Inferring the QoE of HTTP Video Streaming from User-Viewing Activities

Ricky K. P. Mok, Edmond W. W. Chan, Xiapu Luo, and Rocky K. C. Chang
Department of Computing
The Hong Kong Polytechnic University
Hunghom, Kowloon, Hong Kong
{cskpmok|cswwchan|csxluo|csrchang}@comp.polyu.edu.hk

ABSTRACT

HTTP video streaming, employed by most of the video-sharing websites, allows users to control the video playback using, for example, pausing and switching the bit rate. These user-viewing activities can be used to mitigate the temporal structure impairments of the video quality. On the other hand, other activities, such as mouse movement, do not help reduce the impairment level. In this paper, we have performed subjective experiments to analyze user-viewing activities and correlate them with network path performance and user quality of experience. The results show that network measurement alone may miss important information about user dissatisfaction with the video quality. Moreover, video impairments can trigger user-viewing activities, notably pausing and reducing the screen size. By including the pause events into the prediction model, we can increase its explanatory power.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations – *Network monitoring*; H.1.2 [**Models and Principles**]: User/Machine Systems – *Human factors*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Evaluation/methodology*

General Terms

Experimentation, Measurement, Human Factors

Keywords

QoE, HTTP Video Streaming, User-viewing activities

1. INTRODUCTION

The Quality of Experience (QoE) is a multi-dimensional construct, consisting of subjective and objective parameters [19, 23]. For video streaming applications, objective parameters are often the measure of traditional QoSes (Quality of Services), such as Network QoS and Application QoS [11]. Network QoS is a set of network path metrics, including round-trip time (RTT), jitter, and packet loss rate. Application QoS comprises application parameters of the video playback, such as video bit rate, frame rate, and video resolution. Previous works have studied how these QoSes correlate with the QoE [15, 21].

Subjective parameters, such as user expectation and satisfaction, are difficult to evaluate, because of their subjective nature. Although the MOS (Mean Opinion Score) [12, 13] obtained from subjective assessments can provide an overall measurement of the parameters, the assessments are often costly, time-consuming, and not scalable. Moreover, users usually lack the incentive to report accurate scores. Therefore, in this paper we resort to indirect mechanisms for the MOS measurement.

In UDP-based video streaming, poor network conditions may result in frame blocking or even frame freezing, and users do not have control knobs to mitigate the impact. In contrast, HTTP video streaming ensures a reliable delivery of the video stream through TCP and enables playing incompletely downloaded video using the progressive download technology [3]. Although a low TCP goodput could still destruct the temporal structure of the video playback [17], users can mitigate the impact, for example, by pausing the video to buffer more video data. We refer these user actions to as *user-viewing activities*.

In this paper, we propose employing the user-viewing activities to estimate the subjective parameters in the QoE of HTTP video streaming. In particular, we evaluate based on subjective measurement experiments whether user-viewing activities are induced by temporal structure impairments or they are just random actions. Our main findings include:

1. Network quality measurement alone may miss some important information about user dissatisfaction about the video quality.
2. Video impairments could induce user-viewing activities, notably pausing the video and reducing the viewing area.
3. In addition to the application performance metrics proposed in [17], we also incorporate the pause and screen size switching events into the MOS prediction. By using logistic regression analysis, we obtain a better model fit, and the explanatory power increases from 0.24 to 0.32 in terms of the Nagelkerke R^2.

The remaining of this paper is organized as follows. Section 2 describes the user-viewing activities and overall methodology used in this study. Section 3 details the experiment

W-MUST'11, August 19, 2011, Toronto, Ontario, Canada.

setup, whereas section 4 reports the experiment results. After highlighting some related works in section 5, we conclude the paper in section 6.

2. USER-VIEWING ACTIVITIES

User-viewing activities refer to the activities that a user interacts with a player interface. We haved conducted a survey on how users behave when the playback is smooth or jerky. In the survey, we listed all the user-viewing activies in Table 1 except the two mouse movement items. The subjects were asked to rate their level of agreement on whether they will perform the listed activities under a given scenario (i.e., the playback is smooth/jerky). The level of agreement is measured with a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Two sets of ratings are obtained for each user-viewing activity. We compare the two scenarios by subtracting the mean rating of each activity as shown in Eqn. (1). The mean rating difference of activity j, denoted by $\Delta\bar{r}^j$, is given by

$$\Delta\bar{r}^j = \bar{r}^j_{smooth} - \bar{r}^j_{jerky}, \tag{1}$$

where $\bar{r}^j_{smooth}$ and $\bar{r}^j_{jerky}$ are the mean rating of the user-viewing activity j given smooth and jerky playback scenarios, respectively. Using paired samples t-test, we can obtain the level of significance, p, for each activity.

The third column in Table 1 shows the mean rating difference of each user-viewing activity from a survey of 19 people. The activities with positive mean rating difference, $\Delta\bar{r}^j > 0$, means that users prefer those activities when the playback is smooth. Otherwise, users favor the activity in jerky playback scenario. The results show that users choose pausing, switching to a lower picture quality, and watching with normal screen size under jerky playback scenario. In contrast, for smooth playback scenario, only switching up the quality and enlarging the screen size are significant. Users show no significant preference for other activities, such as resuming and time shifting.

On the other hand, some user-viewing activities can help mitigate the temporal structure impairments [17]. Pausing that can increase the time for buffering video data is an example of positive technical impact (i.e., Tech. is +). Other examples include refreshing the page which allows the player to choose another video server from a content delivery network and switching to a lower video quality by reducing the video data size. In contrast, resuming and forward time shifting have negative impact. Resuming the playback consumes the buffered video, and forward time shifting gives up the buffered video and sends a new `HTTP GET` request. The effects of impact is shown in the second column of Table 1 represented by +, ◦, and −.

Table 1 also lists the possible user-viewing activities and their effects of impact, explicit, and implicit meaning, particularly for HTTP video streaming.

2.1 The overall methodology

In this section, we describe the methodology for evaluating whether user-viewing activities are induced by temporal structure impairments or just user's random actions. Our research hypothesis is given in Hypothesis 1. To testify against the null hypothesis which claims that the activities are random, we analyze the activities recorded before and after the impairment events.

HYPOTHESIS 1. *The user-viewing activities are more likely to be triggered after the presence of temporal structure impairments.*

We assume that an impairment event only affects the user-viewing activities nearby. Therefore, we inspect the user-viewing activities within a range around each impairment event. Figure 1 shows an example of a video playback time line. The video starts playing at t_0 and ends at t_{end}. Three impairment events occur at times t_{i-1}, t_i, and t_{i+1}. Two user-viewing activities are recorded at times t_a and t_{a+1}. A time period δ, computed by Eqn. (2), is half of the time between the current impairment event and the nearest impairment events or the start or the end of the video playback. An upper bound for δ is arbitrarily set to 5 seconds to prevent the inclusion of irrelevant activities far away from the impairment event. Two other time periods d_{ai} and $d_{(a+1)i}$ are the time displacements from the activities at t_a and t_{a+1} to t_i, respectively.

$$\delta = \frac{\min(t_{i+1} - t_i, t_i - t_{i-1}, t_i - t_0, t_{end} - t_i, 10)}{2}. \tag{2}$$

Since random activities occur independent of the impairment events, they could occur before or after an impairment with equal probability. For the activities involving mouse clicks, the average time displacement to the impairment events is zero (i.e., $\overline{D_i} = 0$), where D_i is the summation of time displacement within the range between $t_i - \delta$ and $t_{i+1} + \delta$ and n_i is the number of user-viewing activities within the range as shown in (3).

$$D_i = \sum_{j=0}^{n_i} d_{(a+j)i}. \tag{3}$$

On the other hand, the mouse movement at time t is quantified by the speed of the cursor movement, v_t, as shown in (4), where (x_t, y_t) and (x_{t-1}, y_{t-1}) are the current and the pervious recorded coordinates, respectively. The speed of cursor movement is obtained by the Euclidean distance between the two coordinates over the difference in recorded timestamps, Δt. If the mouse movement is impairment-driven, the speed of the cursor movement is expected to be higher after the event.

$$v_t = \frac{\sqrt{(x_{t-1} - x_t)^2 + (y_{t-1} - y_t)^2}}{\Delta t}. \tag{4}$$

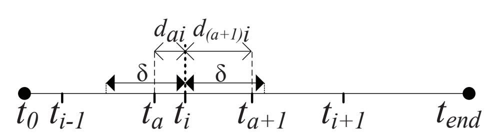

Figure 1: A time line for a video viewing session with impairment events and user-viewing activities.

3. EVALUATION

To validate our hypothesis, we have carried out experiments to record the subjects' video watching activities under various scenarios. We describe the experiment setup in this section and the results in the next.

Table 1: A list of possible user-viewing activities with the effect of technical impact, mean rating difference, explicit and implicit meaning.

Activities	Tech.	$\Delta\bar{r}^j$	Explicit Meaning	Possible Implicit Meaning
Pause	+	-2.37***	Stop playing the video playback for a short period of time.	More time is needed to buffer the video data.
Resume	−	0.26	Continue playing the paused video playback.	Reach the tolerance limit.
Refresh	+	-1.79	Reload the page and video.	The playback quality is unacceptable, and reloading may help.
Switch to a lower video quality	+	-1.79***	Watch the video with lower picture quality.	Scarify the spatial qualify for the playback smoothness.
Switch to a higher video quality	−	2.26***	Watch the video with better picture quality.	The user thinks the current speed is fast enough for watching with a better quality.
Play with full screen	◦	2.42***	Watch the video with a larger size.	The playback is enjoyable.
Return to the normal-size screen	◦	-0.74*	Watch the video with a smaller size.	The impairments are annoying.
Forward time shift	−	0.53	Watch the content after the current video position	The video content is not interesting.
Backward time shift	+	-0.58	Replay the content before the current video position.	Replaying the buffered video can result in a smoother playback.
Lost of window focus	+	-0.47	The browser window is covered or minimized.	The user may not be watching the video.
Frequent mouse movement	◦	n/a	The user moves the mouse over the screen quickly.	The impairments are annoying.
Infrequent mouse movement	◦	n/a	The user does not use the mouse.	The user is enjoying the video.

Note: +, ◦ and − represent positive, neutral, and negative effect, respectively. For $\Delta\bar{r}^j$, *p <0.05, **p <0.01, ***p <0.001.

3.1 Experiment setup

Figure 2 shows the experiment setup which is a simple video delivery system. The video server listening on multiple TCP ports responds with the same content. The Click modular router [16], placed in front of the video server, introduces delay and packet loss to the TCP flows. Table 2 lists the path metrics setting used by the router. The router emulates one of the 9 (3×3) combinations of path metrics on each TCP port. Therefore, connecting the video server via different ports results diverse network path performance. A mild level of cross-traffic with Pareto distributed inter-departure time and fixed size packets is generated by the distributed Internet traffic generator [5]. A workstation installed with a DAG card [1] captures all the traffic between the click router and the video server. The logging server is responsible for recording the information reported by the customized video player (see section 3.2) during the experiment. The traffic between the logging server and the video player, however, will not be captured and manipulated by the click router.

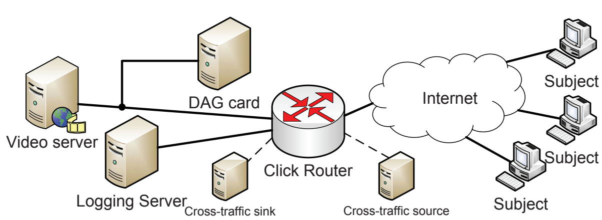

Figure 2: Experiment setup.

3.2 Client software

We have enhanced FlashTrack [17], a customized Adobe Flash video player, to record the application events. The

Table 2: Path quality settings used in the Click modular router.

Path quality metrics	Settings
Additional delay (ms)	0, 40, 80
Round-trip packet loss rate	0%, 2%, 4%

events include the current position of video playback, the buffer status, the number of bytes loaded, buffer-full events, and buffer-empty events. In addition to those events, we record user-viewing activities listed in Table 1. Besides using Flash, we have employed Javascript to capture the cursor coordinates and browser's focus. All the events are logged periodically every 0.25 seconds. The logs are then aggregated and sent to a logging server every three seconds.

The video player provides all the basic functionalities offered by commodity video sharing websites. Users can pause and resume the video playback, changing the video quality, watch in full screen mode and forward/backward shift along the buffered video. Moreover, the downloading progress, current video position, and the length of the video are visible to the subjects. When a buffer-empty event occurs, a small screen is shown over the video displaying the percentage of buffer filled until the buffer is filled up. These visual components help the subjects understand the status of the playback.

3.3 Video materials

Three videos clips are chosen for the experiments. They are labeled as *speech*, *TVshow*, and *sports*, in an ascending order of temporal complexity. Video *speech* shows a person delivering a speech with static background. Video *TVshow* is a portion of a local TV comedy show. Video *sports* is a highlight of basketball games. Table 3 shows the detailed profiles of the video clips used in this experiment. The quality of the source videos is equivalent to 720p. The source videos are then down-sampled into three lower bit rate versions with H.264 codec according to the profiles of 480p, 360p, and 240p.

Table 3: Profiles of the video clips.

Parameters	720p	480p	360p	240p
Video width (pixel)	1280	854	640	400
Video height (pixel)	720	480	360	226
Video bit rate (Mbps)	2	1	0.5	0.25
Video frame rate (fps)	29.97	29.97	29.97	29.97
Audio bit rate (kbps)	128	96	80	32

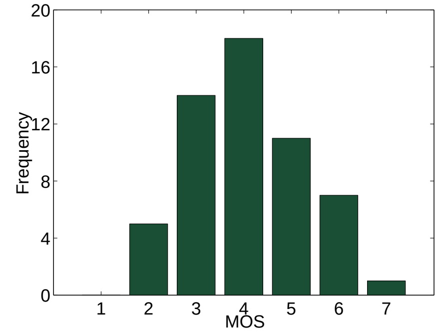

Figure 3: The overall distribution of MOS.

3.4 Subjective assessment

After filling some basic information and answering questions on video-watching habits, each subject was given a list of four videos to watch. They were first instructed to access to a dedicated video to try the platform before starting the experiment. Through this training process, the subjects became more familiar with the testing environment and understood clearly about the functionality provided by the video player. After that, the subjects could freely select the watching sequence of the remaining 3 videos. To mitigate the order effects, the display order and the choice of network path performance of the remaining three videos were randomized, and the initial quality of all the videos was 480p.

The subjects were first informed that they might experience dissimilar performance for different links. Besides, they were also reminded to behave as usual and watch the entire video clips. At the end of each video clip, the subjects could immediately rate their perceived video-watching experience. A 7-point Likert scale of MOS was adopted, from 1 (“Bad”) to 7 (“Excellent”), for obtaining a higher granularity.

4. RESULTS AND ANALYSIS

4.1 User-viewing activities and network path quality

A total of 22 subjects, 16 male and 6 female, participated in the subjective assessment. All of them were non-experts in assessing the video quality. Nine of them carried out the experiment through the campus network, and others accessed the experiment platform through the public Internet. All the subjects reported that they spent at least one hour on surfing the Internet, and 20 of them watched at least one video on the web in the week before performing the experiment. Therefore, it is reasonable to assume that they are familiar with video-watching applications.

Figure 3 shows the frequency distribution of the ratings collected in the assessment. Although we observe very low frequency for the two extreme ratings (1 and 7), unrealistically poor or good performance usually contain less information. Another possible reason is that the subjects, who are all Chinese, avoid giving extreme ratings because of stronger central tendency bias [7]. However, we believe that the frequency of the rating between 2 and 6 can already provide enough variance.

Figure 4 depicts the user-viewing activities recorded from one of the experiments. The bars shown along the x-axis are the cursor speed. The “×”s and “∘”s are the points when the buffer-empty and buffer-full events are triggered, respectively. The time that the subject pressed the pause button and the resume button are denoted by the “+”s and the “□”s, respectively. The “△”s and the “▽”s are the respective times that the subject switched to full screen mode and normal screen mode. The quality switching events are indicated by the “◁”s (for switching down) and “▷”s (for switching up). For a clearer illustration, the events are plotted in different levels.

Figure 4 shows that the application-level events, such as cursor movement, correlate with the user-viewing activities to some extent. In the first 50 seconds, the subject was still adapting to the network condition, and therefore the viewing activities are relatively few. After feeling that the playback is acceptable, she switched to a higher quality and watched in full screen mode just before the second buffer-empty event. However, after the second buffer-empty event occurs, the subject seemed to be annoyed by the event, switching back to the original quality quickly and returning to the normal screen size. She clicked the pause button and returned to normal screen mode until the end of the playback. After 200 seconds, she paused the playback for every buffer-empty event. Moreover, the cursor speed sharply increased whenever some user-viewing activities were captured, because the subject had to move the mouse cursor to press the pause buttons.

Figure 5, on the other hand, shows the round-trip time (RTT) and packet loss rate (aggregated every second) measured from the same viewing session. To facilitate the comparison, the two time axes are perfectly aligned. The median RTT is about 80 ms and the average packet loss rate is 4%. Loss busts are occasionally observed, but the network path was generally unchanged during the whole experiment session. The video was completely downloaded about 10 seconds before the video playback was finished. By comparing Figures 4 and 5, we observe that using only network path measurement could fail to capture the important information about the user perceived QoE of the video, such as her dissatisfaction after the second rebuffering event.

4.2 Hypothesis testing for user-viewing activities

To give a generalized view of the activities, we formulate hypotheses for the user-viewing activities from Hypothesis 1 and test each hypothesis through statistical tests. Using one-sample t test, we can obtain probability p that the null hypothesis is true for given mean and standard error [20]. If a user-viewing activity occurs after the impairment events,

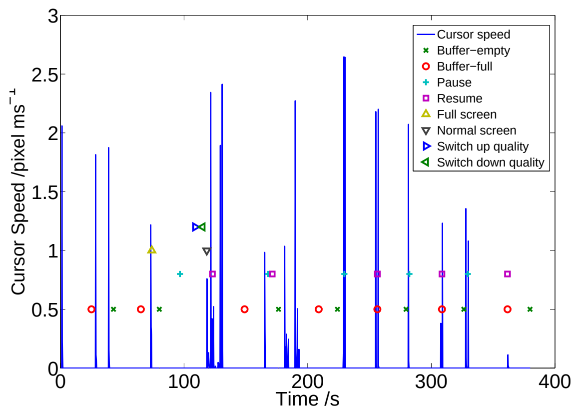

Figure 4: Time series of application-level events and user-viewing activities.

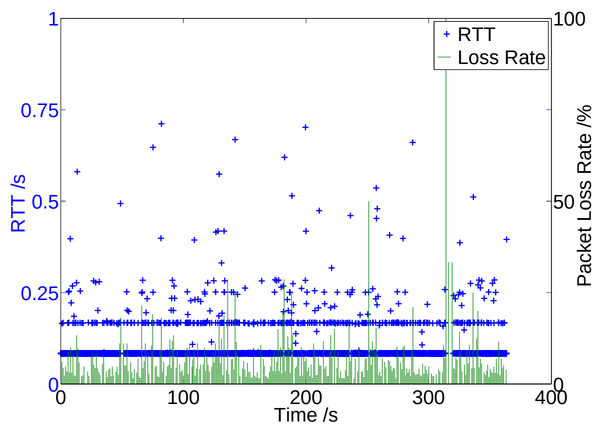

Figure 5: Time series of the RTT and packet loss rate.

the average time distance is positive (i.e., $\overline{D_i} > 0$). We choose rebuffering (buffer-empty) events as the impairment event, because it is the main factor affecting the perceived quality. However, our method can also be applied to other impairment events.

Table 4 shows the average time distance, $\overline{D_i}$, of three user-viewing activities. The results show that the average time distance for the pause activity is significantly larger than zero, which means that users pause the video playback around two seconds after she encounters rebuffering events. Similarly, users change back to the normal-size screen from full screen about three seconds after the occurrence of impairment events.

The few seconds of delay between the impairment events and the activities can be regarded as user's *reaction time*. The average time distance for reducing the screen size is about one second more than pausing, implying that users usually pause the playback before reducing the screen size. As users know that pausing is functional, they regard it as a more critical action than reducing the screen size.

Although the activities of switching to a lower quality have a positive mean, it is not statistically significant due to small sample size ($N = 3$). A small activity count reflects that the subjects prefer pausing instead of switching the quality. The results for mouse movement are also not significant in our analysis, indicating that the average cursor speeds before and after the impairments are very similar.

Table 4: Average time distance for different user-viewing activities and impairment events.

Activities	$\overline{D_i}$ (seconds)
Pause	1.94***
Switch to lower quality	2.19
Reduce the screen size	3.10**
	$\overline{v_t}$ (pixel/ms)
Mouse movement	-0.070

Note: *p <0.05, **p <0.01, ***p <0.001.

4.3 Correlating user-viewing activities with QoE

In [17], we have showed that the rebuffering frequency is the main factor affecting the QoE, in terms of the MOS. Our results show that all three application performance metrics (APMs) impacted the QoE, and we have also found that using log transformation to correct the functional form of the rebuffering frequency and the initial buffering time can obtain a better model fit. As the MOS is ordinal in nature, the ordinary least-square regression cannot be applied. We have therefore adopted the ordinal logistic regression [6] by using SPSS [2] in the analysis below.

The left column of βs in Table 5 shows the regression results of solely using the APMs proposed in [17], where f_{rebuf} is the rebuffering frequency, T_{init} is the initial buffering time, and $\overline{T_{rebuf}}$ is the mean rebuffering duration. The model is significant with a χ^2 of 14.3 on 3 $d.f.$, meaning that the original model better explains the variations in the MOS than an intercept-only model. Among the three APMs, the rebuffering frequency and the initial buffering time are significant, while the mean rebuffering duration is marginally significant. The negative β means that the odds (probability) of obtaining higher MOS categories decrease with the rebuffering frequency and/or the initial buffering time. This implies that a higher rebuffering frequency or a longer initial buffering time has a lesser chance of obtaining higher MOS categories. However, an increase in mean rebuffering duration has a slightly higher chance of obtaining higher MOS categories. We adopt one of the pseudo-R^2 metrics, Nagelkerke R^2 [18], which ranges from 0 to 1, to represent the goodness of fit of the model. The explanatory power is moderate with a Nagelkerke R^2 of 0.24.

Table 5: Regression results for the APM model and modified model.

Predictors	Estimates, βs	
	APM model	Modified model
$\ln(f_{rebuf})$	-0.54*	-0.62*
$\ln(T_{init})$	-0.60*	-0.54*
$\overline{T_{rebuf}}$	0.04 †	0.045*
N_{pause}	-	-0.68†
N_{screen}	-	0.94
Nagelkerke R^2	0.24	0.32
χ^2	14.3**	19.5**
$d.f.$	3	5
Valid N	54	54

Note: †p <0.1, *p <0.05, **p <0.01.

We further incorporate two of the user-viewing activities which show significant results in section 4.1 (i.e., pausing and reducing the screen size). We count the number of pause and screen size reducing activities which are probably triggered by the impairments, denoted by N_{pause} and N_{screen}, respectively. The right column of βs in Table 5 shows the regression results of the modified model which is also significant with a χ^2 of 19.5 on 5 $d.f.$ The rebuffering frequency which shows significant result in the original model is still significant. For the new factors, we have obtained a marginal significance for the pause activities. By adding these two factors, the explanatory power, measured by the Nagelkerke R^2, increases from 0.24 to 0.32. The negative β of N_{pause} means that the probability of obtaining a higher MOS category increases when less pauses are triggered by impairment events. On the other hand, switching the screen size has no effect to the odds of MOS categories.

5. RELATED WORK

Vilas and Paneda et al. modeled the user behavior of a VoD website [22], but they did not correlate the behavior with the QoE. In [24] and [10], a traffic analysis of a VoD system and an IPTV system was performed. However, the user behavior considered there was the access behavior, instead of the video watching behavior in each session.

Apart from video quality assessment, other types of user activities or responses had been investigated for inferring user perceived quality or user satisfaction. For example, Downing [8] used system usage behavior to infer the user satisfaction of the system. Eyetracks was used in [9] to predict the image quality. Clickstream data could also used to improve the ranking in web searching [4, 14].

In [17], we measured the QoE of HTTP video streaming under different network QoS and application QoS. We proposed three APMs, including the initial buffering time, rebuffering frequency, and mean rebuffering duration, to quantify the temporal structure impairments. We investigated the correlation among network QoS, QoE, and the APMs, and concluded that the rebuffering frequency was the main factor affecting the QoE. In this paper, we improve the prediction power of the model by including the pause events.

6. CONCLUSION

In this paper, we studied how the user-viewing activities help evaluate the QoE of HTTP video streaming. We proposed a new methodology of examining the user-viewing activities around the occurrences of impairment events. From our subjective measurement results, we found that the impairments can trigger pause and screen size switching events after two and three seconds, respectively. We then incorporated these triggered activities into the prediction model of the QoE to improve its prediction power. We also found that the pause activities are responsible for the variation of the MOS.

7. ACKNOWLEDGEMENT

We thank the reviewers for their useful comments. This work is partially supported by a grant (ref. no. ITS/355 /09) from the Innovation Technology Fund in Hong Kong and a grant (ref. no. H-ZL17) from the Joint Universities Computer Centre of Hong Kong.

8. REFERENCES

[1] Endace DAG card. http://www.endace.com.
[2] IBM SPSS Statistics. http://www.spss.com.
[3] Adobe. Video Technology Center, Delivery: Progressive download. http://www.adobe.com/devnet/video/progressive.html.
[4] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proc. ACM SIGIR*, 2006.
[5] A. Botta, A. Dainotti, and A. Pescape. Multi-protocol and multi-platform traffic generation and measurement. In *Proc. INFOCOM (DEMO Sess.)*, 2007.
[6] R. Brant. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46(4):1171–1178, 1990.
[7] C. Chen, S.-Y. Lee, and H. W. Stevenson. Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3):170–175, 1995.
[8] C. E. Downing. System usage behavior as a proxy for user satisfaction: an empirical investigation. *Information & Management*, 35:p.203–216, 1999.
[9] K. Fliegel. Eyetracking based approach to objective image quality assessment. In *Proc. ICCST*, 2008.
[10] V. Gopalakrishnan, R. Jana, R. Knag, K. Ramakrishnan, D. Swayne, and V. Vaishampayan. Characterizing interactive behavior in a large-scale operational IPTV environment. In *Proc. IEEE INFOCOM*, 2010.
[11] Y. Ito and S. Tasaka. Quantitative assessment of user-level QoS and its mapping. *IEEE Trans. on Multimedia*, 7 Issue 3:572–584, June 2005.
[12] ITU-R Recommendation BT.500-12. *"Methodology for the subjective assessment of the quality of television pictures"*, Sept 2009.
[13] ITU-T Recommendation P.911. *"Subjective audiovisual quality assessment methods for multimedia applications" Dec 1998.*
[14] T. Joachims. Optimizing search engines using clickthrough data. In *Proc. ACM SIGKDD*, 2002.
[15] J. Klaue, B. Rathke, and A. Wolisz. EvalVid - A framework for video transmission and quality evaluation. *Computer Performance*, 2794:255–272, 2003.
[16] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek. The click modular router. *ACM Trans. Comput. Syst.*, 18(3):263–297, 2000.
[17] R. Mok, E. Chan, and R. Chang. Measuring the quality of experience of HTTP video streaming. In *Proc. IEEE/IFIP IM (pre-conf.)*, 2011.
[18] N. J. D. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78 Issue 3:691–692, 1991.
[19] A. Perkis, S. Munkeby, and O. Hillestad. A model for measuring quality of experience. In *Proc. NORSIG*, 2006.
[20] F. Sani and J. Todman. *Experimental Design and Statistics For Psychology: A First Course*. Blackwell, 2006.
[21] M. Siller and J. Woods. Using an agent based platform to map quality of service to experience in conventional and active networks. *IEE Proc. - Communications*, 153(6):828–840, 2006.
[22] M. Vilas, X. Paneda, R. Garcia, D. Melendi, and V. Garcia. User behavior analysis of a video-on-demand service with a wide variety of subjects and lengths. In *Proc. EUROMICRO*, 2005.
[23] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang. Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In *Proc. ACM MM*, 2009.
[24] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng. Understanding user behavior in large-scale video-on-demand systems. *SIGOPS Oper. Syst. Rev.*, 40:333–344, April 2006.