ASSESSING THE VISUAL EFFECT OF NON-PERIODIC TEMPORAL VARIATION OF QUANTIZATION STEPSIZE IN COMPRESSED VIDEO

Zhili Guo[⋆] *Yao Wang*[⋆] *Xiaoqing Zhu*[†]

[⋆] Polytechnic School of Engineering, New York University
Department of Electrical and Computer Engineering, Brooklyn, NY 11201, U.S.A.
[†]Chief Technology and Architecture Office, Cisco Systems Inc., San Jose, CA 95134, U.S.A.

ABSTRACT

This work investigates the impact of non-periodic quantization step size variation on perceptual video quality. We composed non-periodic temporal variation patterns from periodic variation patterns. We present subjective test results and examine how does the overall quality under non-periodic variation relate to the qualities of individual periodic components, and how does it relate to the instantaneous qualities at different time points. We propose two quality models based on our observations and data analysis. Such quality assessment and modeling are essential for designing video adaptation strategy when delivering video over dynamically changing wireless links.

Index Terms— perceptual video quality, temporal variation, quality metrics

1. INTRODUCTION

In dynamic adaptive video streaming over HTTP (DASH) [1], a video is divided into multiple segments, and each segment is coded into multiple representations with different rates using different encoding parameters (e.g. frame rate, and frame size, and quantization parameter or QP). A receiver requests a particular representation for the next segment, based on the estimated channel throughput between the source and the receiver, as well as the buffer status at the receiver, and possibly other factors (e.g. the computation and display power of the receiver). A main challenge in DASH is how to determine which representation to request for the next segment, under hard-to-predict variation of the channel throughput.

Past approaches for video adaptation often try to avoid short term video rate variation in hope of producing a smooth quality. Therefore, when the available bandwidth experiences positive spikes, the player would choose to stay at the lower rate [2]. Our prior subjective studies have revealed that, contrary to the conventional wisdom, periodic variation between two video rates, obtained by varying either quantization stepsize (QS) or frame rate (FR), can lead to higher perceptual quality than staying at the lower rate, with the improvement depending on the variation magnitude, variation frequency, and video characteristics [3] [4].

Our prior subjective quality study of videos with varying quantization parameter (QP) in [3] was done for periodic variation patterns. In the current work, we composed serval more close-to-reality non-periodic variation patterns from previously used periodic patterns. Through perceptual video quality assessment, we found that the overall perceptual quality of a video with non-periodic QP variations can be largely predicted from the qualities of the composing frequency components. Furthermore, our results shows that the overall perceptual quality can also be predicted well by the median and minimum of the instantaneous qualities over the entire video duration.

This paper is organized as follows. Section 2 describes the design of non-periodic variation patterns to be used for testing and our subjective testing protocol. Section 3 presents the subjective test results and our observations, and also proposes two quality models for videos with non-periodic QP variation. Finally Section 4 concludes the paper.

2. VIDEO QUALITY ASSESSMENT CONFIGURATION

2.1. Design of Temporal Variation Patterns

Although we are mainly interested in assessing the effect of non-periodic variations, our test include both periodic patterns and non-periodic patterns that are composed by averaging of different periodic patterns. With such a design, we can evaluate how is the perceived quality of a non-periodic pattern related to the perceived quality of its underlying periodic components.

In our design, a periodic pattern has periodic pulse pattern as shown in Fig. 1, defined by three attributes, QP_{high}, QP_{low}, and F_z. The latter is equal to half of the period, and is inversely proportional of the variation frequency. We considered only $F_z = 1$ sec and $F_z = 3$ sec and the total sequence length is 10 sec. For a given F_z, we consider a total of 6 combination of QP_{low} and QP_{high}, as indicated in Table 1.

This work was partially supported by Cisco Systems, Inc.

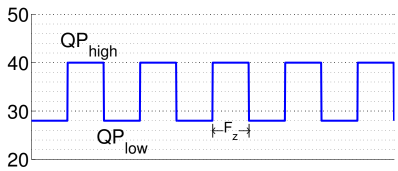


Fig. 1: A sample periodic pattern. Parameters in the figure are: $QP_{low} = 28$, $QP_{high} = 40$, $F_z = 1$ second, $QP_{avg} = 34$, and $\Delta QP = 12$.

Table 1: Periodic Patterns and Subjective Test Results

Frequency	QP		NMOS			
	QP_{low}	QP_{high}				
$F_z = 1$	28	32	0.97	0.96	0.94	0.99
$F_z = 3$			0.95	1.00	0.94	0.94
$F_z = 1$		36	0.71	0.81	0.89	0.80
$F_z = 3$			0.79	0.81	0.89	0.88
$F_z = 1$		40	0.49	0.57	0.82	0.63
$F_z = 3$			0.58	0.69	0.75	0.65
$F_z = 1$	32	36	0.71	0.74	0.86	0.69
$F_z = 3$			0.73	0.80	0.85	0.72
$F_z = 1$		40	0.47	0.60	0.74	0.53
$F_z = 3$			0.60	0.57	0.72	0.63
$F_z = 1$	36	40	0.44	0.48	0.67	0.46
$F_z = 3$			0.53	0.53	0.59	0.58

The last four columns reports the NMOS of the sequences "Stockholm", "Shields", "Soccer", and "News", respectively, from left to right.

We generate non-periodic patterns by linearly combining periodic patterns with different F_z. Although we only included F_z=1 and 3 in the tested periodic sequences, we also consider F_z=2 when composing the non-periodic patterns. Based on our prior study of periodic variation in CIF sequences [3], subjective quality for sequences with F_z=2 is very similar to those with F_z=1. Among all possible patterns that can be generated by averaging three periodic patterns with different QP_{low} and QP_{high}, we generate 20 patterns. The actual variation patterns are shown in Figure 2.

We used four HD(720p) resolution video sources: "Stockholm", "Shields", "Soccer", and "News" (also known as "Kristen and Sara"), each is 10 seconds long. For each source video, we create 37 test sequences. Five of them are with constant QP; 32 sequences are with different temporal variation patterns (12 periodic, and 20 non-periodic, as discussed above).

2.2. Video Perceptual Quality Assessment Platform

In order to make it easier for the human subjects to participate in the subjective video quality assessment, a web-based framework is used. Our platform is based on QualityCrowd2 [5]. We modified the source code, so that every user will be given a unique test session ID with random ordered video sequences (for the same content).

Similar to [3], in order to remove noisy ratings or outliers, we adopted, with some modification, the post-screening method recommended by BT.500-11 [6]. We recruited students in our campus to perform the subjective test. After the screening process there are on average 20 user ratings for each processed video sequence.

3. RESULTS AND ANALYSIS

3.1. Normalized MOS from Subjective Test

The subjective test results for periodic patterns are shown in last four columns in Table 1. They are presented as normalized mean opinion score (NMOS), which is the mean opinion score normalized by the MOS of the sequence with the highest quality (with constant $QP = 28$) for the same video source. Each column of the NMOS in Table 1 represents one video content. For the same variation frequency F_z, if ΔQP is fixed, the NMOS value is monotonically decreasing as QP_{avg} is increasing. When QP_{low} and QP_{high} are the same, higher variation frequency F_z causes worse NMOS except "Soccer" (third column of the NMOS result in Table 1) content which has very high motion.

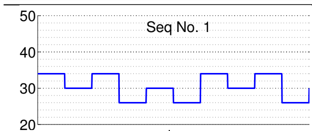
Similarly, in Table 2, the subjective test results for twenty non-periodic patterns are shown. The patterns are ordered by QP_{avg} ascending fashion. It is easy to observe that the NMOS is not monotonically decreasing as QP_{avg} is increasing. If we group these patterns based on their shape, such as grouping sequence no. 1, 10, 3, 6 and 12 as one group and no. 2, 7, 11, 14 and 20 as another group, it can be seen that for patterns within the same group and with the same ΔQP, their NMOS are monotonically decreasing with QP_{avg}. Furthermore, if QP_{avg} is very close, ΔQP and frequency will affect the quality. However, the trends of NMOS vs. ΔQP are not consistent across content. Thus, it is difficult to model the quality of sequences with non-periodic patterns based on the QP_{avg} and ΔQP.

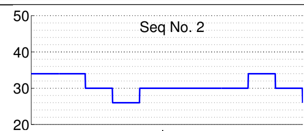
Note that for sequences without variation ($QP_{low} = QP_{high} = QP$), $NMOS$ vs QP can be accurately modeled by the NQQ model reported in [7]. Also, for periodic sequences, the observed trends and quality model proposed in [3] for CIF sequences are also applicable for HD sequences. In the following subsection, we focus on how to model the quality of sequences with non-periodic patterns. We use the first three video content ("Stockholm", "Shields" and "Soccer") as training set and use "News" as testing set for determining the model parameters.

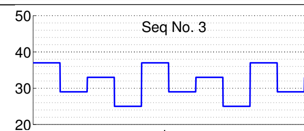
3.2. Overall Quality of a Video with Non-periodic Variation vs. the Qualities of its Periodic Components

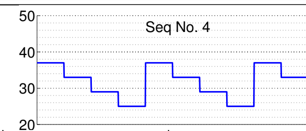
Recall that each non-periodic variation pattern is obtained by averaging two or three periodic variations. In this section, we examine the relationship between the quality of a video with a non-periodic pattern and the qualities of its periodic components.

Fig. 2: Non-periodic Variation Patterns and Subjective Test Results

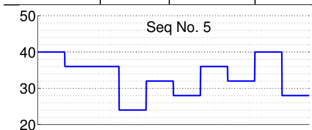


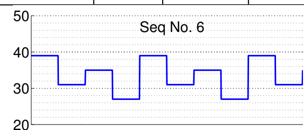


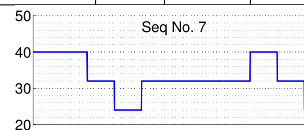


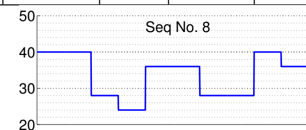


QP_{avg}	30.4	ΔQP	8	QP_{avg}	30.8	ΔQP	8	QP_{avg}	31.4	ΔQP	12	QP_{avg}	31.8	ΔQP	12
NMOS				NMOS				NMOS				NMOS			
0.85	0.89	0.99	0.93	0.85	0.91	0.88	0.91	0.75	0.77	0.87	0.79	0.77	0.73	0.78	0.76

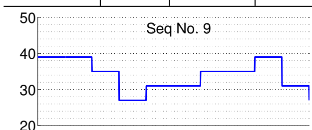


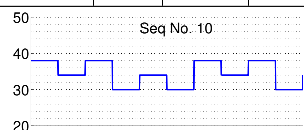


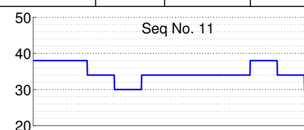


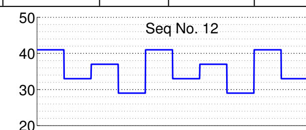


QP_{avg}	33.2	ΔQP	16	QP_{avg}	33.4	ΔQP	12	QP_{avg}	33.6	ΔQP	16	QP_{avg}	33.6	ΔQP	16
NMOS				NMOS				NMOS				NMOS			
0.70	0.61	0.79	0.63	0.69	0.70	0.78	0.68	0.66	0.65	0.69	0.60	0.60	0.63	0.73	0.63

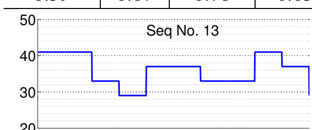


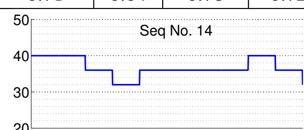


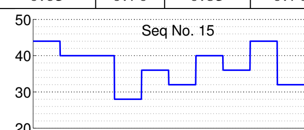


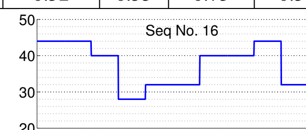


QP_{avg}	34.2	ΔQP	12	QP_{avg}	34.4	ΔQP	8	QP_{avg}	34.8	ΔQP	8	QP_{avg}	35.4	ΔQP	12
NMOS				NMOS				NMOS				NMOS			
0.59	0.67	0.78	0.63	0.71	0.64	0.78	0.72	0.65	0.70	0.83	0.70	0.52	0.55	0.73	0.58

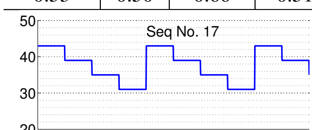


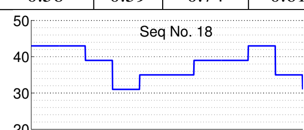


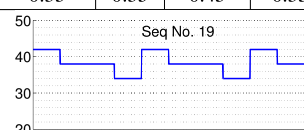


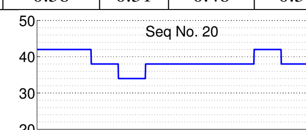


QP_{avg}	36.2	ΔQP	12	QP_{avg}	36.8	ΔQP	8	QP_{avg}	37.2	ΔQP	16	QP_{avg}	37.6	ΔQP	16
NMOS				NMOS				NMOS				NMOS			
0.53	0.50	0.66	0.51	0.58	0.59	0.74	0.61	0.35	0.35	0.45	0.35	0.38	0.31	0.48	0.34









QP_{avg}	37.8	ΔQP	12	QP_{avg}	38.2	ΔQP	12	QP_{avg}	38.4	ΔQP	8	QP_{avg}	38.8	ΔQP	8
NMOS				NMOS				NMOS				NMOS			
0.37	0.39	0.49	0.37	0.37	0.37	0.50	0.40	0.38	0.44	0.58	0.45	0.40	0.42	0.55	0.43

The NMOS values under each pattern are for sequences "Stockholm", "Shields", "Soccer", and "News", from left to right.

Figure 3 shows such relations where the periodic component is ordered by its own quality, with the first dominant component being the video with the lowest quality. Note that among 20 non-periodic test sequences for each source, only two have three components, while the remaining ones have two components only.

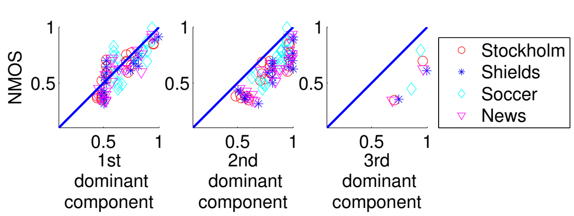


Fig. 3: Relation between the quality of a video with non-periodic variation (vertical axis) and the qualities of its three periodic components (horizontal axis).

As shown in Fig.3, the quality of the non-periodic sequence is quite linearly related to that of the first dominant component with a slope very close to 1. As for the second dominant component, its quality is generally higher than the overall quality. On the other hand, the overall quality is not very related to the quality of the third component. These results suggest that the perceived quality of a non-periodic sequence is mostly dominated by the component with the worst quality, and to a lesser extend, by the quality of the component with the intermediate quality.

Based on the above observation, we propose to model the overall quality by a simple model

$$Q_{v_{DM}} = NMOS_1 \cdot {NMOS_2}^{\kappa} \tag{1}$$

Through least squares fitting using the NMOS values observed for the sequences in the training set, we have found the optimal $\kappa = 0.2$. The scatter plots of the measured quality (vertical axis) and the model-predicted quality (horizontal axis) using the optimal κ are shown in Figure 4. The fitting accuracy of using the optimal κ is $pcc = 0.91$ and $MSE = 0.06$ for the training set, and $pcc = 0.92$, $MSE = 0.08$, for the testing set.

For the proposed model to be useful, one needs to be able to predict the quality of the individual periodic component based on its variation parameters (i.e., QP_{low}, QP_{high}, and F_z). Such a model was developed in our prior work for CIF test sequences [3]. We have validated that the model is still accurate for HD sequences, but the model parameter is resolution and content dependent.

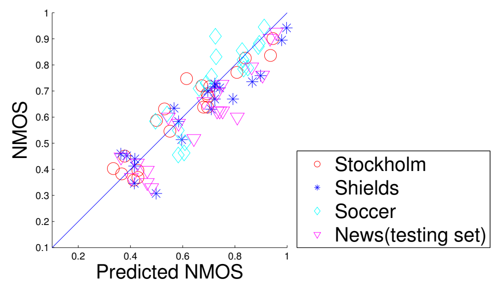


Fig. 4: Comparison between predicted value (using Eq.1) and original NMOS value

3.3. Overall Quality vs. Instantaneous Qualities at Different Time Points

Note that in each non-periodic QP variation pattern tested, the QP value changes at most once every second. In other words, for the entire duration (10 sec) of the test sequence, the QP is constant over every second (cf. figures in Table 2). We assign an instantaneous quality score for each second based on the QP value at that time, using our prior quality model relating the normalized quality with the QP [7]. We investigate how does the overall quality relate to some statistics of the instantaneous quality, including the average, median, and minimum. The relation between the overall quality and these statistics are shown in Figure 5.

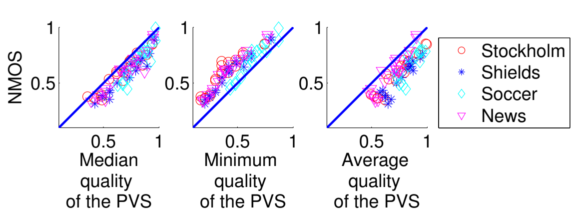


Fig. 5: Relation between non-periodic patterns and their quality statistics. There are twenty non-periodic patterns used in the subjective quality assessment. Note: PVS stands for "Processed Video Sequences"

From Figure 5, It is clear that the overall quality is more consistently related to the median and the minimum, than to the mean quality. Based on this observation, we propose the following model:

$$Q_v = \alpha Q_{median} + \beta Q_{min} \tag{2}$$

We determine a single set of model parameters using least squares fitting based on our training data. The model parameters are $\alpha = 0.68$ and $\beta = 0.33$. Figure 6 compares the predicted quality and the measured quality, using the model in Eq.2. The model accuracy for the training set are $pcc = 0.95$ and $MSE = 0.04$, and for the testing set are $pcc = 0.97$ and $MSE = 0.03$. We have also tried to use the weighted average of median, minimum and average quality to predict the overall quality. But the benefit from including the average quality is insignificant, and yet requires an extra model parameter.

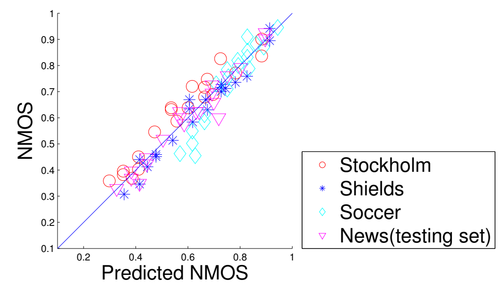


Fig. 6: Comparison between predicted value (using Eq.2, with $\alpha = 0.68$ and $\beta = 0.33$) and original NMOS value

It is quite promising that using the same model parameters, we were able to predict the overall quality of different source videos quite accurately. However, we should note that to derive the instantaneous quality corresponding to each QP, we used the NQQ model in [7], which involves a single content-dependent model parameter.

4. CONCLUSION

In this paper, we present the results of our subjective experiments to investigate the impact of non-periodic quantization step size variation on the perceived video quality. We observed several interesting trends. For example, the frequency component with the worst quality among all the frequency components affects the overall perceptual quality most. Furthermore, the median and minimum of the instantaneous qualities at different time points can be used to predict overall perceptual quality very well.

This paper focuses on the impact of non-periodic temporal variation of the quantization step size on the perceptual quality. Future studies will focus on utilizing our proposed quality model to guide the video representation selection in DASH. More specifically, given the past QP variation and the estimated future bandwidth (and hence feasible QP values for the current segment), how to select the QP of the next segment to optimize the overall perceptual quality.

5. REFERENCES

[1] Thomas Stockhammer, "Dynamic adaptive streaming over http –: Standards and design principles," in *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, New York, NY, USA, 2011, MMSys '11, pp. 133–144, ACM.

[2] Saamer Akhshabi, Sethumadhavan Narayanaswamy, Ali C. Begen, and Constantine Dovrolis, "An experimental evaluation of rate-adaptive video players over http," *Signal Processing: Image Communication*, vol. 27, no. 4, pp. 271 – 287, 2012.

[3] Yen-Fu Ou, Huiqi Zeng, and Yao Wang, "Perceptual quality of video with quantization variation: A subjective study and analytical modeling," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 1505–1508.

[4] Yen-Fu Ou, Yan Zhou, and Yao Wang, "Perceptual quality of video with frame rate variation: A subjective study," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, March 2010, pp. 2446–2449.

[5] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "Qualitycrowd - a framework for crowd-based quality evaluation," in *Picture Coding Symposium (PCS), 2012*, May 2012, pp. 245–248.

[6] "Methodology for the subjective assessment of the quality of television pictures," ITU-R Rec. BT.500-11, 2002.

[7] Yen-Fu Ou, Yuanyi Xue, and Yao Wang, "Q-star: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," *Image Processing, IEEE Transactions on*, vol. 23, no. 6, pp. 2473–2486, June 2014.