A Constant Quality MPEG-1 Video Encoding Scheme and its Traffic Characterization*

İsmail Dalgıç Fouad A. Tobagi

Department of Electrical Engineering
Stanford University
Stanford, CA 94305

ABSTRACT: *In this paper, we introduce a video encoder control scheme for MPEG-1, which maintains the quality of the encoded video at a constant level. This scheme is based on a quantitative video quality measure, and it uses a feedback control mechanism to adjust the parameters of the encoder. We evaluate this scheme by applying it to several types of video sequences, and comparing it with the Constant Bit Rate (CBR) encoder control scheme in terms of quality and traffic. We show that our scheme achieves better quality than the CBR scheme for the same average bit rate, particularly when the scene content varies significantly over time.*

1 Introduction

In order to achieve high compression rates, today's prominent video encoding standards, such as MPEG and H.261, are based on lossy video compression algorithms. Such loss results in tiling, edge busyness, and mosquito noise in the encoded video, thereby reducing the perceived quality. The number of bits required to encode a scene so as to achieve a given quality objective depends on the scene content; the more complex the scene is, the more bits are required. Therefore, in order to achieve a constant video quality over time, the encoder parameters (in particular, the quantization scale) must be appropriately adjusted according to the scene content.

Most of the existing video encoders are designed for *Constant Bit Rate (CBR)* encoding, where the rate of the encoded video is kept constant at all times by dynamically adjusting the encoder parameters (typically the quantizer scale). This is achieved by placing a buffer at the output of the encoder, and using the buffer occupancy level as feedback to dynamically adjust the quantization scale. This scheme is motivated by the fact that some communications technologies, such as ISDN, as well as some storage technologies, such as CD-ROMs, are able to accommodate only constant bit rate streams. It is important to note that in a CBR encoded video stream, the quality varies in time, since the quantizer scale is controlled to achieve a constant bit rate regardless of the scene complexity.

*This work was supported in part by NSF under grant NCR-9016032, and by Pacific Bell.

Many researchers have also considered *Open-Loop Variable Bit Rate (OL-VBR)* encoding, whereby the quantization scale is simply kept at a constant value at all times. The motivation behind this scheme is to presumably provide a more consistent video quality compared to the CBR encoding by taking advantage of the variable bandwidth (and hence, the statistical multiplexing capability) supported by today's LAN technologies, by the newly emerging ATM technology, and by some storage technologies such as magnetic disks. Indeed, with OL-VBR encoding, a more complex scene is encoded using more bits; thus, the quality is less variable in time compared to CBR encoding. Nevertheless, it can be shown that there are still variations in quality.

It is possible to achieve constant quality video encoding if one were to use a quantitative video quality metric and a feedback control mechanism to adjust the encoder parameters. Indeed, we have presented this approach in [1] concentrating on H.261, using a quantitative video quality metric that has been developed at the *Institute for Telecommunication Sciences (ITS)*[2]. We have designed a Proportional-Integral-derivative (PID) feedback control function, and presented a preliminary characterization of the resulting traffic using some short test sequences (i.e., *miss america, susie, and table tennis*).

In this paper, we introduce a constant quality video encoder control scheme (CQ-VBR) for MPEG-1. Given the differences between MPEG-1 and H.261 in terms of the dependencies among frames, we have redesigned the PID feedback function appropriately for MPEG-1, such that it is stable, and has a sufficiently fast response time when reacting to the changes in the scene content. Using several long sequences, we present a characterization of the traffic resulting from the MPEG-1 constant-quality encoding scheme. For a more comprehensive characterization of quality and traffic for the CBR, OL-VBR, and CQ-VBR encoder control schemes using the MPEG-1, H.261, and Motion-JPEG encoding standards, please refer to [3].

The same approach taken in this paper could be applied to MPEG-2. However, the ITS metric is not appropriate for MPEG-2, as it has been calibrated and validated by subjective viewers for a level of quality achieved by MPEG-1 and H.261. Therefore, another quality metric which can work for MPEG-2 is needed, given the better quality achieved and the higher viewer expectation. Such a metric exists, and is presented in [4]. The suitability of this metric for assessing MPEG-2 quality is examined in [5]. Also note that MPEG-2 presents a richer set of parameters controlled, thus allowing a more precise control of quality.

2 ITS Quantitative Video Quality Metric

A quantitative video quality metric has been designed at the Institute for Telecommunication Science (ITS) that agrees closely with quality judgments made by a large number of viewers [2]. To design this metric, the authors first conducted a set of subjective tests. The video impairments used in those tests included digital video compression systems operating at rates around 700 kb/s and lower.

The ITS quantitative video quality metric $\hat{s}$ takes values in the range between 1 and 5 (1 being very annoying and 5 imperceptible) and is a linear combination of three quality impairment measures. Those three measures were selected among a number of candidates such that their combination matched best the subjective evaluations. Those three measures are in turn based upon two quantities, namely, *spatial information* (SI) and *temporal information* (TI). SI and TI are defined on a frame by frame basis. To obtain a single scalar quality estimate for each video sequence, SI and TI values are time-collapsed into three measures m_1, m_2, and m_3, which are then linearly combined to get the final quality measure. m_1 is a measure of spatial distortion, including blurring and false edges. m_2 and m_3 are both measures of temporal distortion. m_2 detects the amount of lost motion energy in the degraded video, such as during frame repetition. m_3 selects the largest amount of added motion in the sequence, which may be in the frame with the maximum jerky motion due to frame dropping, or in the frame with the worst uncorrected block errors due to losses in transmission. For precise definitions of SI, TI, m_1, m_2, and m_3, see [2].

The ITS quality measure $\hat{s}$ is given in terms of m_1, m_2, and m_3 by

$$\hat{s} = 4.77 - 0.992m_1 - 0.272m_2 - 0.356m_3.$$

Note that the definition of $\hat{s}$ given above implies that the quality of each test sequence is represented by a single number. However, for a long sequences, which would most likely contain multiple scenes, it is more meaningful to measure the quality for short time intervals, therefore capturing the quality variations over time. In the sequences we use, there are several scenes which are only one or two seconds long. The interval for measuring the quality should also be chosen large enough to correspond to the response time of the human visual system. With these considerations in mind, in this paper we opt to measure the quality in one-second intervals.

3 MPEG-1 Constant Quality Variable Bit Rate Encoding Scheme

To encode video streams at a constant quality $\hat{s}_{target}$, we have devised a feedback control scheme, where we measure the quality $\hat{s}(k,w)$ at every sampling point k using the last w frames, and use the difference ($\hat{s}_{target}$-$\hat{s}(k,w)$) as feedback to adjust the quantizer scale q. The block diagram of the encoder for this scheme is depicted in Figure 1. The design problem to be solved here is to choose the feedback function f and the quality estimation interval w appropriately so as to cause neither instability, nor a slow response time. Similarly to the H.261 constant-quality scheme [1], here we have also considered a feedback function of the PID (Proportional, Integral, Derivative) type. The PID feedback function is given by

$$\begin{aligned} q(k+1) &= K_p e(k) + K_p \frac{T}{T_I} \sum_{i=1}^{k} e(i) + q(0) + \\ & K_p \frac{T_D}{T}[e(k) - e(k-1)]. \end{aligned}$$

where $e(k) = \hat{s}(k,w) - \hat{s}_{target}$, and T is the sampling period of the system.

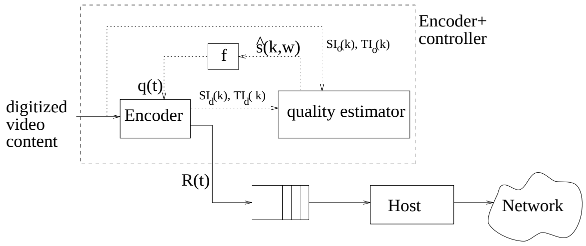
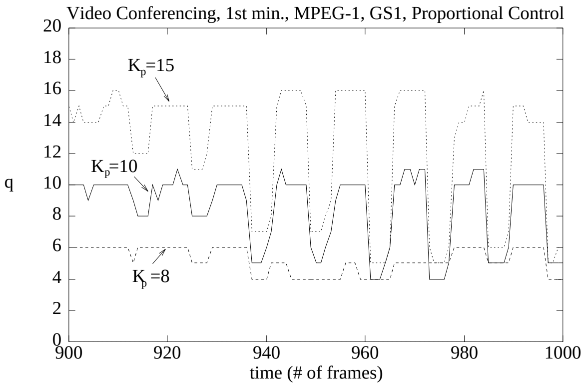
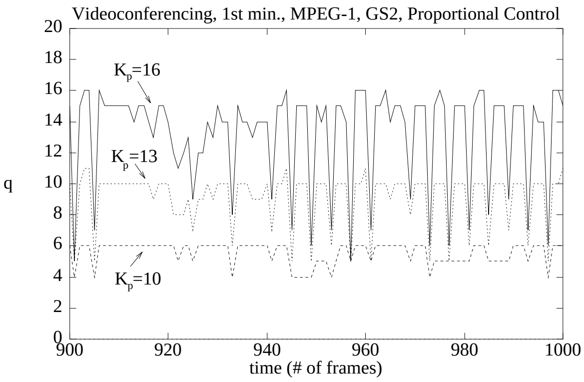


Figure 1: Block diagram of the encoder for Constant Quality VBR encoding

In MPEG, the design of the feedback function depends on the GOP structure. The reason is that the B frames depend on the information in the future frames. Therefore, when there are B frames, the frames are encoded at a different order compared to how they are displayed. As a result of this, the quality metric $\hat{s}$ cannot be computed at every frame interval; instead, it can only be computed at those time instants where the last B frame before an I or P frame is encoded. Therefore, the maximum possible sampling rate of the controller is equal to the number of consecutive B frames plus one. In this paper we have focused on two GOP structures: (i) "IBBPBBPBBPBBI...," which we refer to as GOP Structure 1 (GS1), and (ii) "IPPPPPPPPPPPI...," which we refer to as GOP



(a) GS1



(b) GS2

Figure 2: q versus time for proportional control, Videoconferencing sequence, $\hat{s}_{target}$=4.0, GS1 and GS2.

Structure 2 (GS2). Clearly, the maximum sampling rate for GS1 is equal to one-third of the frame rate, and for GS2 it is equal to the frame rate. Since higher sampling rates result in better performance in a feedback controller, we use the above specified sampling rates in the design of our controller, which are the highest possible.

We have designed the appropriate PID feedback control function for MPEG-1 with these GOP structures by employing the method of "Ziegler-Nichols PID tuning using the stability limit" [6]. The method works as follows (see [6] for details). The system is first controlled using only proportional control. The gain, K_p, is increased until continuous oscillations result, at which point the gain, K_u, and the oscillation period, P_u, are recorded. The PID gains are then determined as follows: $K_p = 0.6K_u$, $T_I = P_u/2$, and $T_D = P_u/8$. We have applied this procedure iteratively for various values of K_p, and for $\hat{s}_{target}$={3.5,4.0,4.5}. In those experiments we have used 3 one-minute video sequences, one taken from an advertisement, the other taken from the movie Star Trek, and the third one being a videoconferencing type sequence. In Figure 2, we show the quantizer scale q as a function of time for the videoconferencing sequence, proportionally controlled with various values of K_p; part (a) of the figure is for GS1, and part (b) is for GS2. It is clear from the figure that for GS1 the continuous oscillations start when the gain K_p is around 10, and for GS2 they start when K_p is around 13. Therefore, for GS1, K_u=10, and P_u=12 frame intervals, and for GS2, K_u=13, and P_u=4 frame intervals. Note that, when the sampling rate of a system is low, it is recommended to use a T_D coefficient larger than that suggested by the Ziegler-Nichols method [6]. For GS1, we have therefore experimented with various values of T_D, and found that $T_D = P_u/4$ gives the best results. Therefore, the PID coefficients for GS1 are given as $K_p = 6$, $T_I = 2T$, and $T_D = T$, and for GS2, they are given as K_p=8, T_I=2T, and T_D=0.5T. (By contrast, the PID coefficients for the H.261 controller are $K_p = 12$, $T_I = 2T$, and $T_D = 0.5T$.) We have also experimented with various values of w, and determined that $w = 3$ frames gives the best results for both GOP structures (similarly to the H.261 controller).

4 Traffic Characterization for the MPEG-1 CQ-VBR Scheme

The CQ-VBR scheme achieves a constant quality; however this comes at the cost of variable-rate traffic. On the other extreme, the CBR scheme achieves a constant bit rate over time at the cost of achieving variable quality. The Open-Loop VBR scheme lies in the middle of these two schemes, having somewhat less variable traffic but more variable quality as compared to the CQ-VBR scheme, and conversely, more variable traffic but less variable quality as compared to the CBR scheme.

In this section, we provide a comparison of traffic and quality for CQ-VBR and CBR by taking the following approach. We first consider video sequences which are encoded with the CQ-VBR scheme at a given target quality. We then consider the same video sequences encoded with the CBR scheme at the rate equal to the average rate of the corresponding CQ-VBR encoded sequence. (Note that the rate-control buffer size for the CBR encoding is chosen at the maximum size allowed by the MPEG-1 constrained parameter set, i.e., 327680 bits.) We compare the CBR and CQ-VBR sequences in terms of traffic and quality.

We present numerical results for five video sequences, each of them one-minute long: (i) a video conferencing type sequence, which consists of a person sitting and talking in front of a camera; (ii) an excerpt from the movie *Star Trek VI* where there is a large amount of motion, including abrupt camera movements, and 29 scene cuts; (iii) an excerpt from the movie *Indiana Jones: Raiders of the Lost Ark*, where again there is a large amount of motion, with some fast panning, and forward camera obstruction; (iv) an excerpt from the movie *Terminator 2*, where the there is not as much motion as in the above two

	CQ-VBR, $\hat{s}_{target}$=4.0										CBR at the same avg. rate				
	$\hat{s}$					Frame Size (kbits)					$\hat{s}$				
	avg	st.d	min	max	span	avg	st.d	min	max	p/a	avg	st.d	min	max	span
Vcnf	4.05	0.05	3.93	4.12	0.19	6.8	6.6	1.0	74.2	10.9	4.06	0.12	3.83	4.47	0.64
T-2	3.96	0.16	3.68	4.32	0.64	7.9	7.4	1.1	90.0	11.4	3.92	0.35	3.22	4.54	1.32
ST	3.96	0.11	3.78	4.18	0.40	11.5	8.7	1.3	96.4	8.4	3.88	0.69	2.91	4.63	1.72
Rdrs	3.98	0.14	3.74	4.33	0.59	11.1	10.8	2.2	124.8	11.3	3.74	0.54	2.27	4.51	2.24
Cms	4.01	0.12	3.72	4.26	0.54	21.4	12.1	1.4	204.7	9.5	3.61	0.47	1.00	4.70	3.70

Table 1: Quality and frame size statistics for all five sequences, MPEG-1, GS1, CQ-VBR, $\hat{s}_{target}$=4.0, and the quality statistics for CBR sequences encoded at the same average rate as their CQ-VBR counterparts.

	CQ-VBR, $\hat{s}_{target}$=4.5										CBR at the same avg. rate				
	$\hat{s}$					Frame Size (kbits)					$\hat{s}$				
	avg	st.d	min	max	span	avg	st.d	min	max	p/a	avg	st.d	min	max	span
Vcnf	4.51	0.07	4.26	4.68	0.42	12.8	11.1	1.0	119.8	9.3	4.49	0.08	4.24	4.63	0.39
T-2	4.50	0.05	4.39	4.60	0.21	23.7	17.8	3.3	129.5	5.5	4.50	0.10	4.22	4.65	0.43
ST	4.48	0.09	4.24	4.66	0.42	20.9	18.9	2.9	136.8	6.5	4.39	0.17	3.96	4.63	0.78
Rdrs	4.53	0.03	4.41	4.60	0.19	51.6	26.3	6.4	151.1	2.9	4.44	0.10	4.09	4.62	0.53
Cms	4.51	0.15	4.26	4.71	0.45	49.4	35.6	1.03	272.9	5.5	4.42	0.16	4.14	4.75	0.61

Table 2: Quality and frame size statistics for all five sequences, MPEG-1, GS1, CQ-VBR, $\hat{s}_{target}$=4.5, and the quality statistics for the corresponding CBR sequences encoded at the same average rate as their CQ-VBR counterparts.

sequences; this sequence contains a mixture of real and synthetic images (with relatively sharp edges), where the T1000 “morphs”, changing its shape from the floor tiles to the human form; and (v) a sequence of three commercial advertisements, some of which contain synthetic images (cartoons), with a very large amount of motion. The sequences are encoded using a software MPEG-1 encoder, developed by the Portable Video Research Group (PVRG) at Stanford University [7]. The picture format used is SIF, the frame rate is 30 per second, and the GOP size is 12 frames.

We have encoded all five sequences using CQ-VBR at $\hat{s}_{target}$=4.0, and $\hat{s}_{target}$=4.5, for both GS1 and GS2. Here we present our results for the GS1, noting that the traffic and quality statistics for GS2 have been found not to be significantly different. In Table 1, we show various quality and traffic statistics for CQ-VBR at $\hat{s}_{target}$=4.0. The particular quality statistics shown are the average, standard deviation, minimum, maximum, and the span between the maximum and minimum values of $\hat{s}$, where $\hat{s}$ is measured over one-second intervals. The traffic statistics shown are the average, standard deviation, minimum, maximum, and peak-to-average ratio of frame sizes. In the same table, we also show the same quality statistics for the corresponding CBR-encoded sequences encoded at the rate equal to the average rate of the corresponding CQ-VBR encoded sequence. Clearly, the CQ-VBR scheme is able to maintain a much more consistent quality level as compared to the CBR scheme. In all cases, the CQ-VBR average quality is very close to the target quality, and deviations from the average are confined to within ± 0.3. By contrast, the corresponding CBR sequences exhibit a large variation in quality; in particular, the minimum $\hat{s}$ values for the *Commercials*, *Raiders*, and *Star Trek* sequences are very low, and accordingly, the span between the minimum and the maximum $\hat{s}$ values are significantly higher than those for CQ-VBR. In these sequences, some scenes have different characteristics from the others. Therefore, in order to maintain a constant level of quality, they should be encoded at different bit rates. To illustrate this, in Figure 3 we plot the $\hat{s}$ and frame size as a function of time for the *Star Trek* sequence, using CQ-VBR at $\hat{s}_{target}$=4.0, and in Figure 4 we plot the same quantities when the sequence is CBR encoded at the same rate as the average rate of the CQ-VBR sequence. It is clearly seen in Figure 3 that there are some parts of the sequence where the resulting frame size statistics (and therefore the bit rate) is significantly larger as compared to the other parts (e.g., around frame 400, and frame 1300). By contrast, since the CBR sequence has to maintain the same average bit rate over time, the resulting quality gets lower for those scenes, especially around frame 1300.

Now consider the *Videoconferencing* sequence. Table 1 indicates that for this sequence the average quality for CQ-VBR and CBR schemes are about the same, and the minimum quality for the CBR scheme is only 0.1 unit smaller than that for the CQ-VBR scheme. This is expected since the contents of this sequence do not vary significantly over time. To show that this is indeed the case, in Figure 5 we show the quality and frame size as a function of time for the CQ-VBR *Videoconferencing* sequence, and in Figure 6 we show the same for the corresponding CBR sequence. It is clear that the traffic produced by the two sequences look remarkably similar, and exhibit little variation over time (except for an initial transient period, which lasts about 40-50 frame times in both cases).

The average frame size for the CQ-VBR sequences encoded at $\hat{s}_{target}$=4.0 show a strong content dependence, varying by a factor of 3 from *Videoconferencing* to *Commercials*. In all cases, the peak-to-average frame size ratios are around 8–11.

We now consider $\hat{s}_{target}$=4.5. In Table 2, we show for $\hat{s}_{target}$=4.5 the same statistics as in Table 1. The CQ-VBR scheme is again able to maintain a consistent level of quality, matching on average the target quality, and generally exhibiting a smaller span compared to $\hat{s}_{target}$=4.0. However, when we examine the quality statistics for the corresponding CBR encoded sequences, we see that the quality differences between the CQ-VBR and CBR sequences are not as large for $\hat{s}_{target}$=4.5 as for $\hat{s}_{target}$=4.0. In this case, smaller quantizer scales are used in order to achieve a better quality, which decreases the variations in quality [3]. However, it is also important to note that the average frame size required to achieve a quality level of 4.5 varies by a factor of 4 from the *Videoconferencing* sequence to the *Raiders* and *Commercials* sequences. If the CBR scheme is used for encoding and transmitting, say, a TV program, then a high bit rate must be chosen to ensure that all the scenes are encoded without excessive quality degradation; the same bit rate must also be used for those scenes which could be encoded at a lower bit rate and still incur a small quality degradation. By contrast, the CQ-VBR scheme allows the bit rate to be automatically adjusted; thus the given quality objective is maintained using only as many bits as required. For example, if the five sequences considered here were to be sent back-to-back with the CQ-VBR scheme, the resulting average rate would be about 950 kb/s. On the other hand, in order to maintain a similar level of quality at all times with the CBR scheme, they would have to be encoded at a rate more than 1.5 Mb/s. Of course, the use of the CQ-VBR scheme requires that the underlying network is capable of accommodating a variable bit rate by means of statistical multiplexing. We are currently evaluating the performance of the CQ-VBR scheme for various networking technologies, including ATM, 10Base-T and 100Base-T Ethernets (single-segment and switched multi-segment), IEEE 802.5 Token-Ring, FDDI, and 100VG-AnyLAN.

5 Conclusions

We have devised a feedback control based constant-quality video encoding scheme for MPEG-1. We have demonstrated that the scheme indeed is able to maintain a consistent quality level. We have shown that at the same average rate, the CBR scheme sometimes attains a much lower quality compared to the CQ-VBR scheme, especially when the scene content is variable in time. Conversely, if a good quality is to be maintained at all times with the CBR scheme, then a large bit rate has to be used; by contrast, the same quality level can be attained by the CQ-VBR scheme using a smaller number of bits on average. The CQ-VBR scheme can be utilized on networks which are capable of accommodating a variable bit rate by means of statistical multiplexing. We are currently evaluating the performance of the CQ-VBR scheme for such networking technologies.

References

[1] İ. Dalgıç and F. A. Tobagi, "Constant Quality Video Encoding," in *IEEE ICC'95*, (Seattle, Washington, USA), pp. 1255–1261, June 1995.

[2] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An Objective Video Quality Assessment System Based on Human Perception," in *SPIE Human Vision, Visual Processing, and Digital Display IV*, vol. 1913, (San Jose, CA), pp. 15–26, Feb. 1993.

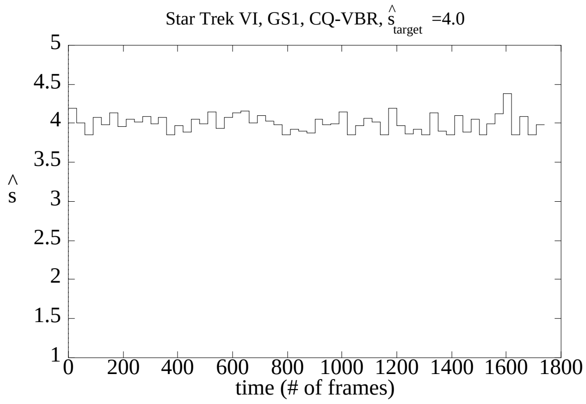
[3] İ. Dalgıç and F. A. Tobagi, "Characterization of Video Traffic and Video Quality for Various Video Encoding Schemes and Various Encoder Control Schemes," Computer Systems Laboratory Technical Report, Stanford University (in preparation).

[4] C. J. van den Branden Lambrecht and O. Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System," in *Proceedings of the IS&T Symposium on Electronic Imaging: Science and Technology,* Digital Video Compression: Algorithms and Technologies 1996, (San Jose, CA), The Society for Imaging Science and Technology, January 1996.

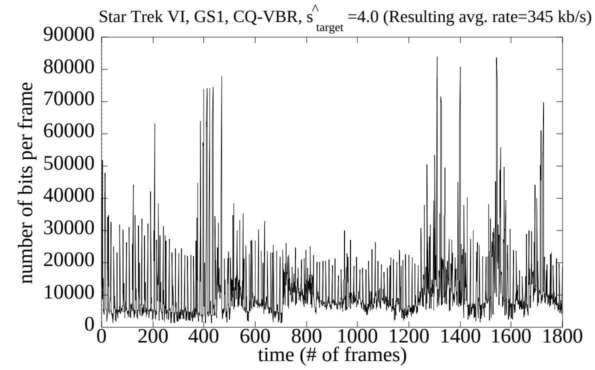
[5] A. Basso, İ. Dalgıç, F. A. Tobagi, and C. J. van den Branden Lambrecht, "Study of MPEG-2 Coding Performance based on a Perceptual Quality Metric," to appear in *1996 Picture Coding Symposium*, Melbourne, Australia, March 1996.

[6] G. F. Franklin, J. D. Powell, and M. L. Workman, *Digital Control of Dynamic Systems (second edition).* Addison-Wesley Publishing Company, 1990.

[7] "The PVRG MPEG-1 encoder is available via anonymous ftp from havefun.Stanford.EDU."

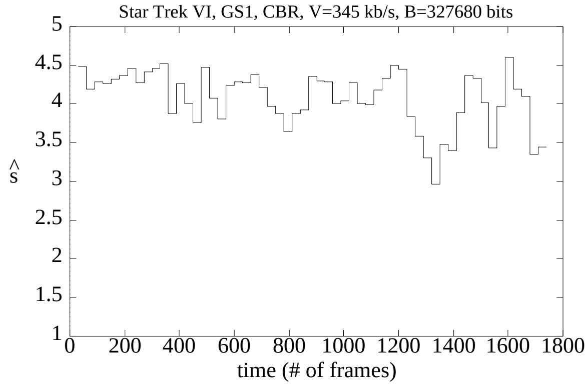


(a) $\hat{s}$ vs. time

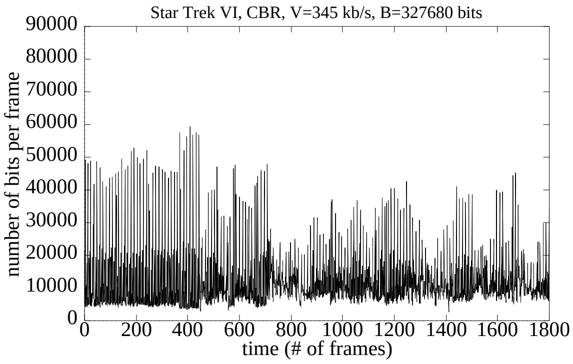


(b) Frame size versus time

Figure 3: Quality and frame size versus time for CQ-VBR, Star Trek VI, $\hat{s}_{target}$=4.0.

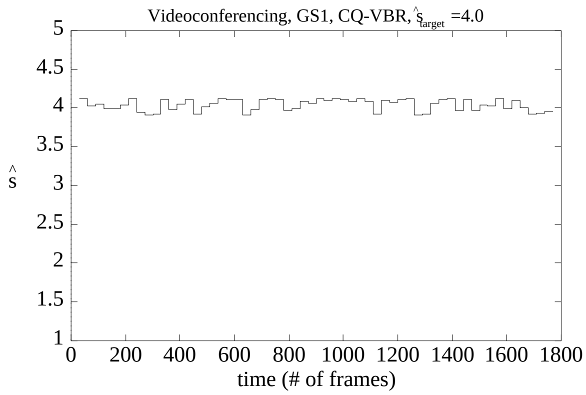


(a) $\hat{s}$ vs. time

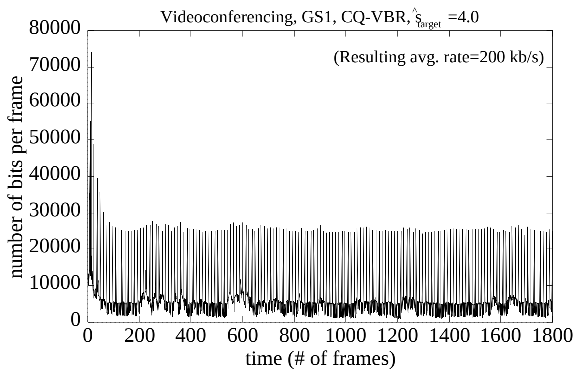


(b) Frame size versus time

Figure 4: Quality and frame size versus time for CBR, Star Trek VI, V=345 kb/s.

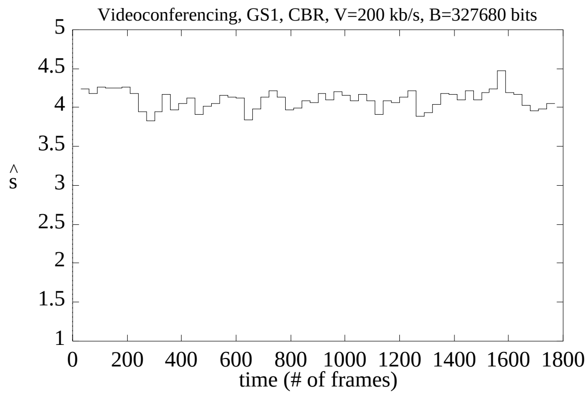


(a) $\hat{s}$ vs. time

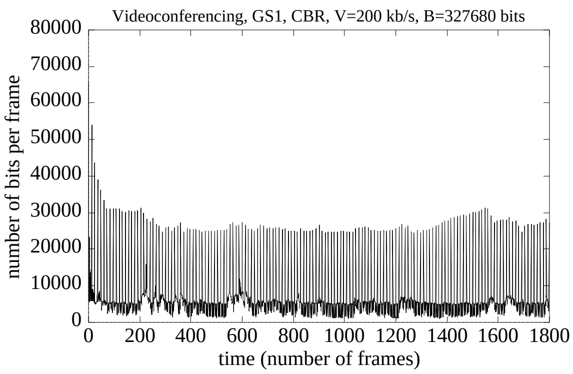


(b) Frame size vs. time

Figure 5: Quality and frame size versus time for CQ-VBR, Videoconferencing, $\hat{s}_{target}$=4.0.



(a) $\hat{s}$ vs. time



(b) Frame size vs. time

Figure 6: Quality and frame size versus time for CBR, Videoconferencing, V=200 kb/s.