Perceptual Quality of Video with Periodic Frame Rate and Quantization Variation–Subjective Studies and Analytical Modeling

Yen-Fu Ou, Wenzhi Lin, Huiqi Zeng, Yao Wang

Abstract—In networked video applications, the frame rate (FR) and quantization stepsize (QS) of a compressed video are often adapted in response to the changes of the available bandwidth. It is important to understand how do the variation of FR and QS and their variation pattern affect the video quality. In this paper, we investigate the impact of temporal variation of FR and QS on the perceptual video quality. Among all possible variation patterns, we focus on videos in which two FR’s (or QS’s) alternate over a fixed interval. We explore the human responses to such variation by conducting subjective evaluation of test videos with different variation magnitudes and frequencies. We further analyze statistical significance of the impact of variation magnitude, variation frequency, video content, and their interactions. By analyzing the subjective ratings, we propose two models for predicting the quality of video with alternating FR and QS, respectively. The proposed models have simple mathematical forms with a few content-dependent parameters. The models fit the measured data very well using parameters determined by least square fitting with the measured data. We further propose some guidelines for adaptation of FR and QS based on trends observed from subjective test results.

Index Terms—perceptual video quality, frame rate, QS, temporal variation, quality metrics.

I. INTRODUCTION

In wireless video streaming, due to the limited sustainable bandwidth of a receiver, a video often has to be coded (or transcoded or extracted from a scalable stream) at reduced frame rate (FR), reduced frame size (FS), and/or increased quantization stepsize (QS), so that each coded frame has adequate quality. A critical issue is how to choose the appropriate spatial, temporal and amplitude resolutions (STAR), so as to achieve the best trade-off between picture quality and motion fluidity in the delivered video (Note that amplitude resolution is inversely related to QS). Another challenging problem is that the sustainable bandwidth of a network link often fluctuates in time, calling for adaptation of STAR. One naive approach would be to find the STAR that optimizes the perceptual quality over each short time duration based on the available instantaneous bandwidth. This may however create a video with rapidly fluctuating STAR, which may be annoying to the viewer. For example, variation in frame rate can cause visually annoying jitter artifacts. It is important to understand how does the variation of the STAR, individually and collectively, affect the perceived quality. Such understanding would enable us to impose proper constraints on the variation of the STAR, when adapting the STAR based on the time-varying bandwidth.

Take for example a hypothetical case where the available bandwidth alternates between \( R_l \) and \( R_h \), and the frame rates that can lead to the best perceived quality for constant rate video at \( R_l \) and \( R_h \) are \( t_l \) and \( t_h \), respectively. In this situation, is it better to code the video with alternating FR’s of \( t_l \) and \( t_h \), or would it be better to stay at \( t_l \)? More generally, one may want to vary not only the FR, but also the FS and QS to meet the instantaneous rate constraints.

There have been several studies regarding the influence of temporal and amplitude resolutions, individually or jointly, on the perceptual quality [1]–[9]. Some of these works (e.g. [1]–[3]) consider the cases where the FR and QS are fixed in the entire video, whereas some (e.g. [4], [5]) consider the impact of FR variation, due to non-uniform and bursty packet losses. Authors in both [4], [5] proposed quality models based on their subjective quality assessment, however, the model in [4] involves too many parameters (a total of 4 fixed constants) and the other [5] predict the quality index by non-linearly combining several sub functions, which can be used to measure different temporal features of video contents. Authors in [6] proposed a variable frame rate control scheme, which adapts the frame rate under fluctuating bandwidth environment. Although this scheme is developed by observing the human preference between fixed and variable frame rate videos they did not provide any mathematical function forms regarding the quality metrics. They provided guidelines for video adaptation instead. In [7], authors reported subjective test results for test videos with periodic variations of QS, while the FR and FS are fixed. However, they only considered QS variation under a fixed, relatively slow variation frequency (change every 5 sec), and no analytical models are proposed.

We have conducted subjective tests evaluating the impact of FR variations when QS and spatial resolution are fixed. Among the many possible patterns of temporal variations, we consider the simple case where the FR alternates between \( t_l \) and \( t_h \), with each FR staying over a constant time duration \( F_z \). See Fig. 1 for an illustration of how the FR and the corresponding video rate changes in time in our test sequences. We study the effect of \( t_h \), \( t_l \), their difference (variation magnitude) as well as the effect of \( F_z \) (inversely related to variation frequency) on the perceived quality. This study directly addresses the questions we raised for the hypothetical example given earlier. But it can also shed lights for more complicated cases where the FR may vary among more than two levels and the variation may
not follow a periodic pattern. We also conducted a parallel study of videos with periodic QS variation between qL and qH, and investigated the impact of qL, qH, and Fz on the perceived quality. Preliminary results of these studies have been reported in [8], [9].

In this paper, we report our subjective test results as well as analytical quality models for videos with FR and QS variations, respectively. Section II describes our subjective test configurations. Section III presents our study for videos with periodic variation of FR, including subjective test results, and proposed model that predicts the perceived quality based on tL, tH and Fz. Section IV presents the similar set of results, but for videos with QS variation. Combining the two subjective test results in aforementioned two sections, Sec. V compares quality of videos with FR and QS variations, respectively, under the same bit rate variation. Section VI investigates the statistical significance of impact of FR/QS variation and video content on perceptual quality using the ANOVA technique. Finally Section VII summarizes our main findings and propose several guidelines for video adaptation based on our findings.

II. SUBJECTIVE TEST SETUP

A. Testing Material

Our experiment is conducted using five video source sequences, Akiyo, Foreman, Football, Ice, Waterfall, all in CIF (352 × 288) resolution and at frame rate 30 fps originally 10 seconds long, which are chosen from JVT (Joint Video Team) test sequence pool [10]. All these sequences are coded using JVT scalable video model (JSVM912) [11]. For each sequence, one bitstream is generated with five quality layers, with corresponding FR of 1.875, 3.75, 7.5, 15, and 30 Hz, and each temporal layer in turn has five quality layers created with QP equal to 28, 32, 36, 40, and 44, respectively.

### Table I

| QS | Fz | tL(Hz) | tH(Hz) |
|----|----|--------|--------|
| 16 | 1/2/3 sec | 30 | 30/15/7.5 |
| 16 | 1/2/3 sec | 15 | 15/7.5 |
| 7.5 | 7.5 |

### Table II

| FR | Fz | QSb | QSs |
|----|----|-----|-----|
| 30 | 1/2/3 sec | 16 | 16/25/40/64/102 |
| 5 sec | 102 | 25/40/64/102 |

*Note that QS values of 16, 25, 40, 64, and 102 correspond to QP levels of 28, 32, 36, 40, and 44, respectively.

Fig. 1. The Variations of QS or FR for a video

---

not follow a periodic pattern. We also conducted a parallel study of videos with periodic QS variation between qL and qH, and investigated the impact of qL, qH, and Fz on the perceived quality. Preliminary results of these studies have been reported in [8], [9].

In this paper, we report our subjective test results as well as analytical quality models for videos with FR and QS variations, respectively. Section II describes our subjective test configurations. Section III presents our study for videos with periodic variation of FR, including subjective test results, and proposed model that predicts the perceived quality based on tL, tH and Fz. Section IV presents the similar set of results, but for videos with QS variation. Combining the two subjective test results in aforementioned two sections, Sec. V compares quality of videos with FR and QS variations, respectively, under the same bit rate variation. Section VI investigates the statistical significance of impact of FR/QS variation and video content on perceptual quality using the ANOVA technique. Finally Section VII summarizes our main findings and propose several guidelines for video adaptation based on our findings.

## II. SUBJECTIVE TEST SETUP

### A. Testing Material

Our experiment is conducted using five video source sequences, Akiyo, Foreman, Football, Ice, Waterfall, all in CIF (352 × 288) resolution and at frame rate 30 fps originally 10 seconds long, which are chosen from JVT (Joint Video Team) test sequence pool [10]. All these sequences are coded using JVT scalable video model (JSVM912) [11]. For each sequence, one bitstream is generated with five quality layers, with corresponding FR of 1.875, 3.75, 7.5, 15, and 30 Hz, and each temporal layer in turn has five quality layers created with QP equal to 28, 32, 36, 40, and 44 (with corresponding to QS = 16, 25, 40, 64, 102), respectively, using the coarse grain scalability (CGS) without QS cascading. The GOP size is set to 16 with only the first frame coded as in the I mode. Hierarchical-B structure is used to provide temporal scalability. For motion estimation, the FastSearch mode is enabled with maximum search range of 16 for full-pel search. We use SAD (Sum of Absolute Difference) as cost function for both full-pel and sub-pel. The entropy coding method is CAVL. The other encoding configurations follow the default settings in JSVM.

Two different experiments, examining quality impact of FR and QS variation, were conducted. For temporal variation, as shown in Fig. 1, we generate videos in which two frame rates switch back and forth periodically through the entire video with changing interval (Fz) of 1, 2, and 3 seconds. The QS is fixed at 16. Let tH and tL denote the higher and lower FR of the video. Table I details all the test configurations, which leads to a total of 90 processed (encoded and decoded) video sequences (PVS). For QS variation, we fix FR to 30 Hz but allow QS to switch back and forth periodically through the entire video with Fz of 1, 2, and 3 seconds. QSb and QSs denote the base (beginning) QS and deviated QS and the combinations of QSb and QSs are summarized in Tab. II. As a result, there are a total of 130 PVS’s.

We choose to examine the variation intervals of 1s, 2s, and 3s only, because this range represents most interesting cases to be studied, at least for an initial investigation. It is worth noting that, in video streaming, through the use of buffers, video rate variation can typically be limited to be no faster than changing every second. It is our hope that this initial study can provide important insight regarding the impact of the variation frequency and frequency magnitude on the perceptual quality.

### B. Subjective Test Configuration

The subjective quality assessment is carried out by using a protocol similar to ACR (Absolute Category Rating) described in [12]. Basically, each viewer is presented a series of video in a random order, and the viewer is asked to give overall rating of each video in the range of 0 to 100. Each test for one subject consists of two sessions, a training session and a test session. The training session (about 2 minutes) is used for the subject to accustom him/herself to the rating procedure and ask questions if any. The PVS’s in the test session (about 12 minutes) are ordered randomly so that each subject sees the video clips in a different order. Each test session contains only a subset of PVS’s with either FR or QS variations only. Most of the viewers are engineering students from Polytechnic Institute of New York University, with age 21 to 33. There are on average 22 ratings for each PVS with QS variations, and 20 ratings for each PVS with FR variations. Details regarding each experiment are can be found in [8], [9].

### C. Data Post Processing

The raw ratings are converted to Z-scores [13] based on the mean and standard deviation of all the scores of each viewer. In order to remove “noisy” ratings or outliers, we adopted, two post screening methods in concatenation. We first perform
BT.500-11 post screening method \cite{14} in Z-score domain to remove all ratings by certain viewers because their ratings are outside the range of the majority of the viewers. On average, one viewer is eliminated for each PVS. We then conduct the second step to the remaining ratings in the raw score domain using a ratio/averaging method. Basically we make use of the fact that a video coded at a lower FR (or higher QS) would not have a rating higher than a video coded at a higher FR (or lower QS), if the viewer’s judgement is consistent. Therefore, we remove all the ratings by a viewer for the same source video if this viewer violated this expected consistency test for more than a certain number of PVS’s for this source. More details can be found in \cite{8}, \cite{9}. After this step, on average, 17 ratings remain for each PVS.

We use MOS\((t_1, t_2)\) to represent the mean opinion score (MOS) for a PVS with two oscillating frame rates \(t_1\) and \(t_2\). In most practical applications, it is the relative MOS compared with some reference maximum MOS that is of importance. Therefore, we report and model the relative quality ratings defined by \(Q(t_1, t_2) = \text{MOS}(t_1, t_2)/\text{MOS}(t_{\text{max}}, t_{\text{max}})\), with \(t_{\text{max}} = 30\)Hz. Similarly, MOS\((q_1, q_2)\) denotes the MOS for a PVS with two oscillating QS’s \(q_1\) and \(q_2\), and \(Q(q_1, q_2) = \text{MOS}(q_1, q_2)/\text{MOS}(q_{\text{min}}, q_{\text{min}})\) with \(q_{\text{min}} = 16\). The test video and test results (in terms of MOS) are available at \cite{15}.

### III. Video Quality under Frame Rate Variation

This section considers the impact of periodic variation of FR on the perceptual quality, while the QS is fixed (QS=16). We first present results for videos with constant FR and show that the quality v.s. FR relation can be captured accurately by a model presented in \cite{14}. We then report subjective ratings for videos with periodic FR variations under different variation magnitudes and frequencies. Based on these results, we further propose a model that relate the quality with both the low and high frame rates.

#### A. Impact of Constant Frame Rate

Figure \ref{fig:fr_const} shows \(Q(t_1, t_h)\) v.s. \(t/t_{\text{max}}\) (here \(t_{\text{max}} = 30\)), of all the testing sequences with \(t_1=t_h=t\). As expected, the MOS reduces as the frame rate decreases. In our prior work \cite{3} that examined the effect of FR and QS on the perceptual quality, when each is held constant over the entire test sequence, we have found that given a fixed QS, the impact of FR can be captured by the following model, known as model for normalized quality vs. temporal resolution (MNQT),

\[
\text{MNQT}_{\text{fr}}(t) = \frac{1 - e^{-\alpha_t t/t_{\text{max}}}}{1 - e^{-\alpha_t}}. \tag{1}
\]

The model parameter \(\alpha_t\) characterizes how fast the quality drops as the frame rate reduces with smaller value corresponding to faster dropping rate. As can be seen from Fig. 2 this model fits with the subset of measured data \((t_h = t)\) very well with Pearson Correlation Coefficient (PCC)=0.995 and Root Mean Square Error (RMSE)=0.013.

#### B. Impact of Frame Rate Variation

We now consider sequences in which the frame rate alternates between \(t_h\) and \(t_l\). We first discuss, under the same average frame rate, \(t_{\text{avg}} = (t_0 + t_l)/2\), how does frame rate variation magnitude \(\Delta t = t_h - t_l\) affects the perceived quality. Figure 3 shows that, when the \(t_{\text{avg}}\) is the same, the MOS for a video with a constant frame rate (as indicated as cyan lines) is higher than that with frame rate variation. The degradation due to frame rate change is more severe when \(\Delta t\) is higher (e.g., quality difference between (30,7.5) and constant frame rate of (30+7.5)/2=18.75 is greater than quality difference between (30,15) and constant frame rate of (30+15)/2=22.5). This result is as expected, as large frame rate variation induces noticeable jitter. It is interesting to note that points corresponding to \(t_l=t_h/2\) are relatively close to the quality-frame rate curves achievable by using constant frame rates, for most of the sequences. But those with \(t_l\) lower than \(t_h/2\) are much below.

Figure 3 shows that using the quality model derived for constant FR, i.e., MNQT\((t_{\text{avg}})\) with \(t_{\text{avg}}=(t_h+t_l)/2\), to represent the quality of video with alternating FR’s \(t_h\) and \(t_l\) is not very accurate. Another intuitive approach would be to model the perceived quality of a video with alternating FR between \(t_h\) and \(t_l\) by the average of the quality of a video with constant FR=\(t_h\), predicted by MNQT\((t_h)\), and that of a video with constant FR=\(t_l\), predicted by MNQT\((t_l)\). The results are also illustrated in Fig. 3 where the navy solid lines and blue dotted lines are the predicted quality using this method when \(t_h\) equals 30 and 15 Hz, respectively. The results indicate that this method is more accurate than using MNQT\((t_{\text{avg}})\), but is still not very accurate especially when \(t_h\) is much larger than \(t_l\).

To examine whether alternating between \(t_l\) and \(t_h\) leads to better quality than staying at \(t_l\), we plots \(Q(t_h, t_1)/Q(t_l, t_1)\) against the ratio \(t_h/t_l\) in Figure 4 where it shows that alternating between \(t_l\) and \(t_h\) generally leads to a quality that is better than or similar to that obtained by staying at \(t_l\) (except for Foreman when \(t_{l}=7.5\) and Fz=1, and for Ice when \(t_{l}=15\) and Fz=2). The slope of improvement depends on Fz, \(t_l\) and the source sequence. However, the quality improvement becomes saturated as \(t_h/t_l > 2\). Furthermore, when \(t_h > 2t_l\), the quality reduces compared to when \(t_h=2t_l\), for some cases, possibly due to the noticeable frame rate variation.

We next look at when \(t_h\) is fixed, how the quality changes with different \(t_l\). Figure 5 shows how \(Q(t_h, t_l)/Q(t_h, t_{\text{avg}})\) decreases with the frame rate ratio \(t_l/t_h\). We can see that the dropping trend depends on \(t_h\), with a higher \(t_h\) leading to a slower dropping rate. The dropping rate is also influenced by Fz, although the trend is inconsistent.

We have found that the quality drop with the frame rate ratio can be modeled well using an inverse exponential function of the form

\[
\text{MNQT}_{\text{fr}}(t_h, t_l) = \frac{1 - e^{-\alpha_{t_l} t_h/t_{\text{avg}}}}{1 - e^{-\alpha_{t_l}}} = \frac{1 - e^{-\alpha_{t_l} t_h/t_{\text{avg}}}}{1 - e^{-\alpha_{t_l}}}, \tag{2}
\]

where the parameter \(\alpha_{t_l}\) depends on \(t_h\). This is based on our earlier observation that the influence of Fz on the dropping rate is inconsistent, and also based on the ANOVA analysis described in Tab. IV and Sec. VI-A. Fig. 6 presents the fitting
curves using (2).

C. Overall Quality Model

In order to predict the overall quality of a video with frame rates oscillating between $t_h$ and $t_i$, we make use of the fact that $Q(t_h, t_i)$ can be written as

$$Q(t_h, t_i) = Q(t_h, t_h) \frac{Q(t_h, t_i)}{Q(t_h, t_h)}.$$  (3)

Then we can use the model in (1) to predict the first term and use the model in (2) to estimate the second term. This yields the proposed quality model for videos with FR variation (to be called QTV) given below

$$QTV(t_h, t_i) = MNQT_\alpha(t_h)MNQT_\alpha(t_i, t_h).$$  (4)

$$= 1 - e^{-\alpha t_i} \left( 1 - e^{-\alpha t_h} \left( \frac{t_h}{t_i} \right) \right)$$  (5)

The first term in (4) predicts the quality achievable with a constant high FR $t_h$, and the second term estimates the quality reduction when the FR fluctuates between $t_h$ and $t_i$. Figure 7 shows the model curves fit the measured data very well. We note that these highly accurate results are obtained by finding best fitting model parameters for each sequence. Nonetheless, the fact that all sequences can be modeled well using the same model form is very encouraging. It suggests that the proposed simple model form reveals the
right trend of how does $t_l$ and $t_h$ each affects the overall quality. Additional study is needed to investigate the relation between $\alpha_q$ and $t_l$ and to develop a closed-form function possibly with content-dependent parameters. For the model to be useful for predicting quality of other sequences outside our test set, one must also investigate how to predict the model parameters from the video content.

IV. VIDEO QUALITY UNDER QS VARIATION

This section presents the impact of periodic variation of QS on the perceptual quality, while the FR is fixed. We first demonstrate results for videos with constant QS and show that the quality v.s. QS relation can be approximated well by a model presented in [3]. We then report subjective ratings for videos with periodic QS variations under different variation magnitudes and frequencies. Based on these results, we further propose a model that relate the quality with both the low and high QS.

A. Impact of constant QS

First we investigate the influence of the QS on the perceptual quality of a video when QS is constant. Figure [5] shows $Q(q_l,q)$ v.s. $q_{\min}/q$ for all the testing sequences with $q_h = q_l = q$. As expected, the quality reduces as $q$ increases (or $q_{\min}/q$ decreases). In our prior work [3], we have examined the impact of QS on the video quality, when the FR and FS are fixed. There we have found that the impact of QS can be captured by the following model, known as model for normalized quality vs. quantization (MNQQ),

$$\text{MNQQ}_e(q) = \frac{1 - e^{-\alpha_q q_{\min}}}{1 - e^{-\alpha_q l}}.$$  \hspace{1cm} (6)

As can be seen from Fig. [5] this model fits with the subset of measured data ($q_l = q_h$) very well, with PCC$=0.990$, RMSE$=0.045$. Note that the parameter $\alpha_q$ characterizes how fast the quality drops as the QS increases, with a smaller value corresponding to faster drop.

B. Impact of QS Variation

We now consider video clips in which the QS alternates between $q_h$ and $q_l$. We first discuss, under the same average QS, $q_{\text{avg}} = (q_h+q_l)/2$, how does QS variation affects the perceived quality. Figure [7] shows that, when $q_{\text{avg}}$ is the same, the quality for a video with a constant QS, $Q(q_{\text{avg}}, q_{\text{avg}})$, is higher than that with QS variation, $Q(q_h, q_l)$. Note that the solid grayish blue curve in Fig. [9] are predicted quality for a video with constant QS based on [4]. It shows that using the quality model derived for constant QS, i.e., MNQQ$_e(q_{\text{avg}})$ to represent the quality of video with alternating QS’s $q_h$ and $q_l$ is not very accurate. Another natural attempt would be to model the perceived quality of a video with alternating QS between $q_l$ and $q_h$ by the average of the quality of a video with constant QS$=q_h$, predicted by MNQQ$_e(q_h)$, and that of a video with constant QS$=q_l$, predicted by MNQQ$_e(q_l)$. The result is also included in Fig. [9] the cyan solid lines and yellow dash lines indicate the predicted quality using this method when $q_l$ equals 16 and 40, respectively. It can be seen that the prediction is not accurate especially when $q_l$ is low, e.g., 16.

We next examine when $q_h$ is fixed due to the lowest available bandwidth, whether alternating between $q_l$ and $q_h$ leads to better quality than staying at $q_h$. We plot $Q(q_l,q_h)/Q(q_h,q_h)$ against the ratio $q_h/q_l$ in Fig. [10]. For $q_h=102$, alternating between $q_l$ and $q_h$ is consistently better than staying at $q_h=102$, and the degree of improvement depends on Fz and the texture details of the video (e.g., Waterfall and Foreman have higher improvement ratio, e.g., up to 2). It can be observed that shorter Fz leads to less improvement. This is as expected as shorter Fz corresponds to more rapid QS switching, which can be more annoying to the human viewers. Interestingly, the slope of improvement saturates and becomes inconsistent as $q_l$ further decreases as $q_h/q_l$ becomes higher than 2.5. When $q_h$ is already low (e.g., $q_h=40$), switching to $q_l$ provides inconsistent gain for most sequences. For Waterfall, we see a consistent quality drop when Fz=1 and 2.
analysis (described in Sec. VI) shows that the quality variation observed for both \(q_i=25\) and \(q_i=16\) is statistically insignificant (see entries for \(P_2\) and \(P_3\) in Tab. VII).

Next, we examine how the quality reduces when the QS varies between \(q_i\) and \(q_t\), compared to the quality achievable with a constant QS=\(q_i\). Figure 11 plots \(Q(q_i,q_t)/Q(q_i,q_i)\) vs. \(q_i/q_t\). Here we can see that the falling rate depends on the video content and \(q_t\). The dropping rate is very similar for \(Fz=1\) and \(Fz=2\), but is typically slower when \(Fz=3\).

We have found that \(Q(q_i,q_t)/Q(q_i,q_t)\) can be modeled quite accurately by the inverted exponential function of the QS ratio \(q_t/q_i\), i.e.,

\[
\text{MNQQ}(q_i,q_t) = \frac{1-e^{-\alpha_{q_t}(Fz,q_t)}(\frac{q_i}{q_t})}{1-e^{-\alpha_{q_t}(Fz,q_t)}}, \quad (7)
\]

where \(\alpha_{q_t}\) is a model parameter reflecting the dropping rate. Based on the observations noted previously, and the ANOVA analysis described in Sec. VI-B and Tab. VII, we choose to use the same model parameter \(\alpha_{q_t}\) for \(Fz=1\) and \(Fz=2\), but a different one for \(Fz=3\). Figure 11 shows that the model predicts the measured data very well.

C. Overall Video Quality

In order to predict the overall quality of a video with alternating QS, we again recognize that \(Q(q_h,q_i)\) can be written as

\[
Q(q_h,q_i) = Q(q_i,q_i) \frac{Q(q_h,q_i)}{Q(q_i,q_i)} \quad (8)
\]

Then we can use the model in (7) to predict the term and use the model in (1) to approximate the second term. This yields the proposed quality model for videos with QS variation (to be called QQV) given below

\[
\text{QQV}(q_h,q_i) = \text{MNQQ}_h(q_t)\text{MNQQ}_t(q_i,q_h) = \frac{1-e^{-\alpha_{q_t}(Fz,q_t)}(\frac{q_i}{q_t})}{1-e^{-\alpha_{q_t}(Fz,q_t)}}, \quad (9)
\]

Figure 12 illustrates that the predicted quality fits with the measured data very well. Note that the first term in (6) is responsible for predicting the quality at a constant QS \(q_i\), whereas the second term accounts for the quality degradation due to QS variation.

V. ADAPTATION OF FR AND QS UNDER THE SAME RATE VARIATION

Recall that the motivation for this study is how to adapt a video when the available bandwidth changes. A natural question to ask is, given the rate variation pattern, which encoding parameter should we change, QS or FR? To answer this question, we find the bit rates corresponding to different QS and FR variations, and repplot the quality in terms of the same bit rate variation.

Let \(r_h\) and \(r_l\) denote the bitrate at \(t_f\) (or \(q_h\)) and \(t_b\) (or \(q_l\)), respectively. We plot the measured data, defined as \(Q(r_h,r_l) = \text{MOS}(r_h,r_l)/\text{MOS}(r_{max},r_{max})\) in terms of bitrate ratio \(r_h/r_l\) in Figs. 13, where the high rate point \((r_h)\) is achieved with \(FR=30\) and \(QS=16\). The rates corresponding to different FR and QS combinations are the actual bit rates when the test
videos are coded using the H.264/SVC encoder using the configuration described in Sec. II. Because the quality ratings between different Fz’s are quite similar, so we only present the cases with Fz=3 only. Figure 13 demonstrates the quality degradation v.s. rate ratio. It is clear that for the same periodic rate variation, videos with QS variation are more preferable than FR variation. We would like to stress that this result is for a particular high rate point when the video is coded at a very high quality. When the video is already at a relatively low rate (coded using a lower FR or higher QS), it is possible that lowering FR may be better than increasing QS, if one has to further reduce the bit rate periodically. Unfortunately, we don’t have subjective ratings for videos starting at a common encoding configuration at a low bit rate. Acquiring such data is beyond the scope of this paper.

VI. STATISTICAL ANALYSIS

The results presented in Sec. III and IV show that FR/QS variation magnitude, variation frequency and video content all affect the perceptual quality. To evaluate whether the changes in quality ratings due to these factors and their interactions are statistically significant, we perform three-way Analysis of Variance (ANOVA) [16]. With ANOVA, we compute the probability (p-value, which is derived from the cumulative distribution function of F based on the F-value) of the event that the difference in MOS when a particular variable is changed is due to chance. If this probability is low (p-value < 0.05), we consider this variable as having statistical significance on MOS.

**TABLE III**

| Cases       | Factors     | F-value | p-value |
|-------------|-------------|---------|---------|
| P1: All Data| Content     | 0       | 0.05    |
|             | Frequency   | 0       | 0.05    |
| P2: (30, 15) and (30, 30) | Δt | 1.74 | 0.187 |
| P3: (15, 7.5) and (15, 15) | Δt | 45.02 | 0.000 |
| P4: (30, 15) and (15, 15) | Δt | 44.52 | 0.000 |
| P5: (15, 7.5) and (7.5, 7.5) | Δt | 31.36 | 0.000 |

A. Analysis for Frame Rate Variation

In Tab. III results given under P1, are obtained by considering all test conditions together. We can see that FR variation magnitude (indicated by Δt=t2 - t1), variation frequency (inversely related to Fz), and video content all have statistically significant impact on the subjective ratings. In addition to conducting ANOVA over all data, we also looked at the significance of variation magnitude for a few specific cases.
The impact of FR variation is significant in all the cases except P2. This means that the quality rating difference between a video with constant FR of 30 Hz, and that alternating between 30 and 15 Hz, is mostly due to viewers’ inconsistency.

Although Tab. III shows that Fz has a significant impact on the absolute quality rating $Q(t_h, t_l)$, we are interested to know the statistical significance of the influence of Fz on the dropping trend of $Q(t_h, t_l)/Q(t_h, t_h)$ data, as mentioned in Sec. III-B. Toward this goal, we conduct a two-way repetition ANOVA test on the $Q(t_h, t_l)/Q(t_h, t_h)$ data. Results in Tab. IV suggest that there is no significant interaction between FR variation magnitude and Fz on the $Q(t_h, t_l)/Q(t_h, t_h)$ data, with $p$-value > 0.05 under all combinations of $t_h$ and $t_l$ examined.

### B. Statistical Significance for QS Variation

Results given under $P_1$ in Tab. V are obtained by considering all test conditions together. We can see that QS variation magnitude (indicated by $\Delta q = q_{h} - q_{l}$), variation frequency, and video content all have significant impact on the subjective ratings. In addition to conducting ANOVA over all data, we also looked at a few specific cases, each comparing ratings for two $(q_{1}, q_{2})$ pairs. The ANOVA results show that, when $q_{1}$ and $q_{2}$ are similar, e.g., P2 and P3, the quality difference due to QS variation is insignificant, which is as expected. In particular, this result confirms that the observed change when switching from $q_{h}=40$ to $q_{l}=25$ or 16 in Fig. 10 is not statistically significant.

To further investigate the dependency of the dropping trend of $Q(q_{1}, q_{2})/Q(q_{1}, q_{2})$ on Fz, we conducted a two-way ANOVA test on $Q(q_{1}, q_{2})/Q(q_{1}, q_{2})$ data. From Tab. VI, we learn that there is no significant difference between Fz=1 and 2 nor between Fz=2 and 3, but the difference between Fz=1 and Fz=3 is significant. This results is quite consistent with the observations from Fig. 11 and this is the basis we choose to use the same model parameter $\alpha_{q}$ for Fz=1 and Fz=2, but a different value for Fz=3 in the QGV model in (9).

### VII. Conclusion

In this paper, we report the results of our subjective experiments to investigate the impact of periodic FR/QS variation on the perceived video quality. We observed following interesting trends. Regarding the FR variation, firstly, under the same average FR the quality for a video with a constant frame rate is higher than that with FR variation, alternating between $t_{1}$ and $t_{3}$; and Secondly, the degradation due to FR change is more severe when $t_{3}/t_{1}$ ratio is higher, especially when $t_{h} > 2t_{l}$; Thirdly, alternating between $t_{1}$ and $t_{h}$ is generally better than staying at $t_{1}$. However, the quality improvement becomes saturated when $t_{h}/t_{1} > 2$. Finally the variation frequency does not have significant impact on the quality decay relative to the quality with a constant FR equal to $t_{h}$. The last two observations are somewhat surprising, and may not be true when Fz < 1 sec.

Similar phenomena have also been observed under QS variation. Firstly, under the same average QS, a video with a constant QS is perceptually more appealing than a video with variable QS under the same average QS. Secondly, the degradation due to QS change is more severe when constant low QS $q_{l}$ is relatively low and $q_{h}/q_{l}$ ratio is high. Thirdly, alternating between $q_{h}$ and $q_{l}$ is generally equal or better than a video with a constant high QS $q_{l}$. However, the improvement becomes saturated when $q_{h}/q_{l} < 0.4$. Unlike the case with FR variation, the variation frequency does affect the quality decay relative to the quality with a constant QS equal to $q_{l}$. Overall, slower variation leads to less quality decay, as expected.

These results provide several important guidelines for rate adaptation in video transmission over dynamically changing networks. For example, when the underlying application allows sufficiently long delay, it is better to code a video at constant FR and QS, and use a large buffer at the sender to deal with the bandwidth variation. When the delay constraint is stringent and one cannot afford to use large buffers, at least
in the range of variation frequency considered here ($Fz \geq 1$ sec), it is better to allow the rate to vary with the available bandwidth (by adjusting FR or QS), rather than staying at the lower rate. This suggests that trying to keep a smooth but low video quality may not be the right strategy for video streaming in dynamic networks. Allowing a certain amount of quality fluctuation may in fact lead to better quality. However, the system should limit the amount of changes in FR or QS. Specifically, when the lowest bandwidth demands a $FR = t_l$ and $QS=q_h$, the system should limit the FR to at most $2t_l$, and QS to at least $0.4q_h$, even if the instantaneous bandwidth allows higher $t$ or lower $q$ occasionally. When the rate reduction can be made by changing either QS or FR, it is suggested that increasing QS is preferred than decreasing FR to achieve the least quality degradation. By comparing video quality under the same periodic rate variation, we further found that adapting QS is better than changing FR, when the high rate point allows the video to be coded at very good quality (e.g. with high FR and low QS). Somewhat surprising, this is observed for video with either low or high motion.

Based on the observed relation between the quality and FR or QS variations, we propose several quality models. For FR variation, the model QT$V$ determines the quality of a video with alternating frames rates of $t_h$ and $t_l$ as the product of two terms. The first term reflects the quality of a video at a constant FR of $t_h$ and the second term indicates the degradation due to the switching between $t_h$ and $t_l$, and depends only on the ratio $t_l/t_h$. Similarly, we have developed QQV model for videos with QS variation. We note that these models are proposed based on the trends observed from the test sequences used in our subjective tests, and they need to be validated using other test sequences. In this study, the model parameters are derived using least square fitting of the model with the mean opinion scores for each sequence. For the model to be useful, one must also study how to predict the model parameters from the video content. Furthermore, rather than predicting the model parameter $a_l$ for each $t_l$, it may be possible to find a simple functional form to model $a_l(t_l)$. Similarly, one may examine and model the underlying relation of $a_q$ with $q_h$ and $Fz$. These are interesting directions for future studies.

This paper focuses only on the impact of periodic frame rate/QS changes on the perceptual quality. One challenging issue in modeling the impact of FR and QS variation on the quality is how to design the subjective test to help understand the effect of infinitely many likely variation patterns. Our subjective study that considers only periodic variations is only the first step towards this challenge, and is motivated by the fact that any pattern can be decomposed into periodic patterns with different frequencies. Our study so far considered only the effect of FR and QS variation, individually, and under a few variation frequencies. Future studies may examine a wider range of variation frequency, variation of spatial resolution, and joint impact of FR, QS and spatial resolution variation.

References

[1] ITU-T Recommendation G.1070, “Opinion Model for Video-Telephony Applications,” Apr. 2007.
[2] R. Feghali et al., “Video quality metric for bit rate control via joint adjustment of quantization and frame rate,” IEEE Trans. on Broadcasting, vol. 53, no. 1, pp. 441–446, Mar. 2007.
[3] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, “Perceptual Quality Assessment of Video Considering both Frame Rate and Quantization Artifacts,” IEEE Trans. On Circuit Syst. Video Technology (CSVT’11), vol. 21, no. 3, pp. 286–298, Mar. 2011.
[4] H.-T. Quan et al., “Temporal Aspect of Perceived Quality of Mobile Video Broadcasting,” IEEE Trans. on Broadcasting, vol. 54, no. 3, pp. 641–651, Sept. 2008.
[5] K.-C. Yang et al., “Perceptual Temporal Quality Metric for Compressed Video,” IEEE Trans. on Multimedia, vol. 9, pp. 1528–1535, Nov. 2007.
[6] V. Baroncini, D. Ciavatta, G. Gaudino, R. Felice, G. Iaconovi, and F. Ubaldi, “Variable frame rate video for mobile devices,” in Proceedings of the 2nd international conference on Mobile multimedia communications, ser. MobiMedia ’06. ACM, 2006, pp. 1–6.
[7] A. K. Moorthy, L. K. Choi, A. Bovik, and G. deVeciana, “Mobile Video Quality Assessment Database,” in IEEE International Conference on Communications (ICC), Jun. 2012.
[8] Y.-F. Ou, Y. Zhou, and Y. Wang. “Perceptual Quality of Video with Frame Rate Variation: a subjective study,” in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP10’), Mar. 2010, pp. 2446 – 2449.
[9] Y.-F. Ou, H. Zeng, and Y. Wang, “Perceptual quality of video with quantization variation: a subjective study and analytical modeling,” in IEEE International Conference on Image Processing (ICIP), 2012.
[10] [Online]. Available: http://wftp3.itu.int av-arch/jvt-site.
[11] Joint Scalable Video Model, “Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-X202,” Jul. 2007.
[12] ITU-R Rec. P.910, “Subjective Video Quality Assessment Methods for Multimedia Applications,” 1999.
[13] A. M. van Dijk, J. B. Martens, and A. B. Watson, “Quality Assessment of Coded Images using Numerical Category Scaling,” in Proc. SPIE, vol. 2451, 1995, p. 90101.
[14] ITU-R Rec. BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” 2002.
[15] NYUPoly Video Lab database:FR/QS Variation subjective quality evaluation. [Online]. Available: http://vision.poly.edu/index.html/index.php?n=HomePage.
[16] G. W. Snedecor et al., Statistical Methods, 8th ed Ames. IA: Iowa State Univ. Press, 1989.