RESEARCH **Open Access**

The impact of video-quality-level switching on user quality of experience in dynamic adaptive streaming over HTTP

Demóstenes Z Rodríguez[1*], Zhou Wang[2], Renata L Rosa[3] and Graça Bressan[3]

Abstract

Dynamic adaptive streaming over HTTP (DASH) has become a promising solution for video delivery services over the Internet in the last few years. Currently, several video content providers use the DASH solution to improve the users' quality of experience (QoE) by automatically switching video quality levels (VQLs) according to the network status. However, the frequency of switching events between different VQLs during a video streaming session may disturb the user's visual attention and therefore affect the user's QoE. As one of the first attempts to characterize the impact of VQL switching on the user's QoE, we carried out a series of subjective tests, which show that there is a correlation between the user QoE and the frequency, type, and temporal location of the switching events. We propose a novel parameter named switching degradation factor (SDF) to capture such correlation. A DASH algorithm with SDF parameter is compared with the same algorithm without SDF. The results demonstrate that the SDF parameter significantly improves the user's QoE, especially when network conditions vary frequently.

Keywords: Video streaming; Adaptive streaming; DASH; Video quality; Subjective test; Quality of experience; Switching degradation factor

1 Introduction

IP network uses the concept of best-effort delivery, where the network does not guarantee the data arrival to the end user at the right time and order, depending on the network traffic load. However, many services, such as video streaming, run over IP networks, where transport layer protocols attempt to improve the IP network performance and, consequently, the end users' quality of experience (QoE). One of these protocols is the widely adopted Transmission Control Protocol (TCP) that supports reliable end-to-end data delivery.

In the last years, video traffic has been increasing dramatically because many video services over the Internet gained popularity. The large number of wireless devices that use video services via mobile networks is one of the major contributors to the growth of video traffic. Currently, most video streaming services run over HyperText Transfer Protocol (HTTP) that uses TCP as the transport layer

protocol, which is not intercepted or blocked by firewalls or network address translation (NAT), as is the case for User Datagram Protocol (UDP). Moreover, HTTP-based delivery provides reliability and deployment simplicity due to HTTP and TCP protocols, which are widely implemented [1].

Video quality assessment and, therefore, users' QoE evaluation is relevant due to the large number of video services offered nowadays. Subjective test of video quality assessment are conducted to determine the user's satisfaction based on which video services may be improved [2]. These tests are generally performed under laboratory conditions. Nevertheless, in recent years, some studies [3,4] show the possibility to perform image or video quality assessment through remote assessors using the Internet.

In recent years, dynamic adaptive streaming over HTTP (DASH) standard [5] has gained popularity. The purpose of DASH is to improve the end user's QoE using a video streaming service. Several video content providers adopted different DASH solutions introducing client and server software, and the most sophisticated consumer electronic devices are expected to support it

* Correspondence: demostenes.zegarra@dcc.ufla.br
[1]Department of Computation Science, University of Lavras, Câmpus Universitário, Caixa Postal 3037, CEP 37200-000, Lavras, Minas Gerais, Brazil
Full list of author information is available at the end of the article

[6]. A performance comparison of different adaptation algorithms programmed in the most popular DASH commercial solution is presented in [7]. It is worth noting that the DASH solution uses a video signal quality level determined at the users' devices. Studies in other areas are making similar efforts. Examples include the 3rd Generation Partnership Project (3GPP) named minimization of drive test (MDT) [8]. The DASH solution uses an adaptation control algorithm to determine the most appropriate video segment to be transmitted according to some network and/or application layer parameters, which reflects the video signal quality to the user's device.

As stated before, DASH intends to improve the users' QoE because they receive the best VQL allowed by network conditions. However, if the network condition changes constantly, the DASH adaptation control algorithms will switch between different VQLs. As a consequence, the user may experience multiple changes in the video presentation in a short time period, thereby affecting the user's QoE. In this research, each VQL is classified by its temporal and spatial resolutions. Different VQL switching types are considered depending on the video encoding characteristics. Switching events between videos with different spatial or temporal resolutions have different impacts on visual attention and user QoE. Hence, different effects on the overall user's QoE are also expected.

The main purpose of this work is to quantitatively determine how the VQL switching events affect the user's QoE in a DASH scenario. This fact stresses the relevance to include, in the DASH algorithms, a decision parameter we named switching degradation factor (SDF) that changes with the VQL switching types, the frequency of VQL switching events, and their temporal locations. Subsequently, improved DASH algorithms are obtained by performing VQL switchings depending on SDF values. Furthermore, this concept can be extended to other bit rate adaptation applications, such as scalable video coding (SVC) [9].

The remainder of this paper is structured as follows: Section 2 presents an overview of the DASH solution, quality adaptation, and visual quality assessment methods. Section 3 describes the quality degradation factors in VQL switching events. Section 4 introduces the proposed SDF parameter. Section 5 illustrates the test environment, implementation, and the results, highlighting the importance of considering the SDF parameter as a decision factor in the DASH algorithm. Finally, Section 6 draws the conclusions.

2 Overview of DASH, quality adaptation, and visual quality assessment methods

DASH is a new standard developed by 3GPP and MPEG [4,5,10] aiming to encode video files using different encoder parameters. Different versions of the same video are obtained and stored in a video server; in which each video version represents a different VQL. In MPEG DASH, the metadata is named media presentation description (MPD). In the DASH solution, the MPD and media are delivered by the HTTP protocol. Each video version stored in the server is logically divided into video segments. A video segment can be represented as a small video file with its own MPD in the file header. The MPD maps the video segment position to the time of the complete video. Thus, the client can access a specific video segment. A general description of a DASH system is shown in Figure 1, in which four versions of the same video with different spatial resolutions are stored in the video server (VQL_A to VQL_D). The video segments are represented by the letter S, for instance, the first segment of VQL_A denoted by SA1. In Figure 1, a DASH control algorithm is employed at the client side. This algorithm uses network parameters as inputs, most commonly the throughput connection, to determine the segment quality level to be downloaded.

In the last 2 to 3 years, a number of adaptation control algorithms have been proposed. These algorithms are typically based on parameters such as available bandwidth [11,12], throughput [13-15], round-trip time (RTT), the average download bit rate, the number and frequency of pauses during a time interval [16,12] that are related with buffering events [17], and the delay associated with user interactivity [18]. In [13], an architecture for DASH in a content distribution network (CDN) scenario is studied. In [19,20], the user perception of adapting video quality is studied. In [19], different test scenarios of a quality upgrade are evaluated in order to determine the optimal adaptation trajectory, but the user QoE degradation is not quantitatively measured; thereby, the results cannot be directly used in a DASH control algorithm. Also, the temporal locations of VQL switching events are not considered.

In the Internet world, smooth transmission of video data has become one of the most challenging problems [21]. If there is a sudden change in video quality during a video streaming session, a common practice in DASH quality adaptation, the visual QoE may be negatively affected. In particular, when the visual system adapts to a specific quality level at specific spatial and temporal resolutions, sudden changes in the quality level may trigger unwilling eye activities such as refocusing and eye movement, which could be distractive to human attention from the video content, resulting in unpleasant QoE. Our preliminary subjective test presented in Figure 2 also suggests that different types of VQLs may have different impacts on visual QoE. Specifically, two types of 1-min videos are shown to the subjects, one contains switching events with different temporal resolutions only and the other with different spatial resolutions
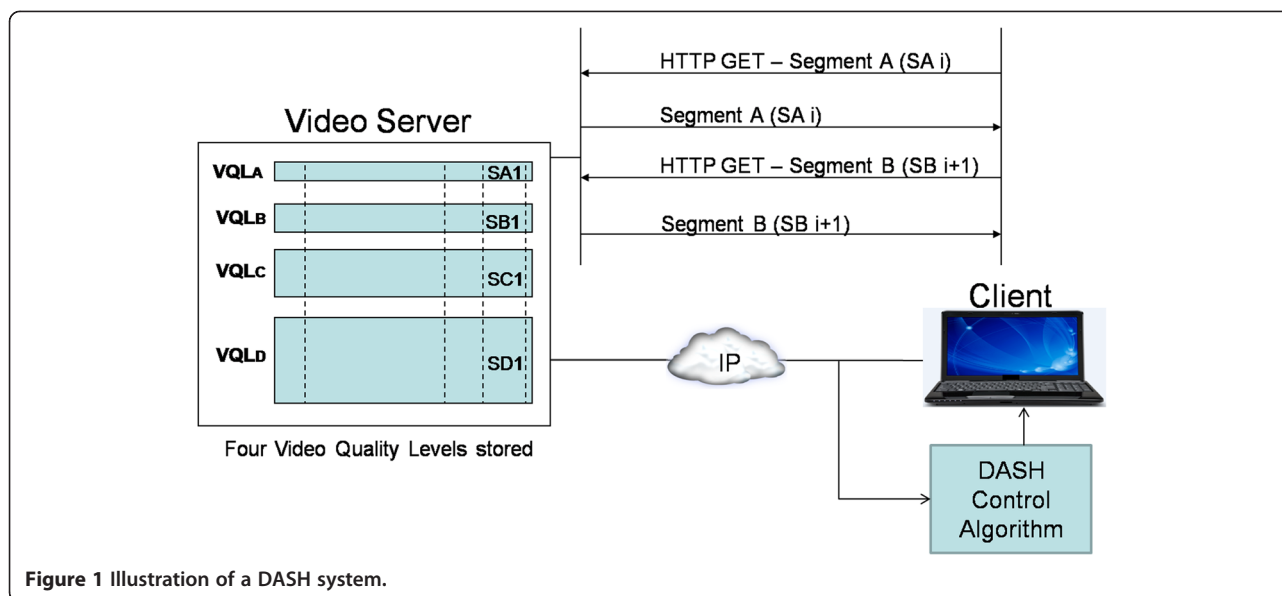
Figure 1 Illustration of a DASH system.

only. There are two useful observations from Figure 2. First, the negative effect on visual QoE, gauged using the mean opinion score (MOS), increases with the frequency of VQL switching. When the frequency is less than 1/16 per second, the effect is minimal, and when the frequency is higher than 1/14 per second, significant drops in MOS values are observed. Second, the negative impact of VQL switching in spatial resolution is much stronger than that in temporal resolution. These observations suggest that to achieve optimal QoE, network quality adaptation techniques should take into account both the frequency and types of VQL switching events. Unfortunately, this has not been well accounted for in state-of-the-art DASH algorithms

Visual attention, context awareness, and assessment of users' expectations play an essential role in determining the user's QoE. The assessment of QoE should include objective human cognitive aspects and incorporate some valid psychological subjective and social approaches [22]; thus, the study is multi-disciplinary in nature, incorporating psychology, cognitive science, sociology, and information technology [23]. It is worth noting that during the subjective test, the evaluators' attention is also predominantly selective to the video content being watched. Hence, the experimental test environment needs to be isolated from external stimuli such as visual or audible noise that could interfere with the evaluators' attention.
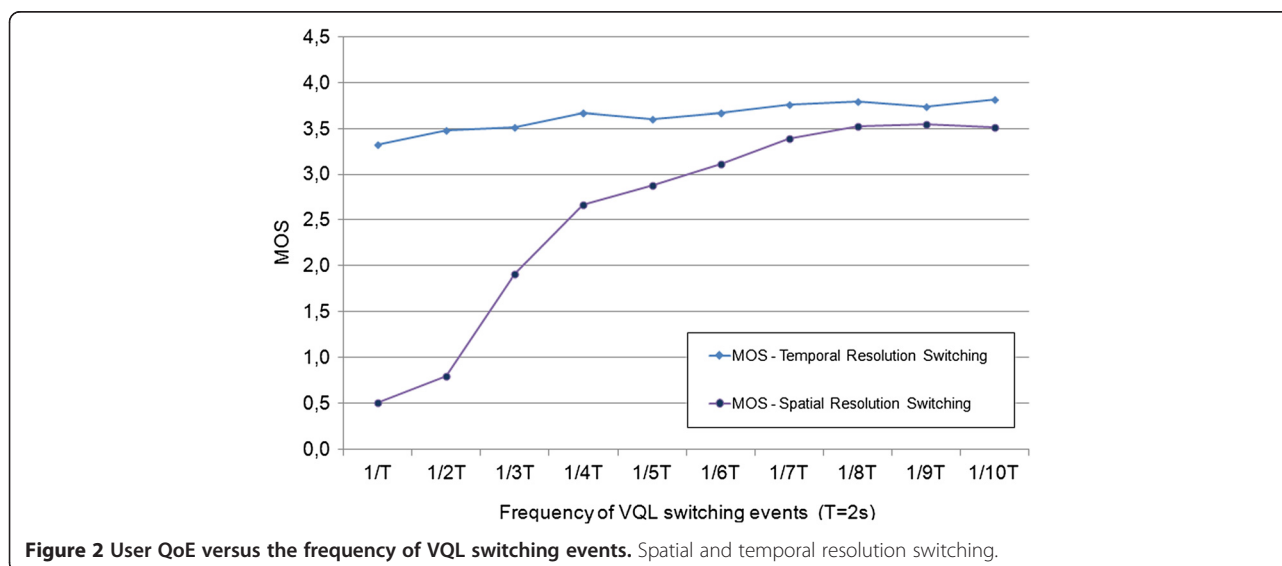

Figure 2 User QoE versus the frequency of VQL switching events. Spatial and temporal resolution switching.

A number of standard subjective testing methodologies recommended by ITU are described in ITU-R BT-500 [24] and ITU-T P.910 [25]. The methodologies in ITU-R BT-500 include double-stimulus continuous quality scale (DSCQS), double-stimulus impairment scale (DSIS), single-stimulus continuous quality evaluation (SSCQE), and simultaneous double stimulus for continuous evaluation (SDSCE). The methodologies in ITU-T P.910 include absolute category rating (ACR), degradation category rating (DCR), absolute category rating with hidden reference (ACR-H), and paired comparison (PC). In this work, we adopt the ACR approach with a 5-point MOS scale recommended in ITU-T P.910, as shown in Table 1.

3 Quality degradation factors in VQL switching

In order to have a better understanding of the impact of VQL switching on visual QoE, here, we elaborate the key issues that have not been fully accounted for in the current DASH quality adaptation control algorithms.

3.1 Frequency of VQL switching events

Considering the changes in network conditions and buffer status, the DASH controller can react in two ways, a switch up (SU) or a switch down (SD) of VQL. The former happens when the bandwidth allows the client to require a higher VQL from the server, and the latter occurs when the bandwidth is not sufficient and it is necessary to perform a downgrade in VQL to avoid interruptions or delays in video transmission.

Figure 3 presents a simple illustrative two-VQL scenario, named Scenario A, where VQL_A and VQL_B represent the high- and low-quality levels, respectively. This scenario contains several VQL switching events and no VQL switching before timestamp T0 is assumed. There are eight time intervals (e.g., the first time interval is from timestamp T0 to T1), each one with t-second duration. Within each interval, the same VQL is maintained, and after this interval, a VQL switching event can occur. In DASH applications, this time interval (t) represents a video segment length that has only a VQL. In order to examine the frequency of VQL switching events, we would need to first define a sliding observation window that shifts with time. For illustrative purpose only, here, we give

an example by defining the size of the sliding window to be

$$T = 4t \tag{1}$$

We have chosen an observation window size of $4t$ because the total time range presented in Figure 3 is $8t$, permitting a good visualization of the first two windows, stressing that this value is only for clarification purposes. Let N_S and F_S denote the number of VQL switching events and their frequency within the sliding observation window, respectively. F_S and N_S are related by

$$F_S = \frac{N_S}{T} \tag{2}$$

In addition to F_S, the network and buffer status can be either good (G), equal (E), or bad (B), and the reaction of the DASH algorithm can be either SU, SD, or no action. Table 2 describes the behavior of scenario A, where the network and buffer status can be complemented with other application layer parameters as inputs to the DASH algorithm.

The current DASH algorithms only consider the network and/or application layer parameters, without taking into account the negative QoE effect caused by VQL switching. As a result, VQL switching is triggered at every timestamp, as can be seen in Table 2. To give an example about how the parameter F_S could be used to avoid too frequent VQL switching events, we define a simple improved algorithm that adds F_S as a decision factor (where a F_S threshold of 1/2 is selected merely to give an example), and the improved algorithm is summarized in Table 3.

Figure 4 plots the case of scenario B when the improved algorithm defined in Table 3 is applied. In addition, Table 4 elaborates the behaviors of the scenario. As expected, the number of VQL switching events is significantly reduced because the DASH algorithm is complemented by prohibiting any SU event as long as the F_S parameter is above the threshold 1/2.

3.2 Types of VQL switching events

In a DASH scenario, there are often more than two versions of the same video available in the video server. Therefore, there could be many more types of VQL switching events, as opposed to only SD and SU in scenarios A and B.

Figure 5 depicts scenario C, in which there are five VQLs and VQL_A and VQL_E represent the highest and the lowest quality levels, respectively. Since VQL switching can occur between any of the five VQLs, there are multiple possible types of switching events, each of which could affect the user QoE in a different way. Therefore, it is desirable to investigate how to quantify the impact of

Table 1 ITU-T 5-point scale - ACR

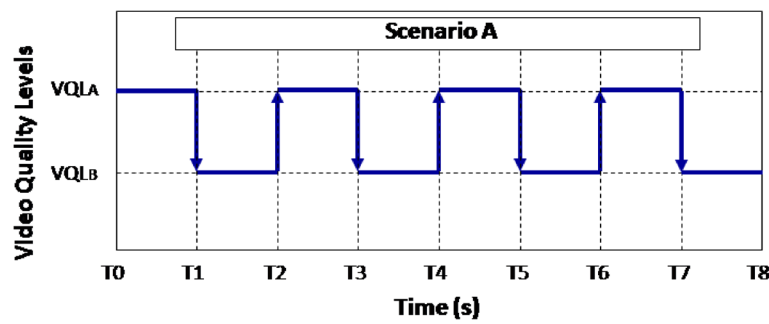| Grading value	Estimated quality	Perceived impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Figure 3 Scenario A. VQL switching events using a DASH algorithm without considering quality degradation caused by VQL switching.

each switching event type on the overall QoE and how to embed such information in the design of DASH quality adaptation algorithms.

3.3 Temporal location of VQL switching events

Another factor that may affect the user QoE is the temporal locations of the VQL switching events. An example is given in Figure 6, where in scenario D, the switching events all occur at the beginning of the session, while in scenario E, all switching events are near the end of the session. Current DASH algorithms do not consider the temporal location of the switching events and give the same degradation weight to both scenarios. This may not be able to precisely account for their actual impacts on the user QoE, which may be affected by psychological factors such as the memory effect.

4 Quality degradation model for VQL switching

Preliminary subjective test results of video quality assessment demonstrated that the users' QoE is affected by the three key quality degradation factors (frequency, type, and temporal location of switching events) related to VQL switching, as elaborated in the three scenarios presented in the previous section. These factors have not been well accounted for in the current DASH algorithms. In this section, we propose a novel SDF, which combines the aforementioned three factors. Parameters in SDF are calibrated using subjective testing data. An improved DASH algorithm is then proposed by incorporating SDF as a decision factor.

Table 2 VQL switching events in scenario A using a DASH algorithm without considering the frequency of switching events

Parameter	Time							
	T0	T1	T2	T3	T4	T5	T6	T7
Frequency of switching events (F_S)	0	0	1/4	1/2	3/4	1	1	1
Network and buffer status	E	B	G	B	G	B	G	B
Current DASH algorithm output	-	SD	SU	SD	SU	SD	SU	SD

For illustration purpose, we will use a specific example in our description of SDF, though the formulation of SDF is applicable to the general scenarios. Assuming there are six versions of the same video, namely V_A, V_B, V_C, V_D, V_E, and V_F in which V_A and V_F represent the highest and the lowest VQLs, respectively. We name a VQL switching between two videos with different spatial resolutions but the same temporal resolution a spatial resolution switching (SRS), and a VQL switching between different temporal resolutions but the same spatial resolution a temporal resolution switching (TRS). Considering the scenario 'C' presented in Figure 5, each VQL switching type i can affect the overall user QoE in a different manner, and we thus associate it with a different weight $w_i^{(T)}$ that quantifies its importance to the user QoE. Table 5 gives an example of six VQL switching types used in our tests.

As presented in Figure 6, the switching events at different temporal locations (e.g., the beginning, middle, and end part of the video) may have different impacts on the overall user QoE; we divide the video into segments, each segment associated with a segmentation index j and a weight $w_j^{(S)}$ that indicates its importance to the overall user QoE.

Table 3 Algorithm 1: frequency of switching events as a decision factor in DASH quality adaptation algorithm

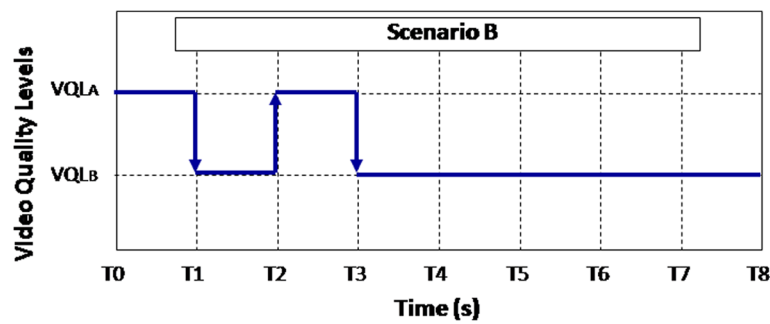| Line	Statement
1	*Fs* < = frequency of switching events
2	VQL$_n$ < = Current Video Quality Level
3	*DASH_Out* < = Output of DASH algorithm (SU, SD or same video quality)
4	*if* (DASH_Out = SD) then
5	$VQL_n = VQL_{n-1}$
6	*if* (***Fs*** < ½ & DASH_Out = SU) then
7	$VQL_n = VQL_{n+1}$
8	*Else*
9	$VQL_n = VQL_n$
10	*end if*

Figure 4 Scenario B. An example of VQL switching events using DASH quality control considering parameter *F*s.

As depicted in Figures 3 and 4 that introduced the scenarios A and B, respectively, during a time period (T) may occur some switching events (N) between different VQLs and located in different instants in the temporal domain.

With all these considerations, we can now define the SDF as

$$\text{SDF} = \frac{1}{T}\sum_{j=1}^{n} w_j^{(S)}\left(\sum_{i=1}^{m} w_i^{(T)} N_{ij}\right) \tag{3}$$

where the parameters are summarized as follows:

- m: number of VQL switching types
- n: number of temporal segments
- N_{ij}: number of VQL switching events of type i during temporal segment j
- $w_i^{(T)}$: weight factor associated with switching type i
- $w_j^{(S)}$: weight factor associated with temporal segment j
- T: duration of the time window being observed

For better understanding of SDF, it is useful to map it to a new scale, so that it can be directly used to predict how VQL switching events change the 5-point scale MOS values. Motivated by previous works on the QoE of multimedia service [26-28], we adopt an exponential function for the mapping, which is given by

$$\overline{\text{SDF}} = Ce^{\text{SDF}} \tag{4}$$

where C is a positive constant that adjusts the speed of the exponential function.

It remains to determine the parameters in the SDF model, including the weighting factors $w_i^{(T)}$ and $w_j^{(S)}$ in (3) for each switching type and temporal segment, as well as the constant C in (4). To do this, we carried out two phases of subjective tests. In the first phase, K test scenarios (specifically, $K = 24$ in our experiment, because we considered six switching types in our tests, resulting an average of four scenarios for each switching type) were used to determine the $w_i^{(T)}$ parameters only. Once the $w_i^{(T)}$ parameters are fixed, a second phase of test is conducted to obtain the $w_j^{(S)}$ parameters. The lengths of the video used in phase 1 were 1 min. In each test scenario, a different set of VQL switching events with different switching types were used.

In phase 1, there is only one temporal segment, i.e., $n = 1$ (though it could still contain multiple VQL switching events). In the k-th scenario, the net impact of VQL switching events on the overall user QoE or the desired $\overline{\text{SDF}}$ factor (denoted by $\overline{\text{SDF}}_k^{(D)}$) would be the difference between the mean of the MOS values of all individual VQLs that are transmitted within the video segment used in the k-th scenario (denoted by $\text{MOS}_k^{\text{mean}}$, which is independent of VQL switching) and the MOS value given to the whole segment (denoted by MOS_k, which is certainly affected by VQL switching, if any). For this, each VQL needs to have an MOS score previously defined, and from this information, only the MOS scores of the VQLs transmitted are used to calculate the $\text{MOS}_k^{\text{mean}}$. Thus, we have

Table 4 VQL switching events in scenario B using a DASH algorithm considering the frequency of switching events

Parameter	Time							
	T0	T1	T2	T3	T4	T5	T6	T7
Frequency of switching events (F_S)	0	0	1/4	1/2	3/4	3/4	1/2	1/4
Network and buffer status	E	B	G	B	G	B	G	B
DASH algorithm with F_S	-	SD	SU	SD	-	-	-	-
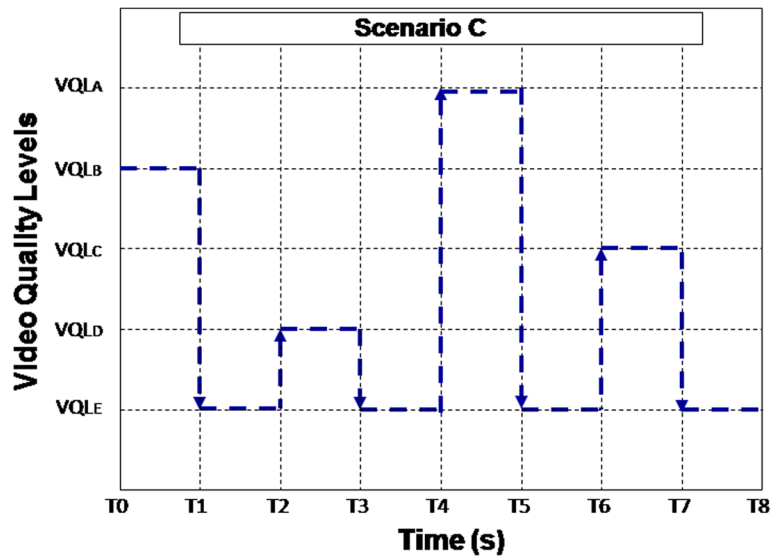
Figure 5 Scenario C. An example of VQL switching events between five quality levels.

$$\overline{\mathrm{SDF}}_k^{(D)} = \mathrm{MOS}_k^{\mathrm{mean}} - \mathrm{MOS}_k \qquad (5)$$

Our purpose here is to pick the optimal values for $w_i^{(T)}$ and C, such that the predicted $\overline{\mathrm{SDF}}$ value for the k-th scenario in (4) is as close to the desired $\overline{\mathrm{SDF}}_k^{(D)}$ value in (5) as possible. A convenient way to resolve this optimization problem is to transform (4) into logarithmic domain (for the case $n = 1$) and solve for a linear regression problem. Specifically, for the k-th scenario, taking the logarithm at both sides of (4), we have

$$\ln\left(\overline{\mathrm{SDF}}_k\right) = \ln(C) + \sum_{i=1}^{m} \frac{N_i^{(k)}}{T} w_i^{(T)} \qquad (6)$$

Pooling this for all K scenarios, we desire to have

$$\mathbf{Q}\mathbf{w}^{(T)} = \mathbf{b}^{(T)} \qquad (7)$$

where

$$\mathbf{Q} = \begin{bmatrix} 1 & q_{1,1} & \cdots & q_{1,m} \\ 1 & q_{2,1} & \cdots & q_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & q_{K,1} & \cdots & q_{K,m} \end{bmatrix}; \qquad q_{k,i} = \frac{N_i^{(k)}}{T} \qquad (8)$$
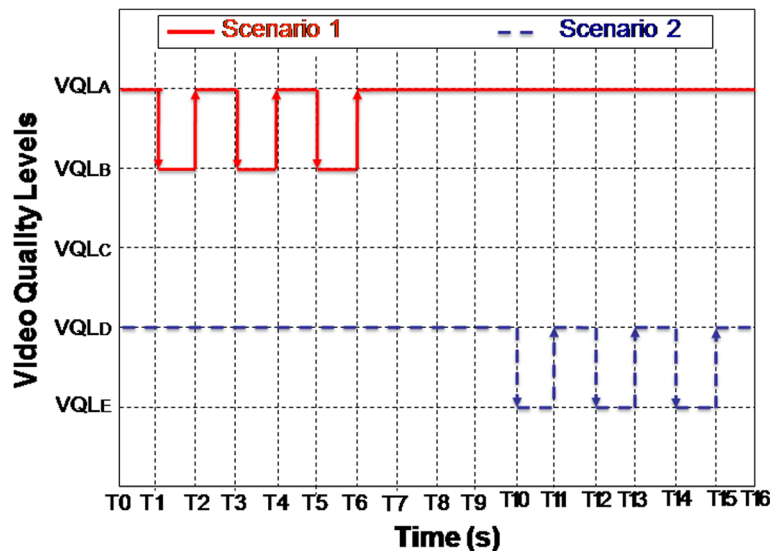

Figure 6 Scenarios D and E. Examples of VQL switching events at different temporal locations.

Table 5 List of VQLS and associated switching types

Video quality level (VQL)	Video characteristics	VQL switching event	VQL switching type
V_A	(SR1, TR1)	$V_A \longleftrightarrow V_B$	TRS (V_{AB})
V_B	(SR1, TR2)	$V_B \longleftrightarrow V_C$	TRS and SRS (V_{BC})
V_C	(SR2, TR1)	$V_B \longleftrightarrow V_D$	SRS (V_{BD})
V_D	(SR2, TR2)	$V_C \longleftrightarrow V_D$	TRS (V_{CD})
V_E	(SR3, TR2)	$V_A \longleftrightarrow V_E$	TRS and SRS (V_{AE})
V_F	(SR3, TR1)	$V_E \longleftrightarrow V_F$	TRS (V_{AE})

$$\mathbf{w}^{(T)} = \begin{bmatrix} \ln(C) \\ w_1^{(T)} \\ w_2^{(T)} \\ \vdots \\ w_m^{(T)} \end{bmatrix}; \qquad \mathbf{b}^{(T)} = \begin{bmatrix} \ln\left(\overline{\mathrm{SDF}}_1^{(D)}\right) \\ \ln\left(\overline{\mathrm{SDF}}_2^{(D)}\right) \\ \vdots \\ \ln\left(\overline{\mathrm{SDF}}_K^{(D)}\right) \end{bmatrix} \quad (9)$$

All unknowns are contained in vector $\mathbf{w}^{(T)}$, which can be obtained using a least square method, specifically, a pseudo-inverse given by

$$\mathbf{w}^{(T)} = \left(\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{b}^{(T)} \quad (10)$$

Thus, the values of the constant C and all $w_i^{(T)}$'s are obtained.

In the second phase, the $w_j^{(S)}$ parameters are estimated assuming that the values of C and $w_i^{(T)}$'s are given (from phase 1). Specifically, a series of $K = 12$ scenarios are tested where each scenario contains three temporal segments ($n = 3$). The lengths of the video used in phase 2 were 3 min. In each test scenario, a different set of VQL switching events with different switching types were used. Similar to the case in phase 1, in the k-th scenario, the net impact of VQL switching events on the overall user QoE or the desired S factor (denoted by $\overline{\mathrm{SDF}}_k^{(D)}$) would be the difference between the mean of the MOS values of all individual VQLs in all the temporal segments (denoted by $\mathrm{MOS}_k^{\mathrm{mean}}$) and the MOS value given to the whole video (denoted by MOS_k), such that

$$\overline{\mathrm{SDF}}_k^{(D)} = \mathrm{MOS}_k^{\mathrm{mean}} - \mathrm{MOS}_k \quad (11)$$

The goal here is to find the optimal values for $w_j^{(S)}$ for the given $w_i^{(T)}$ and C, so that the predicted $\overline{\mathrm{SDF}}$ value

for the k-th scenario in (4) is as close to the desired value $\overline{\mathrm{SDF}}_k^{(D)}$ as possible. For the k-th scenario, taking the logarithm at both sides of (4), we obtain

$$\ln\left(\overline{\mathrm{SDF}}_k\right) = \ln(C) + \sum_{j=1}^{n} w_j^{(S)}\left(\sum_{i=1}^{m} \frac{w_i^{(T)}N_{ij}^{(k)}}{T}\right) \quad (12)$$

Pooling this for all K scenarios, we desire to have

$$\mathbf{P}\mathbf{w}^{(S)} = \mathbf{b}^{(S)} \quad (13)$$

where

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & \cdots & \cdots & p_{1,n} \\ p_{2,1} & \ddots & \cdot & p_{2,n} \\ \vdots & \cdot & \ddots & \vdots \\ p_{K,1} & \cdots & \cdots & p_{K,n} \end{bmatrix}; \quad p_{k,j} = \sum_{i=1}^{m} \frac{w_i^{(T)}N_{ij}^{(k)}}{T} \quad (14)$$

$$\mathbf{w}^{(S)} = \begin{bmatrix} w_1^{(S)} \\ w_2^{(S)} \\ \vdots \\ w_n^{(S)} \end{bmatrix}; \qquad \mathbf{b}^{(S)} = \begin{bmatrix} \ln\left(\overline{\mathrm{SDF}}_1^{(D)}/C\right) \\ \ln\left(\overline{\mathrm{SDF}}_2^{(D)}/C\right) \\ \vdots \\ \ln\left(\overline{\mathrm{SDF}}_K^{(D)}/C\right) \end{bmatrix} \quad (15)$$

All unknowns are contained in vector $\mathbf{w}^{(S)}$, which can be obtained by a pseudo-inverse

$$\mathbf{w}^{(S)} = \left(\mathbf{P}^{\mathrm{T}}\mathbf{P}\right)^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{b}^{(S)} \quad (16)$$

With all the parameters $w_i^{(T)}$'s, $w_j^{(S)}$'s, and C determined, we can now use Equations 3 and 4 to compute the SDF factors as well as the mapped $\overline{\mathrm{SDF}}$ values for the given test scenarios, and $\overline{\mathrm{SDF}}$ can be subsequently employed to predict the drop of MOS value caused purely by VQL switching events.
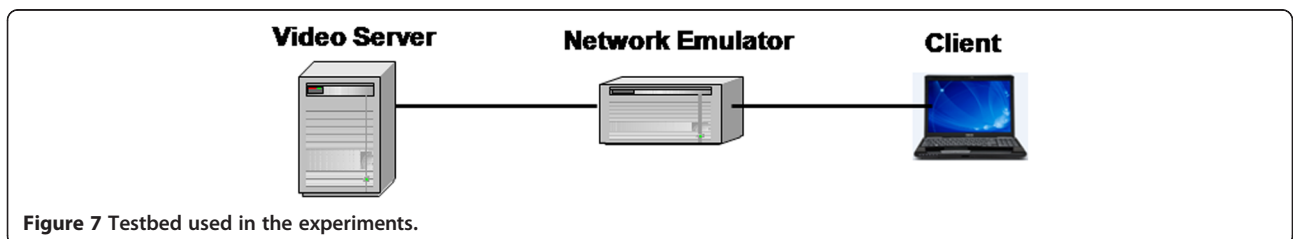

Figure 7 Testbed used in the experiments.

Table 6 Characteristics of videos used as test material

Video quality level (VQL)	Temporal resolution (fps)	Spatial resolution (width × height)
V_A	25	854 × 480
V_B	20	854 × 480
V_C	25	640 × 360
V_D	20	640 × 360
V_E	20	320 × 180
V_F	25	320 × 180

It is worth noting that commercial applications of video streaming services can offer a high number of different spatial and temporal resolutions. In order for the SDF parameter to be useful for real applications, SDF needs to be agnostic to the different video resolutions and consequently works with different switching types.

Based on the $w_i^{(T)}$ and $w_j^{(S)}$ parameters obtained in previous computation, we propose a model to generalize the results to cover a broader range of switching events.

In particular, we define a spatial resolution change parameter

$$R = \sqrt{\frac{\max(W_c, W_n)\ \max(H_c, H_n)}{\min(W_c, W_n)\ \min(H_c, H_n)}} \qquad (17)$$

where (W_c, H_c) and (W_n, H_n) represent the widths and heights of the video before and after switching, respectively. Considering the results obtained by (10) and analyzing different mathematical models, such as exponential and polynomial functions, we find that $w_i^{(T)}$ can be modeled empirically by

$$w_i^{(T)} = \alpha + \beta \log_2(1 + (R - 1)/\eta) \qquad (18)$$

where $\alpha = 2.69$, $\beta = 8.73$, and $\eta = 0.33$ when $R \leq 1.33$, and $\alpha = 11.44$, $\beta = 1.89$ and $\eta = 1.34$ when $R > 1.33$. In a similar way, we find the values of $w_j^{(S)}$ can be well fitted by considering the results obtained by (16):

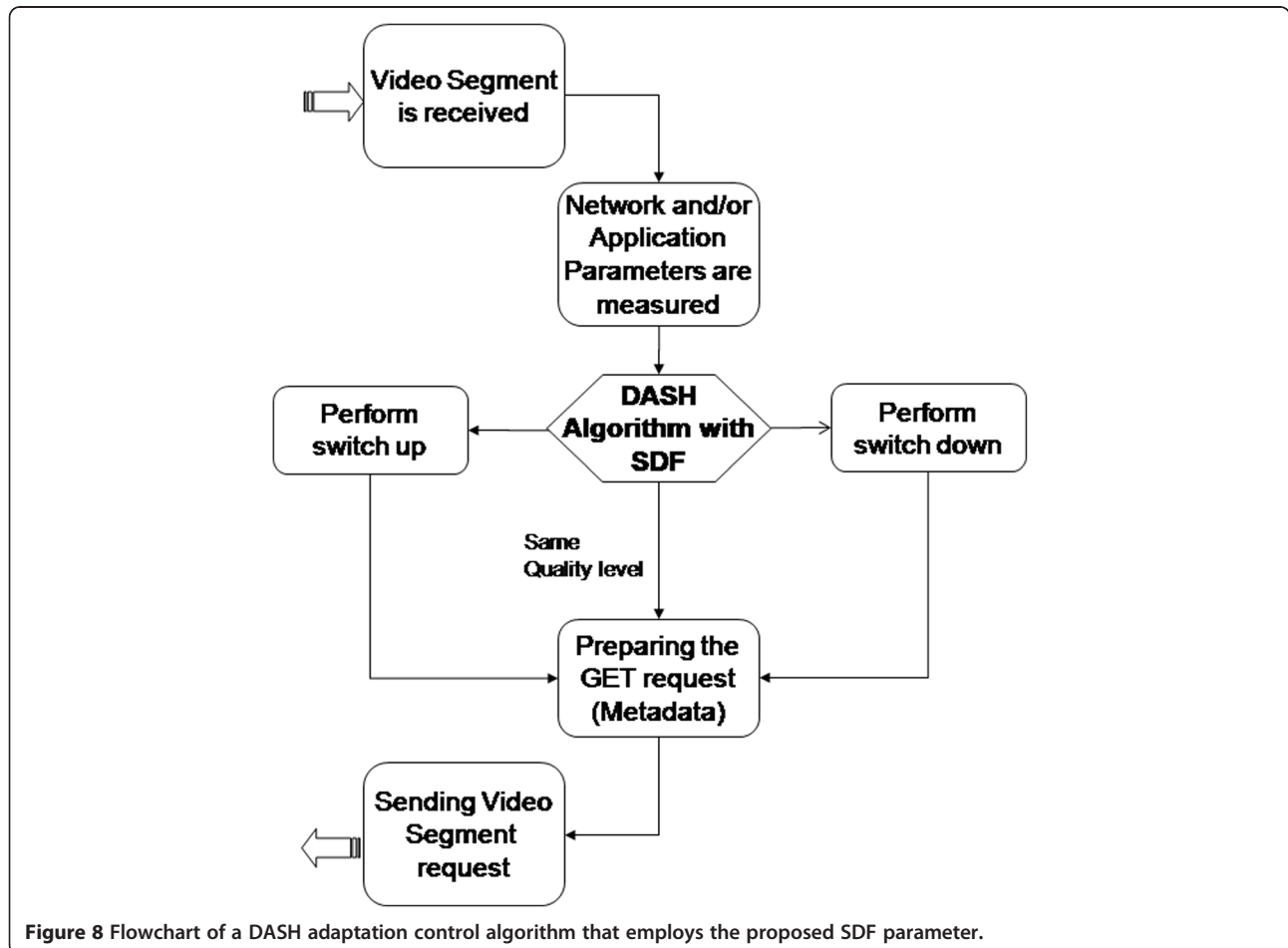$$w_j^{(S)} = \kappa + \lambda \log_2(1 + (n_c - 1)/(n - 1)) \qquad (19)$$

Figure 8 Flowchart of a DASH adaptation control algorithm that employs the proposed SDF parameter.
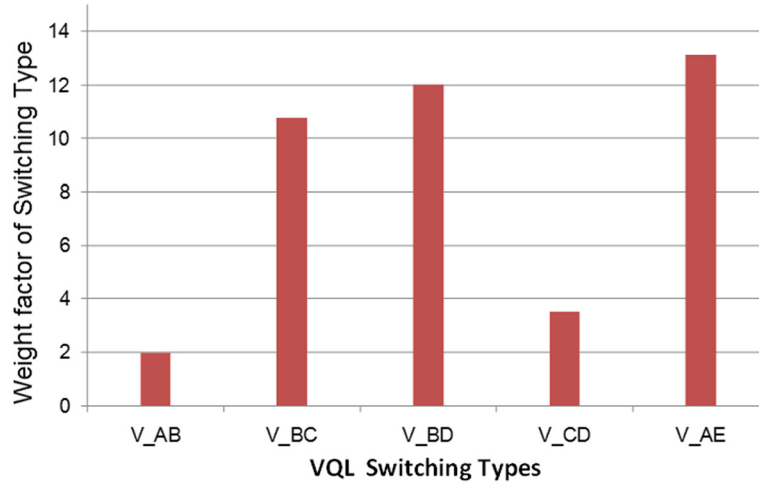
Figure 9 Weighting factor $w_i^{(T)}$ of different VQL switching types.

where $\kappa = 1.42$, $\lambda = -0.38$, n is the total number of temporal segments considered in the video, and n_c is the current temporal segment in which $w_j^{(S)}$ is calculated.

Finally, the MOS value that characterizes the user QoE can be predicted by incorporating $\overline{\text{SDF}}$ into previous QoE models that estimate MOS without taking into account quality degradations due to VQL switching. For example, the video streaming quality metric (VsQM) proposed in [23] provides a model to predict MOS and is specifically useful when pauses exist during video replay. Combining VsQM and SDF, we obtain a model that predicts the overall MOS value by

$$\text{MOS}^{(P)} = \overline{Vs\bar{Q}M} - \overline{\text{SDF}} \qquad (20)$$

where $Vs\bar{Q}M$ and $\overline{\text{SDF}}$ are the VsQM and SDF factors after mapped to the scale that can be directly used to predict MOS in a 5-point scale. This predicted MOS value, denoted by $\text{MOS}^{(P)}$ can then be employed by DASH algorithms for adaptive video streaming.

5 Implementation and testing
5.1 Testing environment and implementation
The testbed used in our experiment is shown in Figure 7, which is isolated with no other processes running in the same computers. A network emulator is implemented based on the open-source tool NETem, which controls the available bandwidth between the client and the video server. The video server is installed with Linux
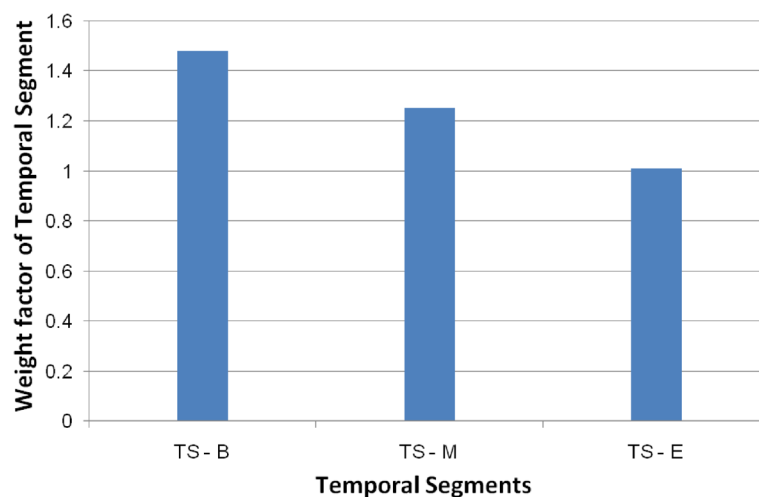

Figure 10 Weighting factor $w_j^{(S)}$ of three segments. TS-B, TS-M, and TS-E are for the temporal segments at the beginning, middle, and end of the video, respectively.
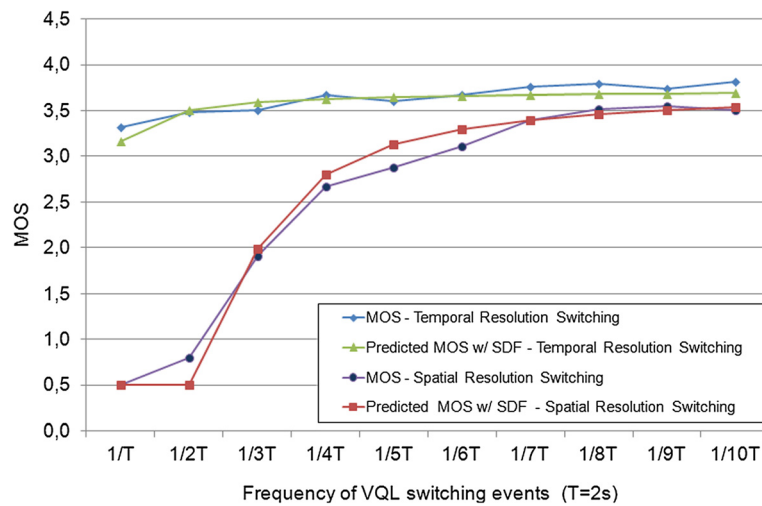
Figure 11 Subjective and objectively predicted MOS for different frequencies of VQL switching events.

and Apache web server version 2.2.21. In addition, a video player is developed using an Open Source Media Framework (OSMF). The initial buffering level requirement is set to 6 s.

Using the information of the metadata MPD, the application is able to know the spatial resolution of the video sequences. This information is obtained from the MPD xml code, specifically from the data contained in the element named 'Representation' and its attributes 'width' and 'height.' Therefore, the ratio of spatial resolutions between the current and next video segment can be calculated using the width or height values.

In the first and second phases of this work, all test videos were 1 or 3 min in length. In the validation phase, videos were 9 and 21 min in length. These videos were compressed using an H.264/AVC video encoder with

different encoding characteristics to obtain six VQLs, as presented in Table 6. The videos are divided into 2-s pieces and are stored in the video server with appropriate identifications. The client sends an HTTP request that contains the URL of a specific video identification, which has been determined by a DASH algorithm running at the client side.

Using our testbed, drastic changes in available bandwidth were emulated. Thus, several test scenarios were created, in which different numbers of VQL switching events and different switching types were inserted. In addition, the temporal locations of VQL switching events vary between different test scenarios.

A DASH control algorithm is implemented based on OSMF in which the SDF parameter was included. The flowchart is presented in Figure 8. In order to assess the
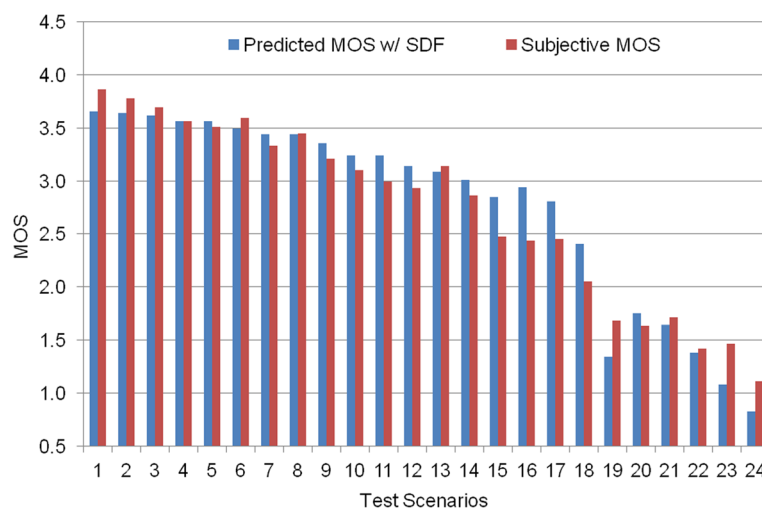

Figure 12 Subjective and objectively predicted MOS for different VQL switching types.
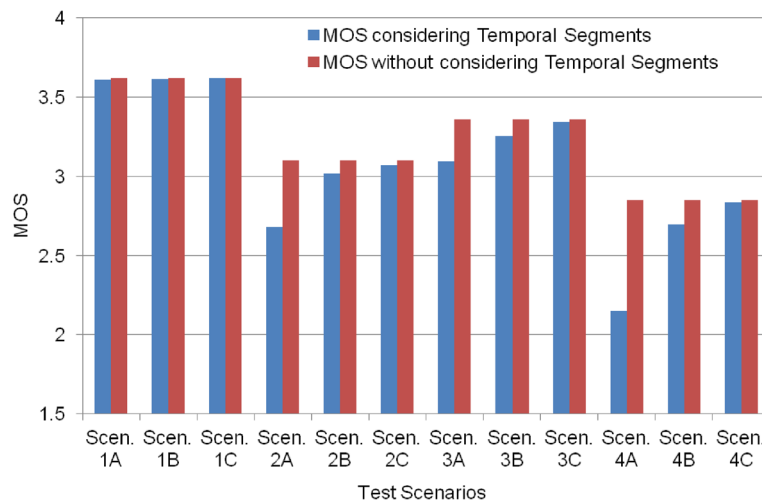
Figure 13 Predicted MOS for different temporal segments. A, B, and C are for the initial, intermediate, and final temporal segments, respectively.

impact of SDF in a DASH control algorithm, the same test scenarios were evaluated using the same DASH algorithm but without using SDF, and the two test cases are compared, as described later.

5.2 Test results

A total of 78 subjects participated in the subjective test, including 44 females and 34 males, aged between 18 and 49 years. None of them presented any sight problems or experience in the quality assessment task. A 21.5-in. LCD monitor was employed with the following characteristics: $1,920 \times 1,080$ pixel resolution, widescreen ratio of 16:9 and brightness of 250 cd/m^2. The test environment had no reflecting ceiling walls or floors or either any disturbing objects. The tests were conducted in

14 weeks, and during this period, the same test room was kept constant. All tests were performed individually and a time limit was not enforced. An instruction session was performed before the tests, in which the assessors were shown sample videos and the experiment process was explained. In the tests, an observation distance of 50 to 60 cm was considered, and assessors used the scale presented in Table 1. Each video received at least 15 scores by the assessors and the scores are averaged to calculate the MOS value. With the test results, a statistical analysis was performed and no observer was identified as an outlier.

Figure 9 presents the results of the $w_i^{(T)}$ values computed using (10), and Figure 10 shows $w_j^{(S)}$ values obtained by (16), respectively.
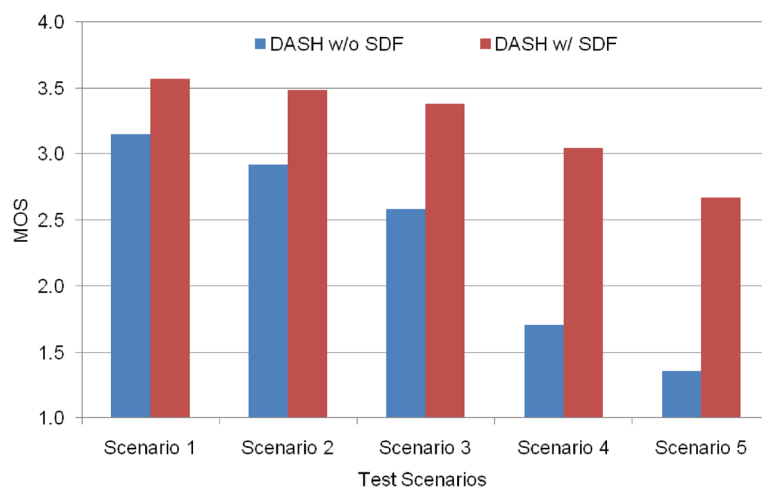

Figure 14 Performance comparison based on MOS values between DASH algorithms with and without considering SDF.

Table 7 Description of the switching events considering their types and temporal distributions used test scenarios

Scenario	Temporal segment	Switching type					
		V_AB	V_BC	V_BD	V_CD	V_AE	V_EF
1	TS-B	5					
	TS-M		3				
	TS-E						
2	TS-B						
	TS-M			5			1
	TS-E						
3	TS-B	1					
	TS-M		3	2			
	TS-E				4		
4	TS-B			3			
	TS-M		4				
	TS-E				2		
5	TS-B			3			
	TS-M					2	1
	TS-E		3				

First temporal segment (TS-B), second temporal segment (TS-M), and third temporal segment (TS-E).

Figure 11 extends Figure 2 by showing how the user's QoE decreases when the frequency of VQL switching events is increased. Results of subjective MOS and the predicted $MOS^{(P)}$ by (20) are presented. The Pearson correlation coefficients for the cases of temporal and spatial resolution are 0.92 and 0.98, respectively.

Figure 12 shows both the subjective MOS and the predicted $MOS^{(P)}$ by (20) for the 24 scenarios considered in the first phase. The Pearson correlation coefficient between subjective and objective MOS values is 0.96.

In order to demonstrate the impact of temporal location, Figure 13 shows how the same impairments located at different time instants degrade the user's QoE. Four scenarios are presented, each with three variations, named A, B, and C, representing the initial, intermediate, and final temporal segments, respectively. Thus, scenarios A's have VQL switching events only in the initial temporal segment, and the same rule for scenarios B and C.

From Figure 13, it can be observed that VQL switching events in the first temporal segment have the highest negative effect on the user QoE, and depending on the test scenario, the QoE can be drastically decreased.

5.3 Applications to DASH algorithms

Five scenarios were used to test a DASH algorithm with and without employing the SDF parameter. In the case that the SDF parameter is adopted, a threshold of 0.6 on the $\overline{SDF}$ value is used. Figure 14 shows the subjective evaluation results. Depending on the test scenario, the difference between using and not using the SDF parameter could vary dramatically. In order to clarify the implementation of test scenarios, Table 7 presents the number and type of switching events that happened in each temporal segment during a video sequence, considering that the SDF parameter was not used in the DASH algorithm. For instance, scenario 5 had the largest quality changes between VQLs, while scenario 1 was the less affected.

In order to validate the generalized SDF parameter in (18) and (19), additional subjective tests were conducted for video lengths of 9 and 21 min. Four versions of the same video were used, all of them with the same
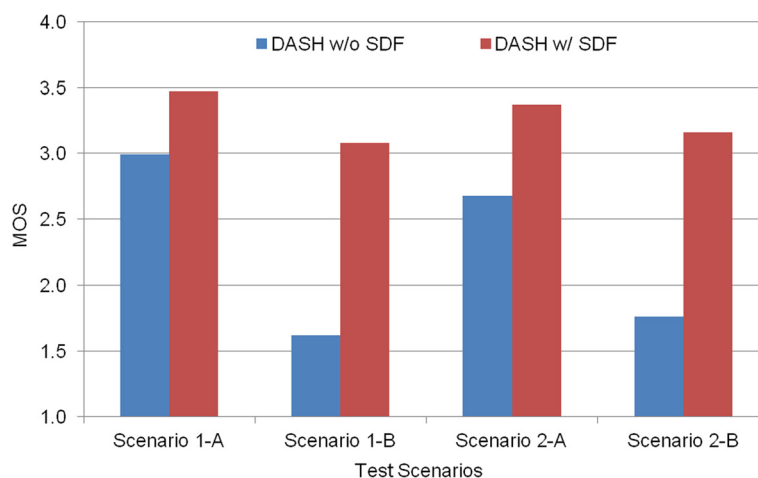

Figure 15 Performance comparison based on MOS value for video lengths of 9 and 21 min.

temporal resolution of 25 fps but with different resolutions of $1{,}136 \times 640$, 960×540, 480×234, and 320×200, respectively. The video sequences used in the experimental tests were built using the same methodology presented in Table 7. Figure 15 shows the results obtained, where scenarios 1-A, 1-B, 2-A, and 2-B represents 9-min video with moderate bandwidth change, 9-min video with frequent bandwidth change, 21-min video with moderate bandwidth change, and 21-min video with frequent bandwidth change, respectively. In the case that the SDF parameter is adopted, a threshold of 0.6 on the $\overline{SDF}$ value is used. These results are similar to those presented in Figure 14.

From Figures 14 and 15, it can be observed that the DASH algorithm that considers SDF substantially improves the user's QoE, especially in the scenarios where the bandwidth varies frequently. Furthermore, the results in Figure 15 demonstrate the generalization ability of the proposed method to the case of long video lengths.

6 Conclusions

Existing DASH solutions do not take into account the impact of VQL switching on the users' QoE. In this study, we make one of the first attempts to address this problem through subjective testing, objective modeling, as well as computer and network configurations to create different scenarios that involved DASH algorithms for adaptive streaming. The major contributions of our work are summarized as follows: First, we find that frequent VQL switching has strong impact on the users' QoE for its disturbance to users' attention to the video content. Second, we find that switchings in spatial and temporal resolutions have significantly different impacts on the QoE. Third, three features in VQL switching, i.e., switching frequency, switching type, and switching temporal location, are identified as the key factors in characterizing the impact of VQL switching on the users' QoE. Fourth, a SDF model is developed to account for the changes caused by VQL switching on the users' QoE. Fifth, a series of subjective experiments are conducted to calibrate the parameters in the SDF model as well as to test the quality prediction performance of objective models on subjective MOS. Sixth, the SDF model is embedded into DASH algorithms and compared with the same algorithms without considering the SDF factor. Validations by subjective test show that the MOSs given by human observers are significantly improved by incorporating SDF in DASH.

Competing interests
The authors declare that they have no competing interests.

Authors' information
DZR received his B.S. degree in Electronic Engineering from the Pontifical Catholic University of Peru and his M.S. degree (2009) and PhD in Electronic Engineering (2013) from the Escola Politécnica of the University of São Paulo (EPUSP). He studied Electronic Systems at USP, with solid knowledge in

Telecommunication Systems and Computer Science based on 13 years of professional experience in important companies. His current interest includes QoS and QoE in multimedia services, digital TV, and architect solutions in Telecommunication Systems. He is currently a professor at the Computer Science Department at Federal University of Lavras (UFLA), Minas Gerais, Brazil.
ZW received his Ph.D. degree from the University of Texas at Austin (2001). He is currently an associate professor at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image/video processing, coding and quality assessment, multimedia communications, computational vision, and biomedical signal processing. He has more than 100 publications in these fields with more than 16,000 citations (Google Scholar). He was a recipient of the 2009 IEEE Signal Processing Society Best Paper Award, 2009 Ontario Early Researcher Award, and ICIP 2008 Best Student Paper Award as a senior author. He is a member of the IEEE Multimedia Signal Processing Technical Committee (MMSP-TC) and has been served now and in the past as an associate editor of IEEE Transactions on Image Processing, IEEE Signal Processing Letters, and Pattern Recognition.
RLR received her B.S. degree in Computer Science from UNIFEI, Brazil and her M.S. degree from the University of São Paulo - USP (2009). She is a Ph.D. student at Escola Politécnica of the University of Sao Paulo (EPUSP). Her current research interest includes computer networks, quality of experience of multimedia service, social networks, and recommendation systems.
GB was granted her Ph.D. in Electronic Engineering (1986) by the Escola Politécnica of the University of São Paulo (EPUSP). Her current research interests include computer networks and digital television focusing on the aspects of distributed systems, distributed middleware, QoS mechanisms, collaborative virtual environment, middleware for digital TV, interactive digital TV, videoconferencing, modeling, and performance analysis of networks, and applications in distance education.

Author details
[1]Department of Computation Science, University of Lavras, Câmpus Universitário, Caixa Postal 3037, CEP 37200-000, Lavras, Minas Gerais, Brazil. [2]Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada. [3]Department of Computer Engineering at the School of Engineering, University of São Paulo, Avenue Prof. Luciano Gualberto, Travessa 3, no. 380, CEP 05508-010 São Paulo, Brazil.

References
1. T Stockhammer, Dynamic adaptive streaming over HTTP - standards and design principles, in *Proc. ACM Conf. on Multimedia Systems (MM'11)* (San Jose, 2011), pp. 133–144
2. H-J Park, D-H Har, Subjective image quality assessment based on objective image quality measurement factors. IEEE Trans. Consumer Electron. **57**(3), 1176–1184 (2011)
3. Q Xu, Q Huang, Y Yao, Online crowdsourcing subjective image quality assessment, in *Proc. of 20th ACM International Conference on Multimedia (MM'12)* (Nara, 2012), pp. 359–368
4. F Ribeiro, D Florencio, V Nascimento, Crowdsourcing subjective image quality evaluation, in *Proc. of 18th IEEE International Conference on Image Processing (ICIP)* (Brussels, 2011), pp. 3097–3100
5. ISO, *ISO/IEC IS 23009-1, Information Technology – Dynamic Adaptive Streaming over HTTP (DASH)* ISO (Geneva, 2012)
6. V Adzic, H Kalva, B Furht, Optimizing video encoding for adaptive streaming over HTTP. IEEE Trans. on Consumer Electron. **58**(2), 397–403 (2012)
7. S Akhshabi, A Begen, C Dovrolis, An experimental evaluation of rate-adaptation algorithms in adaptive streaming over HTTP, in *Proc. ACM Conf. on Multimedia Systems* (San Jose, 2011), p. 157

8. W Hapsari, A Umesh, M Iwamura, M Tomala, B Gyula, B Sebire, Minimization of drive tests solution in 3GPP. IEEE Commun. Mag. **50**(6), 28–36 (2012)
9. Y-M Hsiao, C-H Chen, J-F Lee, Designing and implementing a scalable video-streaming system using an adaptive control scheme. IEEE Trans. on Consumer Electron. **58**(4), 1314–1322 (2012)
10. T Lohmar, T Einarsson, P Frojdh, F Gabin, M Kampmann, Dynamic adaptive HTTP streaming of live content, in *Proc. IEEE World of Wireless, Mobile and Multimedia Networks (WoWMoM)* (Lucca, 2011), pp. 1–8
11. C Liu, I Bouazizi, M Gabbouj, Rate adaptation for adaptive HTTP streaming, in *Proc. ACM Conf. on Multimedia Systems* (San Jose, 2011), pp. 169–174
12. R Mok, E Chan, R Chang, Measuring the quality of experience of HTTP video streaming, in *Proc. of IFIP/IEEE International Symposium on Integrated Network Management (IM)* (Dublin, 2011), pp. 485–492
13. W Pu, Z Zou, C Ch, Dynamic adaptive streaming over HTTP from multiple content distribution servers, in *Proc. of IEEE Global Telecom. Conference* (Houston, 2011), pp. 1–5
14. LD Cicco, S Mascolo, V Palmisano, Feedback control for adaptive live video streaming, in *Proc. of ACM Conf. on Multimedia Systems* (San Jose, 2011), pp. 145–156
15. S Gouache, G Bichot, A Bsila, C Howson, Distributed & adaptive HTTP streaming, in *Proc. IEEE International Conference on Multimedia and Expo (ICME)* (Barcelona, 2011), pp. 1–6
16. T Porter, XH Peng, An objective approach to measuring video playback quality in loss networks using TCP. IEEE Commun. Lett. **15**(1), (2011)
17. K Evensen, D Kaspar, C Griwodz, P Halvorsen, A Hansen, P Engelstad, Improving the performance of quality-adaptive video streaming over multiple heterogeneous access networks, in *Proc. of ACM Conf. on MM. Sys* (San Jose, 2011), pp. 57–68
18. R Huysegems, B De-Vleeschauwer, K De-Schepper, C Hawinkel, T Wu, K Laevens, W Van-Leekwijck, Session reconstruction for HTTP adaptive streaming: laying the foundation for network-based QoE monitoring, in *Proc. IEEE 20th International Workshop on Quality of Service (IWQoS)* (Coimbra, 2012), pp. 1–9
19. N Cranley, P Perry, L Murphy, User perception of adapting video quality. Int. Journal of Human-Computer Studies **64**(8), 637–647 (2006)
20. N Feamster, D Bansal, H Balakrishnan, On the interactions between layered quality adaptation and congestion control for streaming video, in *Proc. 11th International Packet Video Workshop* (Kyongju, 2001)
21. J Kucerova, J Polec, D Tarcsiova, Video quality assessment using visual attention approach for sign language. World Acad. Sci. Eng. Technol. **65**, 194–199 (2012)
22. R Laghari, K Crespi, N Molina, B Palau, QoE aware service delivery in distributed environment, in *IEEE Workshops of International Conference on Advanced Information Networking and Applications* (Biopolis, 2011), pp. 837–842
23. D Rodriguez, J Abrahão, D Begazo, R Lopes, G Bressan, Quality metric to assess video streaming service over TCP considering temporal location of pauses. IEEE Trans. on Consumer Electron. **58**(3), 985–992 (2012)
24. International Telecommunication Union, *ITU-R BT.500-11: Methodology for the Subjective Assessment of the Quality of Television Pictures* (Geneva, 2002)
25. International Telecommunication Union, *ITU-T P.910: Subjective Video Quality Assessment Methods for Multimedia Applications* (Tech. Rec, Geneva, 2008)
26. T Hosfeld, S Biedermann, R Shatz, A Platzer, The memory effect and its implications on Web QoE modeling, in *Proc. of 23rd International Teletraffic Congress (ITC)* (San Francisco, 2011), pp. 103–110
27. D Rodriguez, R Lopes, E Costa, J Abrahão, G Bressan, Video quality assessment in video streaming services considering user preference for video content. IEEE Trans. on Consumer Electron. **60**(3), 436–444 (2014)
28. S Aroussi, T Bouabana-Tebibel, A Mellouk, Empirical QoE/QoS correlation model based on multiple parameters for VoD flows, in *Proc. of Global Communications Conference (GLOBECOM)* (Anaheim, 2012), pp. 1963–1968