Multimedia Systems and Applications II

Andrew G. Tescher
Bhaskaran Vasudev
V. Michael Bove, Jr.
Barbara Derryberry
Chairs/Editors

20–22 September 1999
Boston, Massachusetts

SPIE

P

PROCEEDINGS
SERIES

Volume 3845

Simulation of Graded Video Impairment by Weighted Summation: Validation of the Methodology

John M. Libert, Charles P. Fenimore, and Peter Roitman

National Institute of Standards and Technology[*], 100 Bureau Drive, Stop 8114
Gaithersburg, MD 20899-8114

ABSTRACT

The investigation examines two methodologies by which to control the impairment level of digital video test materials. Such continuous fine-tuning of video impairments is required for psychophysical measurements of human visual sensitivity to picture impairments induced by MPEG-2[1] compression. Because the visual sensitivity data will be used to calibrate objective and subjective video quality models and scales, the stimuli must contain realistic representations of actual encoder-induced video impairments. That is, both the visual and objective spatio-temporal response to the stimuli must be similar to the response to impairments induced directly by an encoder. The first method builds a regression model of the Peak Signal-To- Noise Ratio (PSNR) of the output sequence as a function of the bit rate specification used to encode a given video clip. The experiments find that for any source sequence, a polynomial function can be defined by which to predict the encoder bit rate that will yield a sequence having any targeted PSNR level. In a second method, MPEG-2-processed sequences are linearly combined with their unprocessed video sources. Linear regression is used to relate PSNR to the weighting factors used in combining the source and processed sequences. Then the "synthetically" adjusted impairments are compared to those created via an encoder. Visual comparison is made between corresponding I-, B-, and P-frames of the synthetically generated sequences and those processed by the codec. Also, PSNR comparisons are made between various combinations of source sequence, the MPEG-2 sequence used for mixing, the mixed sequence, and the codec-processed sequence. Both methods are found to support precision adjustment of impairment level adequate for visual threshold measurement. The authors caution that some realism may be lost when using the weighted summation method with highly compression-impaired video.

Keywords: video quality, video impairment, MPEG-2, video compression, PSNR, simulated impairment

1. INTRODUCTION

Because bandwidth is likely to remain at a premium, it also is likely that digital video will continue to be compressed as much as possible, limited only by the tolerance of the human viewer for a degraded picture. Accordingly, the digital video industry recognizes a need for objective quality metrics which have been calibrated to human subjective quality assessments. The need to support objective computational methods with human visual data has spurred several major research efforts, including those described in [2] and [3]. While both of these projects support improvement of video quality computational models, they address the quality issue at quite different levels of abstraction.

The study organized and executed by the Video Quality Experts Group (VQEG), described in [2] uses an approach of the television industry for subjective picture quality assessment. The methods, detailed in [4], generally involve assigning values of a numerical category scale, discrete or continuous, to video sequences based on each viewer's personal opinion of its quality relative to either an explicit or implicit reference. Generally, training is provided in an effort to "calibrate" internal scales of the viewers. But the precise nature of the scale by which each viewer assigns ratings is not observable directly. Also, unknown are the relative importance each viewer gives to each spatio-temporal distortion. Moreover, such an ordering may vary among viewers, or even may vary for a single viewer over the duration of the testing period.

However, such procedures have the advantage that they are efficient. They integrate a number of disparate quality elements, spatial and temporal, explicit and implicit, into a single value. Furthermore, the quality rating procedures tend to use test material in which picture distortions and their context are identical to or close to that which the viewers will see on television screens or multimedia monitors. Such stimuli retain an element of realism that may be lacking in simpler, highly-controlled

254

Part of the SPIE Conference on Multimedia Systems and Applications II
Boston, Massachusetts • September 1999 SPIE Vol. 3845 • 0277-786X/99/$10.00

stimuli. But the trade-off in yielding control is that the investigator does not know either the image characteristics that elicit a particular rating, or the functional form of the viewer's internal scale. These features of conventional subjective quality measurement present some problems to the builder of objective quality models. But with careful design and execution of experiments, training of subjects, and proper use of statistical analysis such studies continue to contribute useful data to video quality measurement.

Approaching the quality issue with a more "bottom-up" approach is the ongoing study described in [3]. This effort, known as "Modelfest" [5] aims to develop a comprehensive database of human visual threshold responses to elemental visual stimuli. Such stimuli are systematically varied with respect to the spatial and temporal parameters understood to be involved in human vision. Such data, collected via controlled psychophysical methods, are expected to assist in the development of increasingly refined and robust computational models of human vision. Such models, in turn, can be directed to the objective measurement of video quality.

Though not itself the topic of the present investigation, another subjective quality measurement activity has been proposed [6] that addresses the human visual aspect of quality modeling at an intermediate level of abstraction. The proposed study will attempt to define a scale of video impairment in terms of multiple measurements of the just-noticeable-difference (JND) of compression-induced video impairments. Through a series of paired comparisons the investigators will attempt to define a quantitative function that describes human visual sensitivity to video impairments at varying amplitudes. In this effort, quality assessment would involve visual perception of actual video impairments in their normal context rather than more atomic visual stimuli used in most vision experiments. Yet the experimenters will employ quantitative psychophysical techniques and exert greater control over the stimuli than that characteristic of subjective testing of the sort used in the VQEG study.

The present investigation is conducted to test candidate methodologies by which to prepare video test materials for the proposed JND study. In general, psychophysical measurement of a sensory threshold requires that the stimulus characteristic under study can be varied continuously by the experimenter and that the attribute under study can be gauged on some interval scale [7]. In the planned JND experiment, it will be necessary to produce video compression artifacts at various levels. Several methods for doing this were considered by the present investigators.

One approach is to model the spatio-temporal frequency characteristics of various impairments sufficiently so that synthetic impairments can be inserted into reference image sequences. Such a method is described in [8]. This method is interesting in the degree of control it offers with respect to the amplitude, distribution, and mix of various impairments. However, unless the spatio-temporal models of the impairments is very good, the necessary realism, both subjective and objective, would not support the JND study.

A second approach is suggested by the notion that video impairments are related directly to the parameters used in controlling the MPEG-2 encoding process. Variation of parameters, such as the target bit rate and group of pictures (GOP) can result in greater or lesser video impairment. However, the precise effect of these parameters depends on the spatio-temporal characteristics of the video sequence. Thus, while it is difficult to make general predictions as to the degree of impairment that might result from a given target bit rate, it should be possible to make such predictions for any particular video sequence. One objective of the present study is to determine the degree to which the encoder can be directed to output video having a specified degree of impairment.

A third approach, also to be examined here, is the degree to which video sequences can be generated at specified impairment levels by linear combination of an MPEG-2 processed (impaired) sequence with its unprocessed source, a weighted summation of the sequences. This technique has been used in various forms by a number of investigators [9, 10, 11]. On the surface, it appears obvious that such a linear combination should produce a sequence bearing impairments at a magnitude somewhere between the two end-members. However, a question remains as to whether sequences generated by combination of a highly compressed sequence and its source are a realistic simulation of the encoder output at the higher bit rate. Moreover, to be most useful, the weighted summation scheme should permit combination of scenes without computationally intensive registration of the sequences. The present investigation examines the effectiveness of the weighted summation method as a simulation of MPEG-2 encoded video.

2. PROCEDURE

1. Test Material

Three video clips were selected for the experiments. A sequence of sixty frames was taken from each of the sequences referred to as Duck, Water, and Mobile and Calendar. A single field of each sequence is shown in Fig. 1, 2, and 3, respectively.

Figure 1. Single field of DUCK sequence.

Figure 2. Single field of WATER sequence.

Figure 3. Single field of Mobile & Calendar sequence.

The three sequences were shown in an earlier study [12] to vary in their criticality, or the ease with which they can be encoded, where higher criticality indicates more difficulty encoding and greater impairment at a given bit rate. The criticality, C, of the three sequences can be ordered as $C_{mobile \& Calendar} \gg C_{Water} > C_{Duck}$. Each of the 60-frame clips was processed via the MPEG-2 [1] software codec Test Model 5 (TM-5) [13]. The encoding was done at main profile and main level with GOP=15, M=3, and at target bit rates of 15 Mb/s, 12 Mb/s, 10 Mb/s, 8Mb/s , 6 Mb/s, 4 Mb/s, and 2 Mb/s. The TM-5 decoder was applied to expand the compressed bit streams back into their original Rec. 601 [**Error! Reference source not found.**] component video format, Y,Cb,Cr.

2. Measurement of Relative Distortion

Any attempt to "tune" the level of a stimulus attribute requires some continuous measure by which to quantify the attribute[7]. In the present instance, we elected to use the average Peak Signal-to-Noise Ratio (PSNR) of the video sequence, where the PSNR for each frame is computed according to the expression:

$$PSNR = 20\log_{10}\left[\sum_{a=1}^{3} w_a\left(\frac{b_a^{2}}{m_a}\right)\right],$$ (1)

where b = peak value for channel (235 for luminance, Y'[1], 240 for chrominance channels, Cb and Cr)
       m = mean squared difference between processed and source channel
       a = 1...3, corresponding to component video channels, Y', Cb, Cr
       w = weight, 0.5 for Y', and 0.25 for each chrominance channel

PSNR is comparatively easy to compute and makes no assumptions as to the structure of video impairments or their perceptibility. It is thus appropriate for control of impairment level in the planned visual threshold experiment. An alternative might be to use a more sophisticated computational vision model, but this might later compromise the use of the threshold data for calibration of the vision model.

3. Method 1: Fit-Directed Encoding

Linear models up to a 4th order polynomial were fit to the average PSNR vs. target bit rate for each of the three video sequences. Figure 4 exhibits polynomial fits of orders 1 to 4 for the Duck sequence. The dotted curves mark 95 % percent
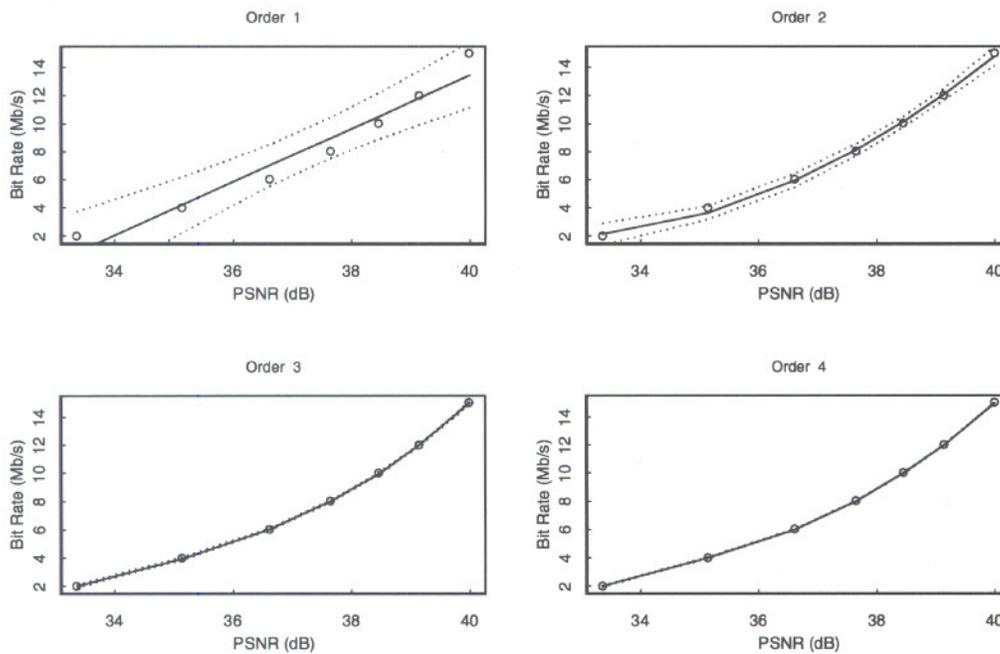

Figure 4. Polynomial fits of bit rate vs. PSNR for Duck sequences.

---

[1] Please note that in this paper, we adopt the convention of Poynton [15] in referring to non-linear encoded video luma as Y' and the luminance, i.e., the luminous flux per unit area, as Y.

confidence bounds for each fit. Plots obtained for the Water and Mobile & Calendar sequences were so similar that they are not shown here.

For the Duck sequence, Table 1 summarizes the standard errors of the predicted bit rate values plotted in Fig. 4. The coefficients of the fitted 4th order polynomial are also shown with measures of their standard error and the results of a t-test of their statistical significance.

Table 1. Standard Error of Bit Rate Predictions from PSNR - Duck

Original Bit Rate	15.0000	12.0000	10.0000	8.0000	6.0000	4.0000	2.0000
SE Order 1	0.6854	0.5636	0.4853	0.4279	0.4336	0.5777	0.8526
SE Order 2	0.1997	0.1259	0.1130	0.1278	0.1432	0.1384	0.2287
SE Order 3	0.0332	0.0203	0.0217	0.0209	0.0215	0.0297	0.0351
SE Order 4	0.0115	0.0089	0.0077	0.0078	0.0092	0.0114	0.0117

Coefficients (4th order fit):	Value	Std. Error	t value	Pr(>ltl)
Intercept	8.1429	0.0044	1845.2161	0
Order 1	10.8959	0.0117	933.2187	0
Order 2	2.4293	0.0117	208.0640	0
Order 3	0.4817	0.0117	41.2535	0.0006
Order 4	-0.0589	0.0117	-5.0471	0.0371

Table 1 indicates that on the scale of the PSNR, the prediction error is quite small, particularly with the 3rd and 4th order polynomials. A t-test indicates that even the 4th order term is significant, but its magnitude is small enough relative to that of the lower order terms to question its practical significance at the 5 % level.

Similar results are observed with the Water and Mobile & Calendar sequences summarized in Tables 2 and 3, respectively. As with the Duck sequence, the standard error of bit rate predictions to yield targeted PSNR are extremely small.

Table 2. Standard Error of Bit Rate Predictions from PSNR - Water

Original Bit Rate	15.0000	12.0000	10.0000	8.0000	6.0000	4.0000	2.0000
SE Order 1	0.5655	0.4483	0.3819	0.3381	0.3496	0.4663	0.6588
SE Order 2	0.1300	0.0778	0.0729	0.0819	0.0878	0.0841	0.1408
SE Order 3	0.0964	0.0605	0.0627	0.0577	0.0631	0.0807	0.0995
SE Order 4	0.0073	0.0059	0.0047	0.0050	0.0056	0.0071	0.0074

Coefficients (4th order fit):	Value	Std. Error	t value	Pr(>ltl)
Intercept	8.1429	0.0028	2918.3186	0
Order 1	10.9987	0.0074	1489.8650	0
Order 2	1.9477	0.0074	263.8390	0
Order 3	0.2496	0.0074	33.8115	0.0009
Order 4	-0.1742	0.0074	-23.6002	0.0018

Table 3. Standard Error of Bit Rate Predictions from PSNR - Mobile & Calendar

Original Bit Rate	15.0000	12.0000	10.0000	8.0000	6.0000	4.0000	2.0000
SE Order 1	0.8085	0.6708	0.5827	0.5167	0.5166	0.6556	1.0735
SE Order 2	0.1410	0.0904	0.0790	0.0876	0.1010	0.1030	0.1687
SE Order 3	0.0110	0.0067	0.0072	0.0070	0.0069	0.0102	0.0118
SE Order 4	0.0028	0.0021	0.0018	0.0019	0.0022	0.0028	0.0028

Coefficients (4th order fit):	Value	Std. Error	t value	Pr(>ltl)
Intercept	8.1429	0.0011	7639.7098	0
Order 1	10.7684	0.0028	3818.5999	0
Order 2	2.9624	0.0028	1050.5076	0
Order 3	0.3482	0.0028	123.4776	0.0001
Order 4	0.0200	0.0028	7.1057	0.0192

Thus, it appears that if one is able to generate a sample of encoded sequences over a range of bit rates of interest, it is quite feasible to define, for a given video sequence and encoder, a function that will enable precise control over the PSNR of additional output sequences.

4.   Method 2: Weighted Summation Method

In some cases direct encoding may not be an option. In the anticipated JND study described previously, for example, a relatively large volume of video test material is available with associated subjective quality ratings. Source sequences are available as well as a number of processed versions of the source created by various means including, but not restricted to, MPEG-2 encoding. The sampling of impairment levels in the available data covers a wide range, but is not sufficiently dense to support threshold measurements. Replicating the processing to add the required additional sequences is not possible. Also, the investigators wish to retain the option of relating the threshold measurements to the existing subjective quality ratings. Such joint analysis would not be compromised by additional encoding, e.g., by using the fit-directed method described above. Accordingly, a weighted summation method is considered here. In particular, the investigators are interested in evaluating the departures of the synthetically graded impairments from those induced directly by encoding.

For each of the three source sequences described previously, a series of mixed sequences was produced by linear combination of the source sequence with a corresponding sequence having undergone compression. Thus, the frames of an uncompressed source sequence, having luma and chroma components (Y', C_b, C_r), were combined with corresponding frames of a selected processed sequence (MPEG-2) according to the linear equation

$$(Y', C_b, C_r)_{sim} = (1-t) \cdot (Y', C_b, C_r)_{source} + t \cdot (\overline{Y}', C_{b_0}, C_{r_0})_{MPEG-2} . \qquad (2)$$

The non-dimensional contrast parameter, t, ranges on the interval $[0,1]$. At $t = 1$, one would have the uncompressed source sequence, while for $t=0$, one would have the MPEG-2 impaired sequence. The value of t was varied in increments of 0.1 to generate nine impaired sequences from each source combined with an MPEG-2 processed mate. This mixing procedure was performed using both the 2 Mb/s and the 4 Mb/s compressed versions of each source sequence.
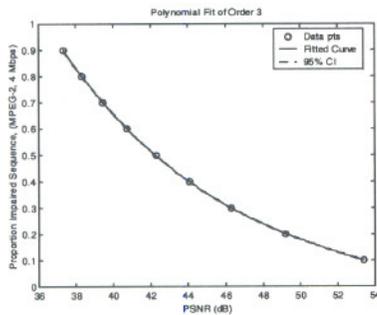

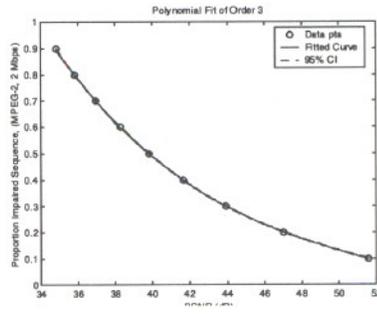**Figure 5. Duck: Cubic fit to mixing factor (*t*) of 4 Mb/s MPEG-2 sequence vs. PSNR.**

Figure 6. Duck: Cubic fit to mixing factor (*t*) of 2 Mb/s MPEG-2 sequence vs. PSNR.
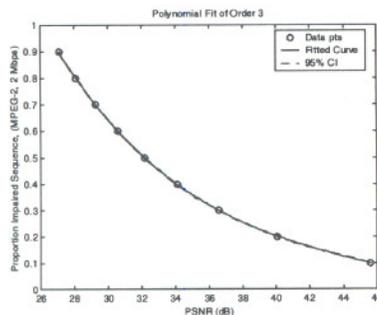

Figure 7. M&C: Cubic fit to mixing factor (*t*) of 2 Mb/s MPEG-2 sequence vs. PSNR.
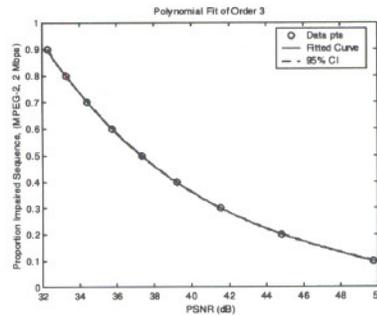

Figure 8. Water: Cubic fit to mixing factor (*t*) of 2 Mb/s MPEG-2 sequence vs. PSNR.

259

The average PSNR was computed for each series of mixed sequences according to Eq. 1. For each series of the synthetically impaired sequences, a polynomial was fit to the computed PSNR values and the corresponding t values. Figures 5 and 6 exhibit the cubic fits for the Duck sequence, with compressed components 2 Mb/s and 4Mb/s, respectively. Figures 7 and 8 are cubic fits to the Mobile & Calendar and Water weighted summation sequences using only the 2 Mb/s MPEG-2 material. For each of the three source video sequences, the mapping functions enable control over the mixing so as to yield any degree of impairment, as gauged by the PSNR. The parameters of the fitting functions are summarized in Table 4. As may be observed, the 3rd order term is quite small for each of the fits, but was found to reduce the width of 95 % confidence bounds about the fitted lines in each case.

Table 4. Polynomial Fits

Sequence	Intercept	x	x^2	x^3
Duck 4Mbps	16.355072	-0.858686	0.015467	-0.000095
Duck 2 Mbps	14.173794	-0.774368	0.014545	-0.000093
M & C 2 Mbps	8.841076	-0.551443	0.011941	-0.000089
Water 2 Mbps	12.211709	-0.696499	0.013697	-0.000092

Insofar as MPEG-2 is designed to adjust its encoding mechanisms to target a bit rate selected by the user, it is not surprising that a pixel level metric such as PSNR should be so tightly coupled to the bit rates of sequences linearly added to one another. The similarity of the functional forms depicted in Fig. 5 – 8 and parametrized in Table 4 reinforce this notion, but also indicate that the precise relationship between bit rate and PSNR is dependent upon the particular video sequence.

Given that the weighted summation method can be used to generate video sequences having any targeted level of distortion, the investigators performed a qualitative examination to verify that synthetically impaired sequences were visually similar to MPEG-2 sequences matched according to PSNR. Selected I-frames, B-frames, and P-frames were examined to verify the weighted summation frames were similar in appearance to those created by encoding. The PSNR contrasts between sequences provide a quantitative measure of similarity.

3. RESULTS AND DISCUSSION

Fig. 9, 10, and 11 exhibit enlarged portions of I-, P-, and B-frames sampled from the Duck sequences. In each figure, the source image (a) has been linearly combined with the MPEG-2 encoded frame (c) to produce the weighted summation frame (d). As explained above, the mixing parameters were selected so that the mixed frame (d) would be similar to (b) both visually and with respect to PSNR. Figure 12 displays a similar arrangement of corresponding I-frames of the Mobile & Calendar sequence.

For the Duck sequence, the mixing of the source was done with the 2 Mb/s compressed sequence. The Mobile & Calendar source sequence was mixed with its 4 Mb/s counterpart. In all cases, the mixed sequences are acceptably similar to the directly encoded versions, though some small differences may be seen in the Mobile & Calendar images.

Quantitative comparisons of the images sequences shown were made among the sequences using PSNR. The results are summarized in Table 5. While PSNR might not fully capture the potential structural differences between the images, the investigators thought it a reasonable comparison to make for their purposes. It does confirm, for example, that the global error relationships between the source sequence and the other three sequences are in the proper order. As expected, the error

Table 5. PSNR Comparisons of Test Sequences

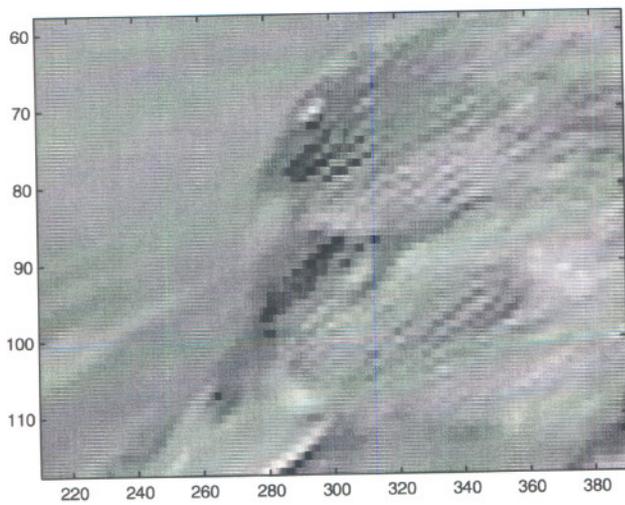| Interpolated Clips	a x b	a x c	a x d	b x d
Duck (2 Mbps)	33.4843	28.5762	39.7288	37.5358
Duck (4 Mbps)	34.4634	32.2573	36.6335	45.7107
Mobile (4 Mbps)	26.0395	24.7771	26.6539	30.4058

a = Source
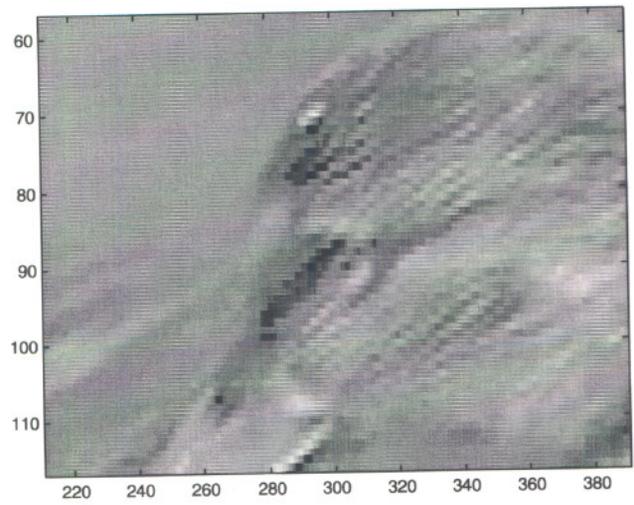b = MPEG-2 Sequence at 6 Mb/s
c = MPEG-2 at 2 Mb/s
d = Source + MPEG-2 such that PSNR == 6 Mb/s
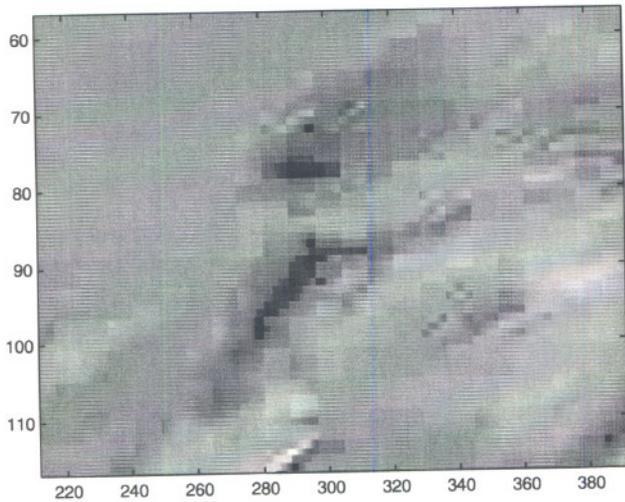Duck mixed with both 2- and 4- Mb/s sequences
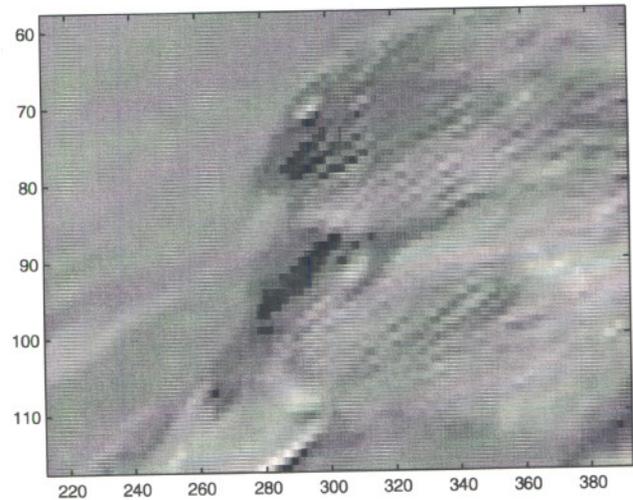PSNRs are averaged over 1 GOP

(a) Source

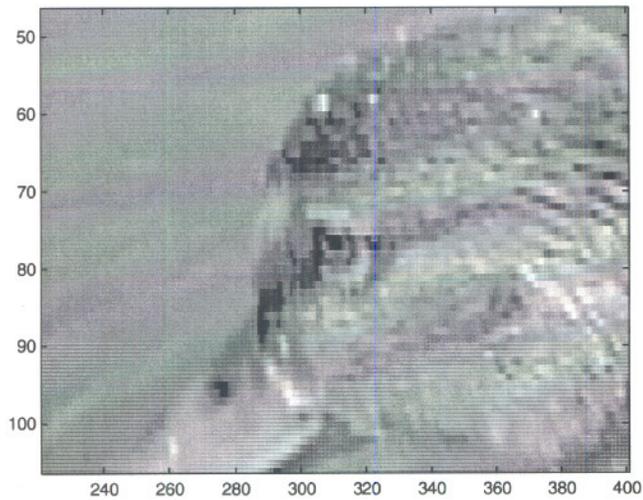(b) MPEG-2 Compressed at 6 Mb/s.
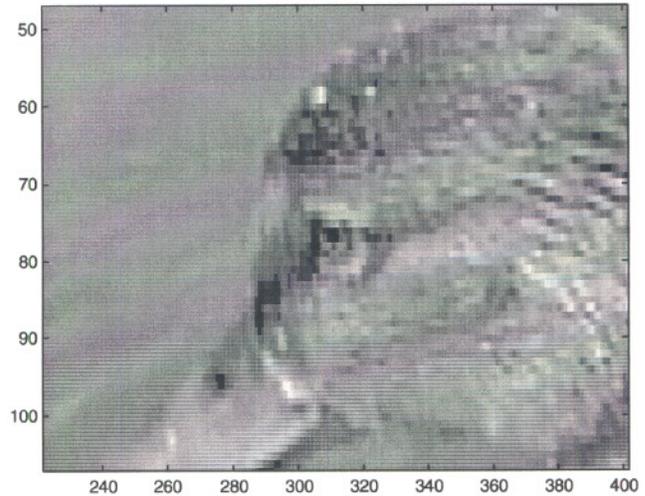

(c) MPEG-2 Compressed at 2 Mb/s.

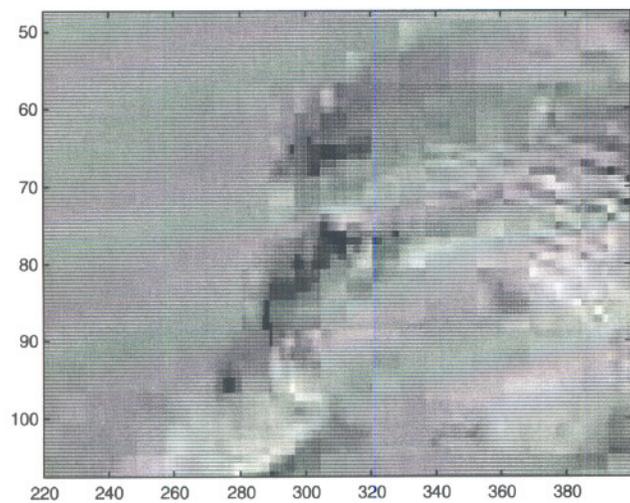(d) Weighted sum of source and 2 Mb/s sequence
targeted PSNR of 6 Mb/s.

Figure 9. Enlarged portions of Source and processed I-frames. Appearance of weighted summation frame (d) generated from (a) and (c) to targeted PSNR of 6 Mb/s is similar in appearance to MPEG-2 frame encoded at 6 Mb/s. (Pixel coordinates are shown along image margins.)
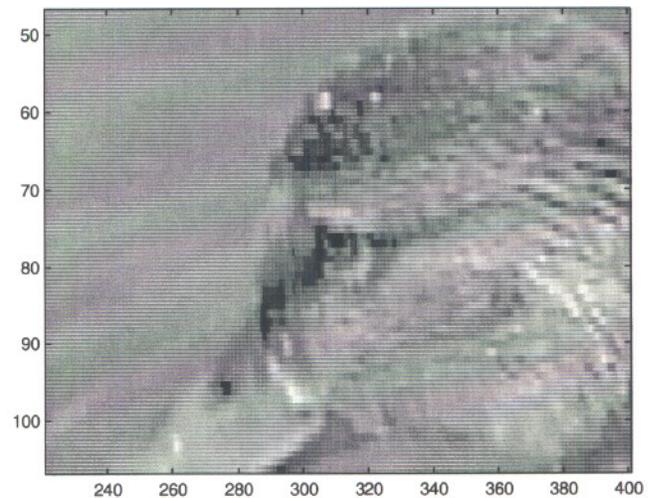
(a) Source.

(b) MPEG-2 sequence compressed at 6 Mb/s.

(c) MPEG-2 compressed at 2 Mb/s.

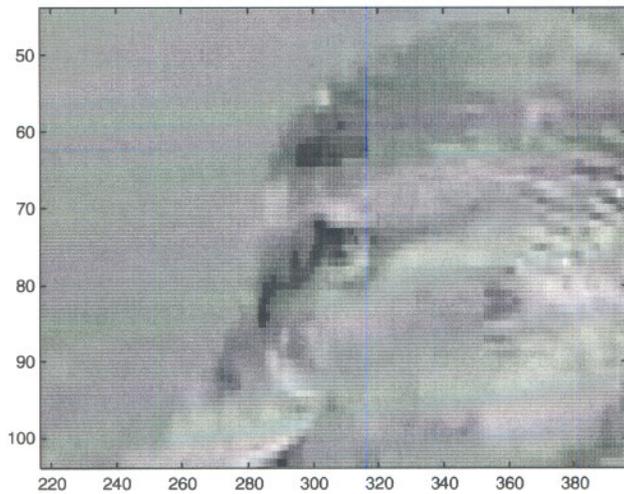(d) Weighted sum of source and MPEG-2 at 2 Mb/s at targeted PSNR of 6 Mb/s.

Figure 10. Enlarged portions of Source and processed P-frames. Appearance of weighted summation frame (d) generated from (a) and (c) to targeted PSNR of 6 Mb/s is similar in appearance to MPEG-2 frame encoded at 6 Mb/s. (Pixel coordinates are shown along image margins.)
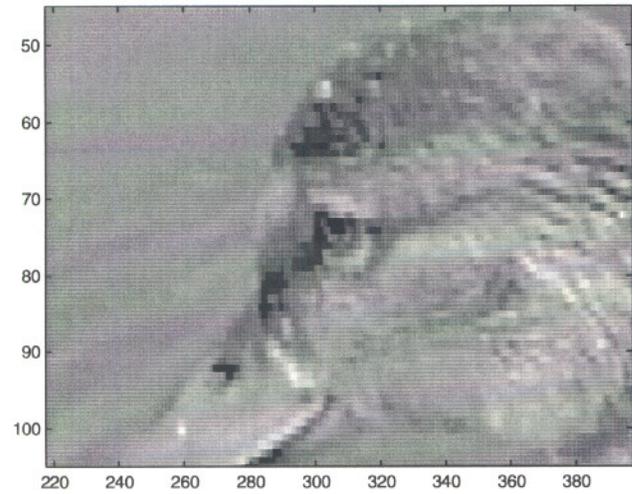
(a) Source.

(b) MPEG-2 sequence compressed at 6 Mb/s.

(c) MPEG-2 compressed at 2 Mb/s.

(d) Weighted sum of source and MPEG-2 at 2 Mb/s at targeted PSNR of 6 Mb/s.

Figure 11.   Enlarged portions of Source and processed B-frames. Appearance of weighted summation frame (d) generated from (a) and (c) to targeted PSNR of 6 Mb/s is similar in appearance to MPEG-2 frame encoded at 6 Mb/s. (Pixel coordinates are shown along image margins.)
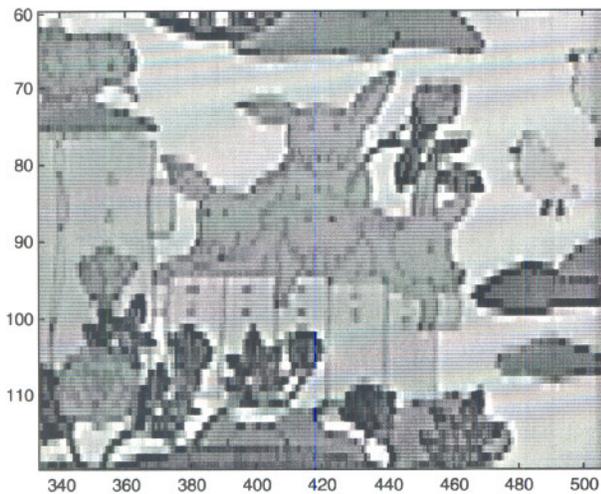
(a) Source I-Frame

(b) I-frame encoded at 6 Mb/s

(c) I-frame encoded at 4 Mb/s

(d) Weighted sum of source and 4 Mb/s with PSNR of 6 Mb/s

Figure 12. Enlarged portions of Source and processed I-frames. Appearance of weighted summation frame (d) generated from (a) and (c) to targeted PSNR of 6 Mb/s is similar in appearance to MPEG-2 frame encoded at 6 Mb/s. (Pixel coordinates are shown along image margins.)

is greatest (i.e., lower PSNR) in the comparison of the source with the lower bit rate MPEG-2 sequence (a x c). It also is encouraging that the lowest error is found in the comparison of the 6 Mb/s MPEG-2 sequence and the mixed sequence produced as a simulation of 6 Mb/s encoding (b x d). The unexpected result is that the error between the source and encoded sequence (a x b) is greater than that between source and simulation (a x d). In this regard, it is interesting as well that the effect is less pronounced when the 4 Mb/s sequence is used in the weighted summation.

The latter result implies that when the source is linearly combined with a severely impaired sequence, the result more closely resembles the source than the encoded sequence it was intended to simulate. We might assume, for example, that the impairment pattern itself remains relatively constant except for an increase in relative amplitude as bit rate is reduced. We would expect, in this case, that the simulation of a 6 Mb/s sequence should be a close approximation to the encoded 6 Mb/s

sequence. Moreover, we would expect the simulated sequence to bear no greater similarity to the source sequence that that exhibited by the encoded sequence. If, on the other hand, the impairment pattern were to change markedly at very low bit rates, the mixing based on PSNR would fail to behave linearly. The disruption of the impairment pattern would inflate the PSNR so that a greater weight would be assigned to the source sequence than that which would be predicted by the linear relationship. Hence, the simulation of the impaired sequence would be more like the source and less like the MPEG-2 impaired sequence. This is consistent with the results for the Duck sequence observed in Table 5.

4. CONCLUSION

We find that either of two methods may be used to generate variably impaired test material suitable for threshold studies of impairment visibility. For any source sequence, simple curve fitting results in a polynomial function by which to set the encoder bit rate parameter to produce an output sequence having a targeted PSNR level. Moreover, such control over the encoding error rate is highly precise, provided that the function derivation is restricted to each sequence of the source video.

In the case that fit-directed encoding is not possible, linear combination of an impaired sequence with its source can be used to generate video at intermediate levels of impairment. The linearity of this process appears to break down to some extent, however, when the weighted summation with the source is done with a low bit rate (< 4 Mb/s) MPEG-2 sequence. In this case, some realism may be lost.

5. REFERENCES

1. ISO/IEC DIS 13818-2. International Standard (MPEG-2), Moving Picture Experts Group, – Generic coding of moving pictures and associated audio information, Part 2 Video.
2. P. Corriveau and A. Webster. "The Video Quality Experts Group evaluates objective methods of video image quality assessment." *Proceedings of the 140th Annual SMPTE Technical Conference and Exhibit,* Pasadena, CA, October 28-31, 1998. pp. 509-516.
3. Thom Carney, Stanley A. Klein, Christopher W. Tyler, Amnon D. Silverstein, Brent Beutter, Dennis Levi, Andrew B. Watson, Adam J. Reeves, Anthony M. Norcia, Chien-Chung Chen, Walter Makous, and Miguel P. Eckstein. "The development of an image/threshold database for designing and testing human vision models." *Human Vision and Electronic Imaging IV*, SPIE Vol. 3644, 25-29 January 1999, San Jose, California, 542-551.
4. Recommendation ITU-R BT.500-9, *Methodology for the subjective assessment of the quality of television pictures,* ITU-R 1974-1998.
5. Thom Carney, *et. al.* Modelfest. http://www.neurometrics.com/projects/Modelfest/IndexModelfest.htm
6. J. M. Libert, A. B. Watson, A.M. Rohaly, and L. Stanger. "Toward development of a JND-based scale of digital video impairment: contribution of the IEEE Broadcast Technology Society Subcommittee on Video Compression Measurements." [in preparation for *Human Vision and Electronic Imaging V*, SPIE Photonics 2000, January 2000, San Jose, California.]
7. Personal Communication, Geoffrey Iverson, Inst. for Mathematical Behavioral Sciences, University of California at Irvine.
8. Recommendation ITU-T P.930, *Principals of a reference impairment system for video,* ITU-T 1996.
9. C. Fenimore, B. Field, and C. Van DeGrift, "Test patterns and quality metrics for digital video compression." *Human Vision and Electronic Imaging II,* SPIE Vol. 3016, 10-13 February, San Jose, CA, 269-276, 1997.
10. Ann Marie Rohaly, Albert J. Ahumada, and Andrew B. Watson. "Object detection in natural backgrounds predicted by discrimination performance and models." *Vision Research,* Vol. 37, No. 23, 3225-3235, 1997.
11. J. M. Libert and C. Fenimore. "Visibility thresholds for compression-induced image blocking: measurements and methods." *Human Vision and Electronic Imaging IV*, SPIE Vol. 3644, 25-29 January, San Jose, California, 197-206, 1999.
12. C. Fenimore, J. M. Libert, and S. Wolf. "Perceptual effects of noise in digital video compression." *Proceedings of the 140th Annual SMPTE Technical Conference and Exhibit,* Pasadena, CA, October 28-31, 1998. pp. 472-484.
13. ISO/IEC JTC1/SC29/WG11/N0400, *Test Model 5 (draft),* MPEG93/457, 1993 (Software is available by FTP on the World Wide Web at ftp://crs4.it/mpeg/programs/).
14. Recommendation ITU-R BT.601-5, *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios.*
15. C. A. Poynton, *A Technical Introduction to Digital Video,* John Wiley and Sons, New York, NY, p. 24, 1996.
16. J. M. Neter, M. H. Kunter, C. J. Nachtsheim, and W. Wasserman. 1996. *Applied Linear Statistical Models,* 4th Edition, WCB/McGraw-Hill, Boston, MA.