutive segments that are right before segment i and have VQM value within range $[VQM_i - \mu, VQM_i + \mu]$. Parameter μ is heuristically set to be 0.05.

Fig. 17 shows an example of bit rate trace and the corresponding D_i values for a 20-second DASH video. In the y-axis we have converted bit rate into VQM value. As shown in Fig. 17, D_i is an integer that will accumulate if VQM remains constant or vary within a very small range. For example, for the 10th segment (when i equals 10), there are 3 consecutive segments before it that have VQM value between $VQM_{10} - \mu$, $VQM_{10} + \mu$, therefore D_{10} equals 3.

We model the first part of impairment (caused by low level itself) as:

$$P_1 = \frac{1}{N}\sum_{i=1}^{N} VQM_i^* e^{k^*T^*D_i} \tag{5}$$

As shown in equation (5), the P_1 value (impairment due to low level) is a weighted average of the VQM values of each video segment. The exponential term in equation (5), e^{k*T*D_i}, is used to comply with our observation (e) that the annoyance caused by a low level grows exponentially with the duration that the 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 (5) is used to control how fast the annoyance grows with time. The value of k is determined experimentally and listed in Table VI.

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

$$P_2 = \frac{1}{N}\sum_{i=1}^{N-1} |VQM_i - VQM_{i+1}|^{2} {}^* sign(VQM_{i+1} - VQM_i), \tag{6}$$

where

$$sign(x) = \begin{cases} 1, & x > 0 \\ 0, & otherwise \end{cases} \tag{7}$$

The value of P_2 is the average of the square of VQM differences between adjacent segments. According to our observation (f), the impairment caused by increasing switch is much smaller than that caused by decreasing switch. Therefore

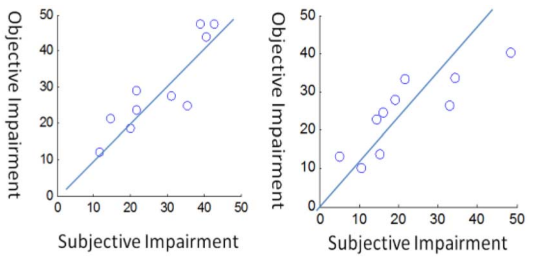


Fig. 18. Relationship between subjective and objective impairments: (a) left, first round of validation for impairment function I_{LV}; (b) right, second round of validation for impairment function I_{LV}.

in equation (6) we use sign function (equation (7)) to only consider 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$_2$:

$$I_{LV} = B_1{}^*P_1 + B_2{}^*P_2 \tag{8}$$

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 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 the impairment function, and use the rest for validation.

Fig. 18(a) and (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 lead to higher correlation. The corresponding coefficient values, B$_1$ and B$_2$ are derived using linear regression technique and are listed in Table VI.

V. Overall User Experience Model

In this section, we develop a DASH User Experience (DASH-UE) model which quantitatively measures the overall user experience, incorporating the impairment functions that we had developed in the previous section. We present results of another round of subjective experiments conducted to derive and validate the DASH-UE model.

We define DASH Mean Opinion Score (DASH-MOS) as a measurement metric for DASH-UE. Since DASH-MOS is determined by initial delay, stall and level variation factors as shown in Fig. 2, we attempt to formulate it using the impairment functions of these factors, similar to the framework of ITU-T E-Model [13]. ITU-T E-model is developed for audio transmission, where the multiple impairments (such as network delay impairment, audio distortion impairment, etc.) occur simultaneously. E-model offers a way to quantify the combined impact of these audio transmission impairments. Therefore, we borrow the framework of the E-model because our DASH-UE model also needs to quantify the combined effect of multiple impairments.

In ITU-T E-model, the Mean Opinion Score (MOS) is formulated by a transmission rating factor R [13]. We duplicate this function for our DASH-MOS formulation:

$$DASH - MOS = 1 + 0.035R + 7 \times 10^{-6}R(R-60)(100-R) \tag{9}$$

In (9), the transmission rating factor R takes value from range [0, 100] (the higher R, the better DASH-UE). DASH-MOS is related with R through nonlinear mapping, and it is within the range of [1, 4.5].

Although the framework of ITU-T E model is helpful for our study, the formula to compute R factor specified in ITU-T E model is specific to audio transmission and not suitable for DASH video streaming. Therefore, in this paper we propose to formulate the R factor as:

$$\begin{aligned} R &= F(I_{ID}, I_{ST}, I_{LV}) \\ &= 100 - I_{ID} - I_{ST} - I_{LV} \\ &\quad + \sum_{\substack{i,j \in \{ID,ST,LV\} \\ i \neq j}} f_{ij}\big(I_i, I_j\big)(R > 0) \end{aligned} \tag{10}$$

In (10), R is composed of impairment functions due to initial delay, stall, and level variation. These impairment functions have been derived in previous section. Term $f_{ij}(I_i, I_j)$ indicates the cross-effect between two different impairments and is used to compensate and adjust the R factor, because when several impairments happen simultaneously, the overall impairment will be different from the sum of each impairment.

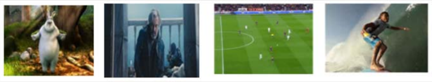
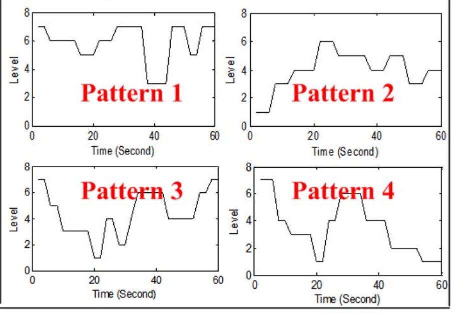
In order to derive the formula of function $f_{ij}(I_i, I_j)$, and also validate the accuracy of the overall DASH-UE model, we conduct another set of subjective quality assessment experiments with a new group of participants. In the following subsections, we will introduce the new subjective tests, analyze the collected test results, and derive and validate the DASH-UE model.

A. Second Round of Subjective Test

Another set of subjective tests has been carried out using a new panel of 47 subjects from UCSD and Qualcomm. The subjective test is still conducted in a controlled lab environment. Unlike the first round of test, this time each viewer is watching videos where the three artifacts (initial delay, stall and level variation) happen simultaneously.

Table IX lists the parameter values we use in this round of subjective tests. Similar to the first round of tests, each test video is about one minute long. We have selected 4 different video contents, including medium motion videos like animation and movie, and high motion videos such as soccer and surfing videos. The initial delay values vary between 2 to 10 seconds, which is a reasonable range considering the whole test video is one minute long. We tried 4 different stall distributions which lead to a wide range of stall impairment

TABLE IX
PARAMETERS FOR SECOND ROUND OF SUBJECTIVE TEST

Video Content	
Initial Delay	2s, 6s, 10s
Stall	(No Stall) , (2 stall, 2 sec) (2 stall, 8 sec) , (4 stall, 12 sec)
Level Variation	

values, from no stall to 4 stalls which add up to 12 seconds. Moreover, we include 4 different level variation patterns which exhibit very distinct impairments.

The experiment is divided into two sessions with a 10-minute comfort break between them. This adheres to the ITU-T recommendations that a continuous time period should not exceed half an hour to ensure a subject does not experience fatigue during the test.

The subjects evaluate the overall user experience of video quality (represented by the R-factor shown in equation (10)) according to the judging criteria shown in Table V. The R data obtained from the test was scanned for unreliable and inconsistent results. We used the ITU-T [12] criteria for screening subjective ratings which led to two subjects being rejected and their evaluations are eliminated. The scores from the rest of the subjects (45) were averaged to compute the overall user experience (R value) for each test condition.

B. Model Derivation and Validation

In this subsection, we will first present and analyze the test results. Based on the observations drawn from the results, we will then derive the DASH-UE model. Finally we will present validation results.

Fig. 19 shows the subjective R values for different stall impairment, I_{ST}, and different level variation impairment, I_{LV}, for a fixed initial delay value (here we show result when initial delay equals 2 seconds or 10 seconds). Note the values of I_{ST} and I_{LV} used in Fig. 18 are derived from the actual stall and level variations used in each test condition, using the stall and level variation impairment functions (equation (2) $\sim$ (8)). We can see that the overall user experience (R value) will drop as I_{ST} or I_{LV} increases. For a fixed level variation pattern (fixed I_{LV}), the R value will monotonically decrease when I_{ST}increases, and the same for I_{LV}. From Fig. 18, we have the following observation.

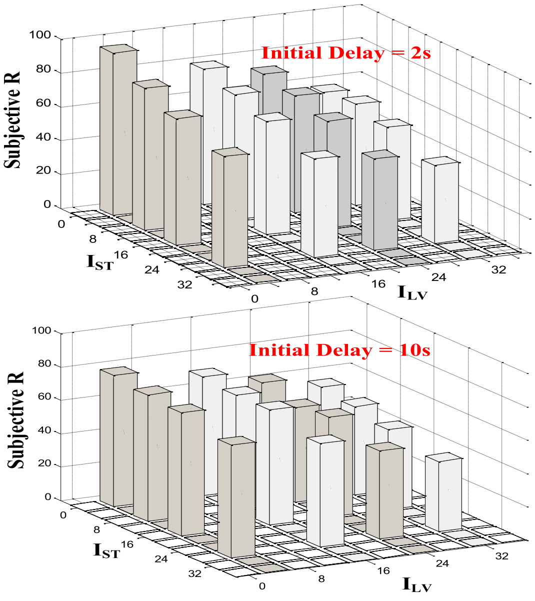


Fig. 19. Relation between subjective R scores and I_{ST} and I_{LV}.

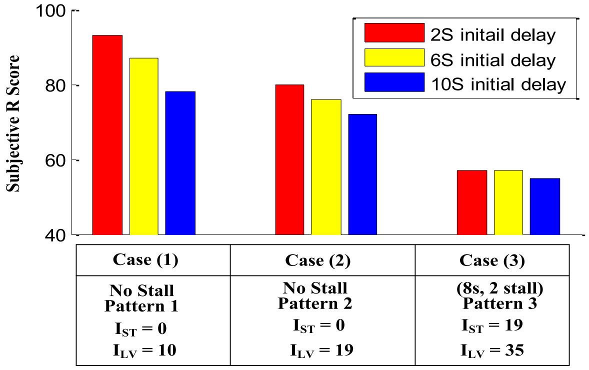


Fig. 20. Subjective R score under different initial delay, stall and level variation.

Observation (g): For a certain initial delay, both stall and level variation will affect the overall UE, no matter how big the initial delay is.

Fig. 20 shows the results from a different perspective. It shows how subjective R scores vary for different initial delay values under 3 sets of values of stall and level variation: (1) no stall, level variation pattern 1 (the shape of pattern 1 is shown in Table IX); (2) no stall, level variation pattern 2; and (3) two stalls which add up to 8 seconds, and level variation pattern 3. We can see that for case (1), when I_{ST} and I_{LV} are small, the R value will drop significantly when initial delay increases. For case (3), when I_{ST} and I_{LV} are large, there will not be significant difference in R value when initial delay increases from 2 seconds to 10 seconds. We can see that as I_{ST} and I_{LV} increase (from case (1) to cases (2) and (3)), initial delay will have less influence on overall user experience. This is due to the fact that people actually have higher tolerance for initial delay than stall and level variation. From Fig. 19 we find that:

Observation (h): When stall and level variation impairments are prominent, the subject may pay less attention to the impact caused by initial delay. On the other hand, when stall

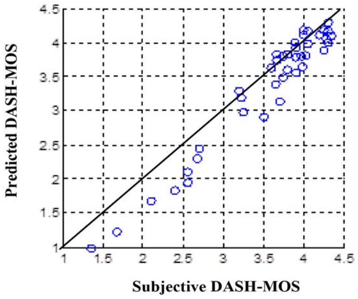


Fig. 21. Relation between predicted and subjective DASH-MOS.

and level variations are marginal, the impact caused by initial delay is more noticeable.

From observations (g) and (h), I_{ID}, I_{ST} and I_{LV} should not be treated as equally important in the formula $f_{ij}(I_i, I_j)$ of equation (10). The formulas of the compensation terms, $f_{ij}(I_i, I_j)$, should be derived such that: when I_{ST} and I_{LV} are large, the compensation terms associated with I_{ID}, ($f_{ID,ST}(I_{ID}, I_{ST})+ f_{ID,LV}(I_{ID}, I_{LV})$), should approximately cancel out the term $(-I_{ID})$ in equation (10), such that the impact of initial delay on overall user experience is marginal; on the other hand, when I_{ST} and I_{LV} are small, the compensation terms associated with I_{ID} should also be small, such that initial delay will have big impact on overall user experience.

Next, we randomly select 60% of the test results, and use them to train the model for R (equation (10)). During the training, we use different types of functions for $f_{ij}(I_i, I_j)$, including $(I_i + a \times I_j)^n$, $I_i^n \times I_j^m$, and $e^{(I_i + a \times I_j)^n}$, and use non-linear regression to compute the coefficients for the functions. Then we use the other 40% of test results to validate the proposed R model with all possible $f_{ij}(I_i, I_j)$ functions. Finally we select the function shown in equation (11), since it achieves the highest correlation in the model validation process.

$$R = 100 - I_{ID} - I_{ST} - I_{LV} + C_1 {}^* I_{ID}\sqrt{I_{ST} + I_{LV}} + C_2 {}^* \sqrt{I_{ST} {}^* I_{LV}} \quad (11)$$

In equation (11), coefficients C_1 equals 0.15, C_2 equals 0.82. Note that in the term $C_1 {}^* I_{ID}\sqrt{I_{ST} + I_{LV}}$ we have taken I_{ID} out of the square root to give more compensation for initial delay, which conforms to observation (h) that when I_{ST} and I_{LV} are large, people will ignore the impairment caused by I_{ID}.

Having derived the complete DASH-UE model (equations (1) ∼ (8), (11) and Table VI), we now use the rest of the subjective tests to validate the model. Fig. 21 shows the correlation between predicted DASH-MOS scores computed by the DASH-UE model (y-axis) and subjective DASH-MOS scores (x-axis) for each of the tests. From Fig. 21 we observe a high correlation of 0.91, and hence conclude that the proposed model can accurately predict user experience of DASH video.

VI. Application of DASH-UE Model to Long Videos

One limitation of the DASH-UE model is that it is derived based on 1-min long test videos, due to the limitation of number of test videos that each subject can watch. In this section, we propose a method to apply the DASH-UE model to long videos without modifying the model, and provide preliminary validation results for the accuracy of this method.

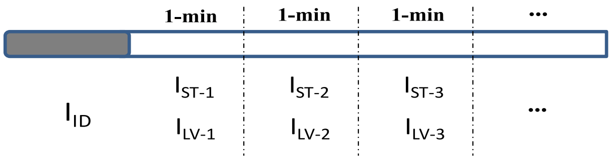


Fig. 22. Impairment values of each minute of video.

TABLE X
Two Approaches to Compute User Experience of Long Video

Approach A	$R = \frac{1}{N}\sum_{i=1}^{N} R_i$ Where: $R_i = \begin{cases} F(I_{ID}, I_{ST-1}, I_{LV-1}) & (if\ i = 1) \\ F(\ 0, I_{ST-1}, I_{LV-1}) & (if\ i \neq 1) \end{cases}$
Approach B	$R = F(I_{ID}, I_{ST-ave}, I_{LV-ave})$ Where: $I_{ST-ave} = \frac{1}{N}\sum_{i=1}^{N} I_{ST-i}$ $I_{LV-ave} = \frac{1}{N}\sum_{i=1}^{N} I_{LV-i}$

Since our DASH-UE model is derived based on 1-minute test videos, in the new approach, we propose to divide each video into 1-minute intervals, and record the stall and level variation pattern of each minute and calculate the corresponding impairment value. As shown in Fig. 22, we denote the stall impairment and level variation impairment during *i*-th minute as I_{ST-i} and I_{LV-i}, respectively. We propose two different approaches to measure the UE for the entire video, as listed in Table X. Approach A computes the entire video's user experience by taking the average of every minute's user experience. Alternatively, in approach B, we first calculate the average stall impairment (I_{ST-ave}) and level variation impairment (I_{LV-ave}) by taking the average of the impairments of each minute. The overall user experience (R) is then computed using the average stall and level variation impairment, together with the initial delay impairment.

Another problem is that the initial delay impairment (I_{ID}) is derived based on 1-min short video, and therefore it is not applicable for long form videos. For the same initial delay, the impairment on short video and long video would be different.

Therefore, we have conducted another round of subjective test with long videos using 24 subjects from UCSD. This test consists of two parts. The first part is used for deriving initial delay impairment function (I_{ID}) for long video. The second part is used for validating the proposed approaches A and B.

The first part of the test is similar to the test for deriving I_{ID} (described in Section IV-A). We ask the viewers to watch a 1-min video with different initial delay values including 3, 6, 10, 15, 25 seconds. But this time instead of evaluating the initial delay impairment on the 1-min test video, we ask the viewers how big the impairment would be assuming they

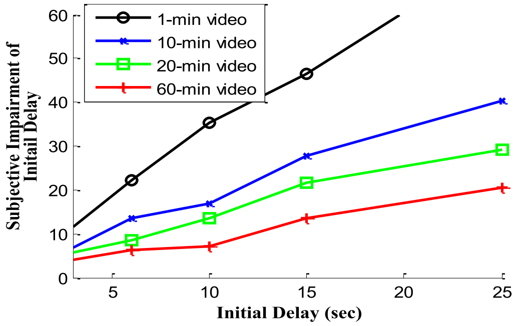


Fig. 23. Subjective evaluations of initial delay impairment for different video length.

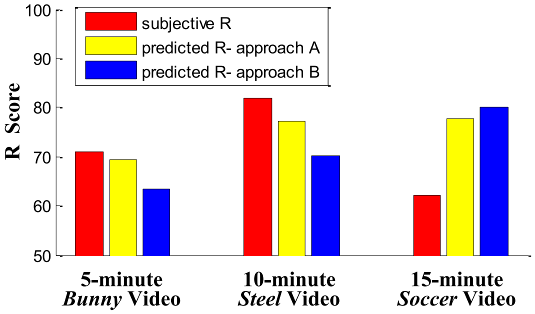


Fig. 24. Relation between subjective R scores and predicted scores using two approaches for test videos.

are watching a 10-min, 20-min or 60-min video. This evaluation is done immediately after the viewers finish watching the 1-min test video and still clearly remember the amount of annoyance they experienced due to initial delay. The results are shown in Fig. 23.

Based on the results shown in Fig. 23, we apply regression technique and adjust the initial delay impairment function $\mathrm{I_{ID}}$ (originally proposed in equation (1)) as:

$$I_{ID} = \min\left\{3.2 * \frac{L_{ID}}{1 + \ln(0.8 + 0.2 * \mathrm{L_{total}})}, 100\right\} \quad (12)$$

After deriving $\mathrm{I_{ID}}$ for long video, we conduct the second part of subjective test in which we ask each subject to watch 3 videos, with duration of 5 minutes, 10 minutes, and 15 minutes respectively. For these 3 videos, we have distributed all 3 artifacts (initial delay, stall, level variation) simultaneously in every minute of the video. We have also selected different content for the 3 videos, including *Bunny (for 5 minutes video)*, *Steel (for 10 minutes video)* and *Soccer (for 15 minutes video)*, which are described in Table II. After collecting all the participants' evaluations, we then compare the subjective evaluation of the overall UE (R value), with the predicted R values using approaches A and B.

Fig. 24 shows the mean value of the subjective R value (given by subjects) and the predicted R values using approach A and approach B. We can see that: 1) approach A will always lead to a predicted R value closer to the subjects' evaluation compared to approach B; 2) for the 5 and 10 minute videos, the difference between the predicted R value (using approach A) and subjective R value are both below 10, out of a 100 scale, showing high correlation; 3) for the 15 minute video, neither of the approaches A or B can lead to accurate prediction.

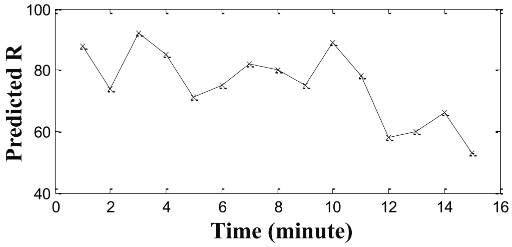


Fig. 25. Trace of predicted R scores for 15-minute *Soccer* test video.

The validation results show that for videos up to 10 minutes long, approach A can provide adequate prediction accuracy. According to surveys conducted by comScore, the average length of online videos is about 6.4 minutes long [20]. Therefore we claim that the proposed method (approach A) can be applied to most of the online videos.

However, for videos that are longer (more than 10 minutes long), our approach may not be applied directly. For these videos, instead of providing one UE prediction score for the entire video, we could use the DASH-UE model to provide a minute-by-minute UE score trace. For example, we show in Fig. 25 the predicted UE score trace of the 15-min *Soccer* test video.

Fig. 25 also provides us insight into why the prediction accuracy of approaches A and B may be low for longer videos, like the 15-minute Soccer video (shown in Fig. 24). We notice from Fig. 25 that the last few minute intervals have much higher impairment (lower predicted R) than the previous intervals. Since the quality (impairment) experienced in the latter intervals of the video may have higher impact on a viewer's assessment than the quality experienced in the earlier parts, this helps explain the low subjective score 62 seen in Fig. 24. However, our proposed approaches A and B give equal weight to the impairment of each minute interval, and hence predict much higher scores than the subjective score as seen in Fig. 24. Hence, one way of improving prediction accuracy for long videos may be to use unequal weights for each time interval, with the weights increasing with increasing intervals. The latter may provide a single UE score for the entire video with sufficient accuracy, as an alternative when a single score is more desired than a minute-by-minute UE scores as suggested earlier for long videos. In general, quality assessment of long videos should be investigated further as part of future work, including the above suggested possible approach.

VII. Conclusion

In this paper, we have presented a novel user experience model which can quantitatively measure the user experience of DASH video, by taking into account both spatial and temporal artifacts. We first investigate 3 factors which will impact user experience: *initial delay*, *stall* and *level variation*. Secondly, we design and conduct subjective experiments to derive the impairment function for each of the factors. Thirdly, we combine the 3 impairment functions to formulate an overall user experience model by conducting another round of subjective tests in which subjects evaluate video quality when they experience combined artifacts. Finally, we demonstrate applicability of the proposed model to videos up to 10 minutes, longer than the average length of online videos.

The proposed user experience model does not need to access the uncompressed video source and hence can be conveniently incorporated into DASH client software to quantify user experience in real time. Moreover, the proposed model can be used by a DASH service provider to monitor and control the quality of service, as well as optimize the DASH rate adaptation algorithm.

Although the proposed DASH-UE model has considered some video content features such as motion, there are other factors related to the video content and the context of the video viewed, such as the popularity of the video and the type of device the video is watched on, which may impact user experience. For example, it is possible that a viewer will have a different level of tolerance with a video that is interesting to him/her, versus some other less appealing videos. In the future, we plan to study how these other factors will affect user experience and extend our DASH-UE model to consider them. Also, as suggested in the previous section, we plan to study and determine user experience modeling for long videos, including the unequal weight based approach suggested earlier.

Furthermore, another possible extension of this DASH-UE model is to consider the saliency information of video frames such that the important regions of a frame have larger impact on user experience than non-important regions. More specifically, we can first apply the techniques proposed in [21]–[23] to detect the salient regions of video frames. Subsequently, based on the saliency information, we can develop a saliency-based video frame quality metric and then replace the VQM metric in I_{LV} function (equation (5)∼(8)) with this new metric. One can expect the modeling accuracy will be improved by incorporating the saliency information into the DASH-UE model.

References

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast*, Cisco, San Jose, CA, USA, 2013.

[2] *Dynamic Adaptive Streaming Over HTTP, w11578*, ISO/IEC JTC 1/SC 29/WG 11 Standard CD 23001-6, 2010.

[3] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[4] J. Ozer, "Adaptive streaming in the field," *Stream. Media Mag.*, vol. 7, no. 6, pp. 36–49, Jan. 2011.

[5] 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.

[6] M. Ries, O. Nemethova, and M. Rupp, "Video quality estimation for mobile H.264/AVC video streaming," *J. Commun.*, vol. 3, no. 1, pp. 41–50, Jan. 2008.

[7] Y.-C. Lin, D. Varodayan, and B. Girod, "Video quality monitoring for mobile multicast peers using distributed source coding," in *Proc. 5th Int. Mobile Multimedia Commun. Conf.*, London, U.K., 2009, pp. 1–6.

[8] R. Mok and E. Chan, "Measuring the quality of experience of HTTP video streaming," in *Proc. IEEE Int. Symp. Integr. Netw. Manage.*, Dublin, Ireland, 2011, pp. 485–492.

[9] K. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC," in *Proc. Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, 2012, pp. 127–131.

[10] P. Ni, R. Eg, and A. Eichhorn, "Flicker effects in adaptive video streaming to handheld devices," in *Proc. ACM Int. Multimedia Conf. (MM)*, Scottsdale, AZ, USA, 2011, pp. 463–472.

[11] P. Ni, R. Eg, and A. Eichhorn, "Spatial flicker effect in video scaling," in *Proc. Int. Workshop Qual. Multimedia Exp.*, Mechelen, Belgium, 2011, pp. 55–60.

[12] *Methodology for Subjective Assessment of the Quality of Television Picture*, document BT-500-11, Int. Telecomm. Union, Geneva, Switzerland, 2002.

[13] *The E-Model: A Computational Model for Use in Transmission Planning*, document G.107, ITU-T, Geneva, Switzerland, Mar. 2005.

[14] A. Khan, L. Sun, and E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 431–442, Apr. 2012.

[15] G. W. Cermak, "Subjective video quality as a function of bit rate, frame rate, packet loss rate and codec," in *Proc. 1st Int. Workshop Qual. Multimedia Exp. (QoMEX)*, San Diego, CA, USA, Jul. 2009, pp. 41–46.

[16] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 513–516, Apr. 2010.

[17] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[18] K. Pessemier, K. Moor, and W. Joseph, "Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 47–61, Mar. 2013.

[19] M. Mok, E. Chan, and R. Chang, "Measuring the quality of experience of HTTP video streaming," in *Proc. IEEE Int. Symp. Integr. Netw. Manage.*, Dublin, Ireland, May 2011, pp. 485–492.

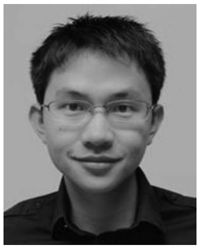
[20] *Video Length Statistics by ComScore*. [Online]. Available: http://www.tubefilter.com/2012/05/12/average-length-online-video/, accessed Feb. 12, 2014.

[21] Y. Tong, F. A. Cheikh, H. Konik, and A. Tremeau, "Full reference image quality assessment based on saliency map analysis," *Int. J. Imag. Sci. Technol.*, vol. 54, no. 3, 2010, Art. ID 030503.

[22] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[23] Y. Tong, F. A. Cheikh, F. F. E. Guraya, H. Konik, and A. Tremeau, "A spatiotemporal saliency model for video surveillance," *J. Cogn. Comput.*, vol. 3, no. 1, pp. 241–263, 2011.

[24] *Ffmpeg Software*. [Online]. Available: https://www.ffmpeg.org, accessed Mar. 25, 2014.



Yao Liu is currently pursuing the Ph.D. degree of electrical and computer engineering with the University of California, San Diego. His industry experiences include internship at Qualcomm in 2010 and Yahoo in 2013. His research interests include mobile multimedia, wireless communication, and mobile cloud computing.

Sujit Dey (SM'03–F'14) received the Ph.D. degree in computer science from Duke University, Durham, NC, USA, in 1991. He is a Professor with the Department of Electrical and Computer Engineering, University of California, San Diego (UCSD), where he is the Head of the Mobile Systems Design Laboratory. He is the Director of the UCSD Center for Wireless Communications. He also serves as the Faculty Director of the von Liebig Entrepreneurism Center. He founded Ortiva Wireless in 2004, where he served as its Founding CEO and later as CTO till its acquisition by Allot Communications in 2012. He has co-authored over 200 publications, including journal and conference papers, and a book on low-power design. He is the co-inventor of 18 U.S. patents, resulting in multiple technology licensing and commercialization.

Fatih Ulupinar received the Ph.D. degree of electrical engineering with the University of Southern California in 1991. Since then he has worked with Bilkent University, Turkey, Advanced Computer Science Corporation, Los Angeles, and Siemens Corporate Research, New Jersey. He is currently with Qualcomm Research, where he has worked on the design and implementation of many of the wireless WAN systems, such as EvDO, UMB, and LTE. He is working on end-to-end large data transfer and efficient video streaming. He holds over 90 patents.

Michael Luby received the B.Sc. degree in applied mathematics from the M.I.T. and the Ph.D. degree in theoretical computer science from the UC Berkeley. He is the Vice President of Technology, Qualcomm Inc., focusing on advanced research, including broadcast multimedia delivery, Internet streaming, and reliable distributed storage. He was a recipient of the IEEE Richard W. Hamming Medal, the ACM SIGCOMM Test of Time Award, the IEEE Eric E. Sumner Communications Theory Award, and the ACM SIAM Outstanding Paper Prize, for his work in coding theory, cryptography, and content delivery technologies. He is a member of the National Academy of Engineering.

Yinian Mao received the Ph.D. degree in electrical engineering from the University of Maryland in 2006, and the B.S.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2001. He is a Senior Staff Engineer with Qualcomm Research, San Diego. His research interests are in the broad area of mobile communications and computing, multimedia signal processing, and information security. He has co-authored over ten peer-reviewed papers and holds over 30 U.S. and world patents.