Progress Report of Task Force on Selecting Test Material and Test Labs for Defining a Unit of Measure and a Means of Calibration for Video Impairment[1]

Author: John M. Libert
21 July, 1999

1 Scope and Objectives

1.1 This project complements that of the *Task Force on Defining A Unit of Measure & a Means of Calibration for Video Impairment*[2] (UCTF), chaired by Leon Stanger, and aims to resolve issues related to the implementation of the plan proposed by that task force. Insofar as the subjective test protocol is somewhat entwined with the test material selection and manipulation, the scope of the Materials & Test task force (MTTF) was broadened to address details of test protocol.

1.2 Objective 1: Identify a methodology for the production of video test materials exhibiting only a single type of impairment at graded levels of expression with respect to an unprocessed reference sequence;

1.3 Objective 2: Identify subjective test laboratories able and willing to execute visual threshold testing in support of the calibration effort;

1.4 Objective 3: Detail a subjective testing protocol for the determination of a series of just-noticeable-differences (JND) with respect to the visibility of selected video compression-induced impairments;

1.5 Objective 4: Using the JND measurements as a calibration measure, compile a set of video test materials exhibiting each of the selected impairments at levels corresponding to 1 – n JND levels beyond the reference sequence. Make these sequences available for distribution on suitable media to the video industry to calibrate either visual or objective measures of impairment.

2 Technical Progress

2.1 <u>Test Material Production</u>

2.1.1 *Selection of test materials* – As the ultimate objective of our effort is to be the distribution of a set of visually calibrated video sequences, it was decided early in the project that any material selected would have to be free of copyright restrictions on its distribution. Unfortunately, much of the video material that has been used in video quality measurement and in the development and testing of video codecs has been made

---

[1] In the remainder of this document, this task force will be referred to as the *Materials & Test* task force (MTTF.)

[2] This task force will be referred to as the *Unit & Calibration* task force (UCTF.)

available by its owners for restricted use in research, but not for broad distribution without granting of permission or payment of royalties.

The author circulated an email request to the G-2.1.6 membership for contributions of either test materials that could be entered into the public domain or volunteers to shoot some appropriate video sequences. The request generated little or no response. Noting the status of a concurrent effort on the part of a SMPTE subcommittee to secure test materials, the author concluded that perhaps the best immediate action would be to shoot some material himself.

It should be noted that the majority of video quality measurement efforts, including that of SMPTE, have tended to select video materials that maximize the stress on video codecs and maintain a relatively high level of interest of the viewer and to be representative of actual television images. In the opinion of the author, such criteria have merit, but such materials may not suit the purposes of our proposed visual threshold measurements. "Interesting" materials are more likely to generate a complex mixture of video impairments that would complicate the interpretation of threshold measurements. Part of the aim of the pilot study being planned by G-2.1.6 is to begin to resolve the visual sensitivity to several of the most bothersome and common video defects.

2.1.2 *Blocking Impairments* – As suggested in the report of the UCTF, running water has been recognized to induce relatively high rates of blocking impairments when processed by MPEG-2 codecs. While recognized to be less desirable than video acquired with a studio quality digital camera, the author and a colleague took a Sony Betacam SP camera into the field to capture a few sequences of video. The component video signal was input to a video disk recorder for A/D conversion to a sequence of YCbCr (4:2:2) interlaced frames. Figure 1 exhibits a single frame from the sequence.

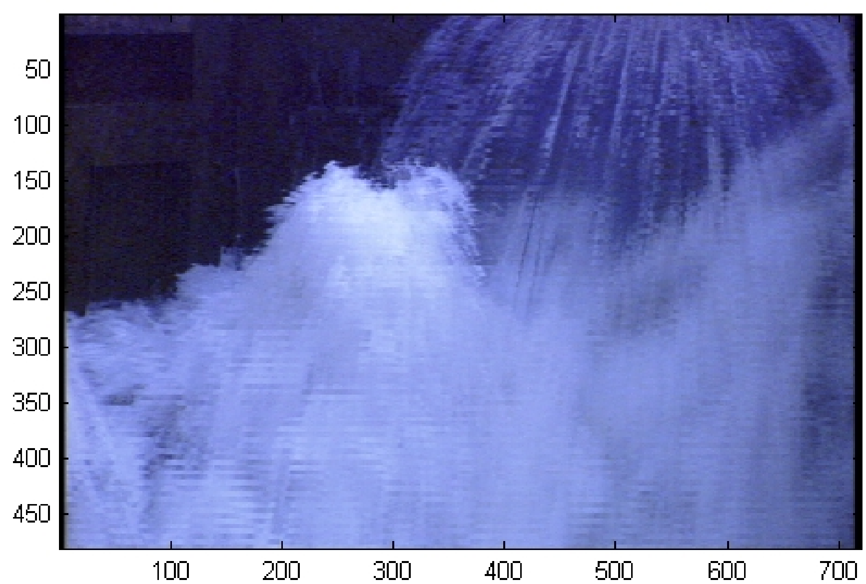

Figure 1 – Single frame of video sequence of water pouring through lock on the C&O canal at Great Falls, MD. (Displayed as a still frame, the interlacing error is quite apparent.)

Figures 2 and 3 show a single frame of the "spillway" sequence subjected to MPEG-2 encoding at 8- and 2 Mbps, respectively.

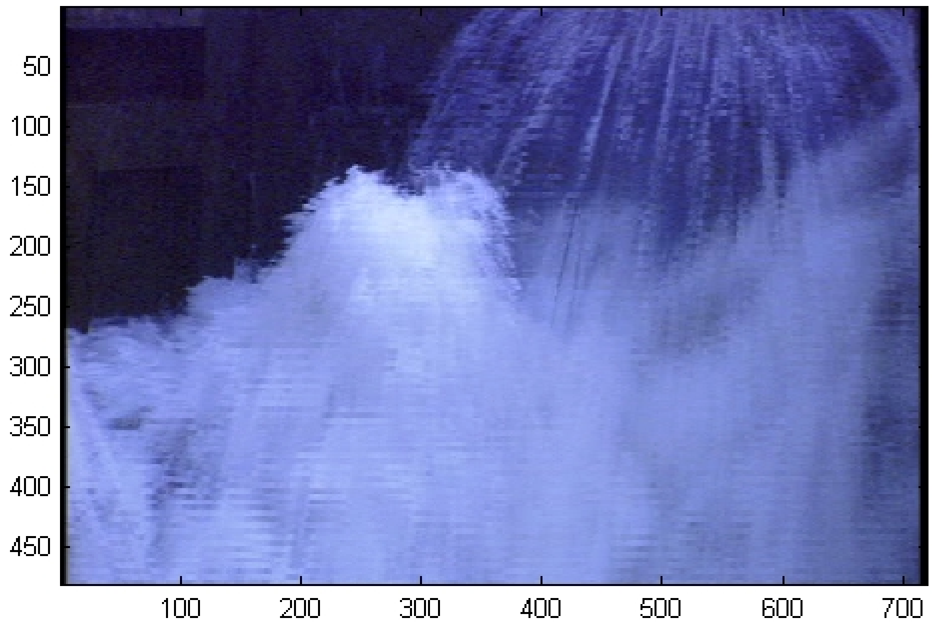

Figure 2 – I-frame of MPEG-2 encoded sequence shown in Fig. 1 after encoding at 8 Mbps.

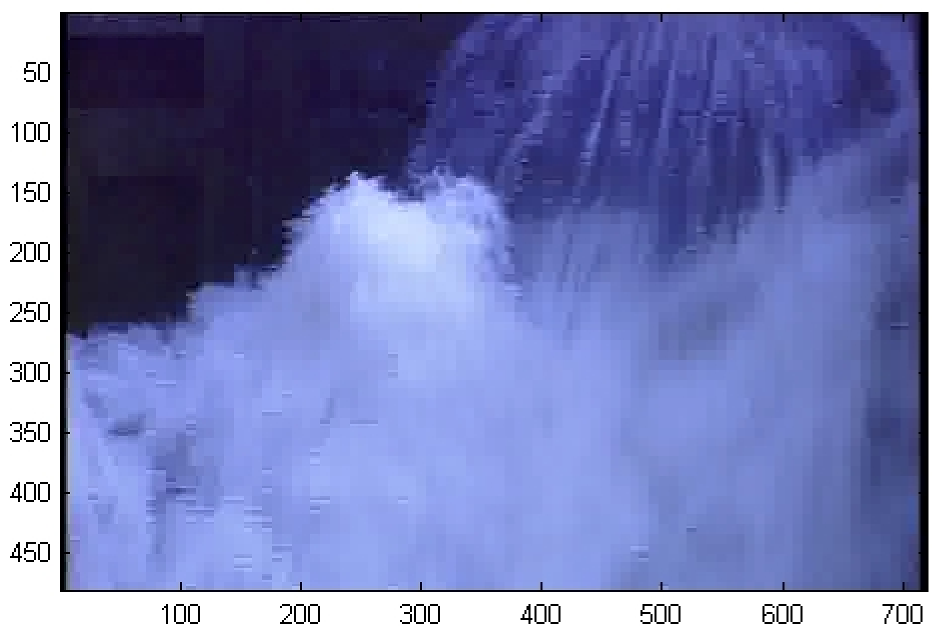

Figure 3 – I-frame of sequence depicted in Fig. 1 after MPEG-2 encoding at 2Mbps.

The camera work is decidedly short of an Academy Award, but the MPEG-2 encoding of the sequence does appear to induce predominantly the expected blocking impairments.

The enlarged sub-samples of Figures 4-6 exhibit more clearly the familiar pattern, or its absence.

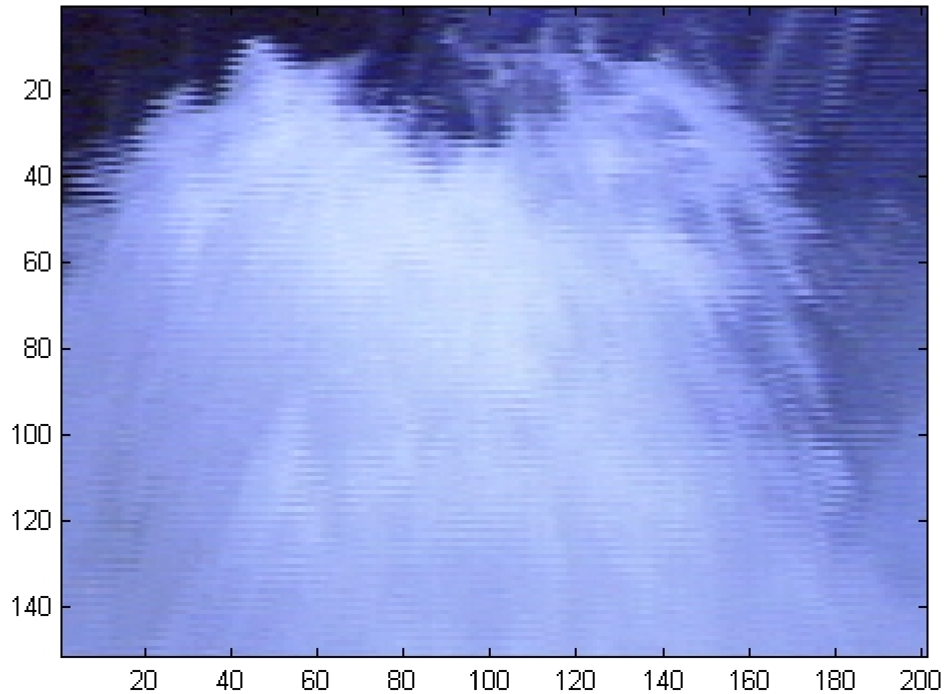

Figure 4 – Sub-sample of frame from spillway source sequence.

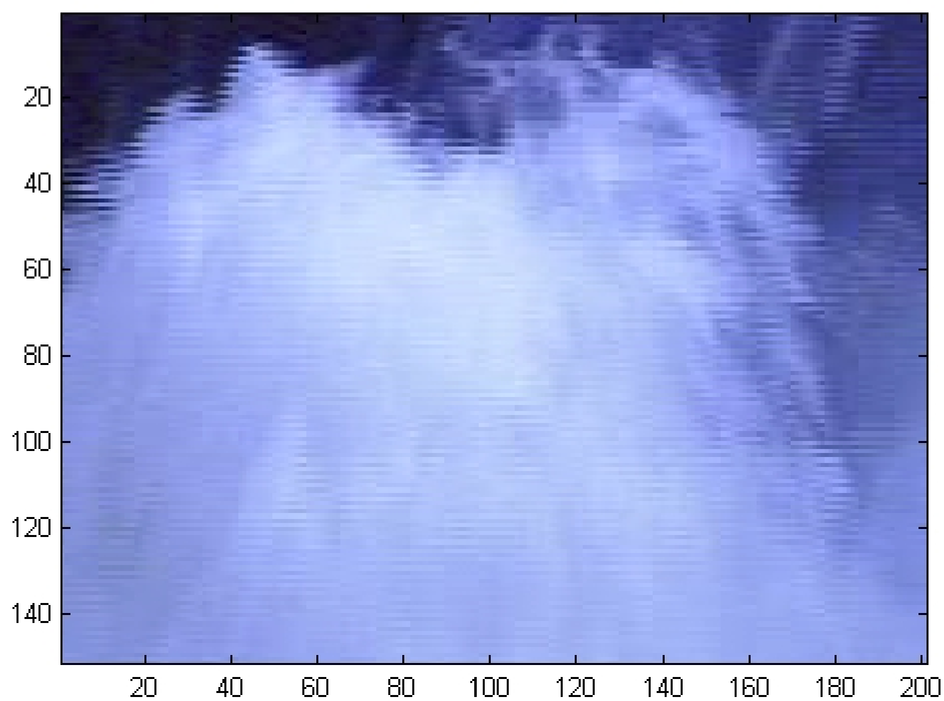

Figure 5 – Sub-sample of I-frame from sequence encoded at 8 Mbps.

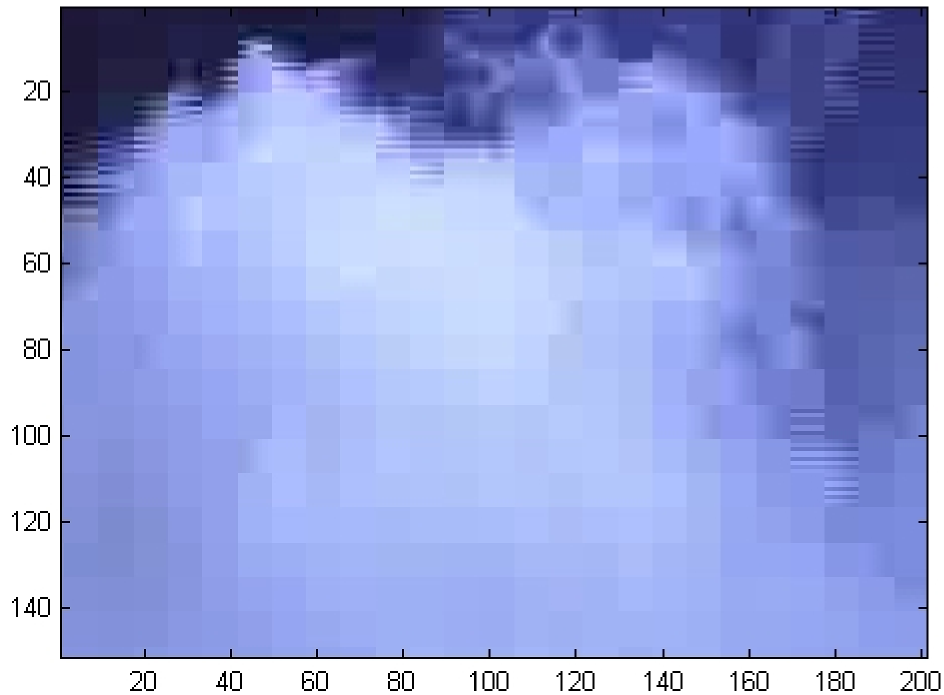

Figure 6 – Sub-sample of I-frame encoded at 2Mbps.

To form a basic working data set, the reference sequence was compressed using the public domain MPEG-2 codec "Test Model 5" at a sampling of target bit rates from 15 Mb/s to 2Mb/s. Peak signal to noise ratio (PSNR) was computed for each decoded sequence according to the expression:

$$PSNR = 20 \cdot \log_{10} \frac{b}{rms}$$

where b is the largest possible value of the signal, 235 for Y and 240 for Cb and Cr channels, and rms is the root mean squared difference between two images. In computation of the total rms error, weights of 0.5 for luminance and 0.25 for each of the color channels were applied. (The PSNR is given in decibel units (dB), which measure the ratio of the peak signal and the difference between two images. An increase of 20 dB corresponds to a ten-fold decrease in the rms difference between two images.)

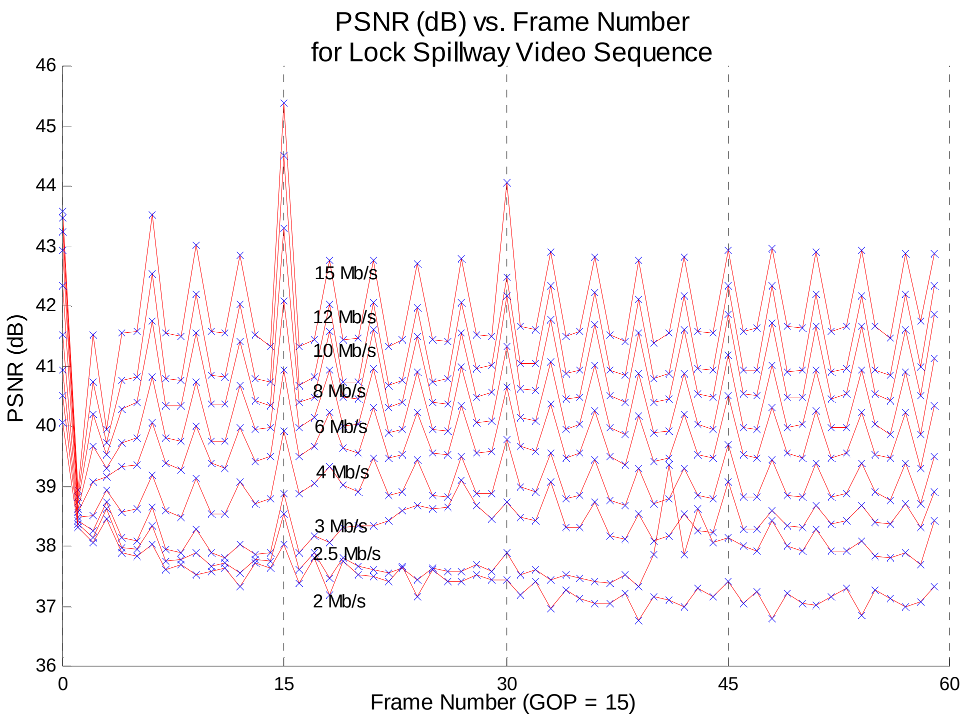

Figure 7 – PSNR for 60 frames of sequences processed by TM5. Target bit rates are indicated for each trace.

Figure 7 - plots peak signal to noise ratio for the 60 frames of each of the processed sequences. Target bit rates are indicated for each trace. The MPEG-2 processing was done using a Group of Pictures (GOP) of 15, with P-frames separated by two intervening B-frames. Thus, the GOP pattern is: I-B-B-P-B-B-P-B-B-P-B-B-P-B-B, beginning with frame 0, an I frame. The PSNR pattern probably should not be taken as stable until frame 15. The largest peaks in the PSNR coincide with the I-frames, e.g. at 15, 30, and 45, and the smaller peaks coincide with P-frames.

Evident in the plot is the parallelism of the PSNR pattern frame-to-frame down to a target bit rate of around 4 Mb/s, beyond which the pattern degrades, presumably as error processes beyond those due to quantization alone dominate.

Mean PSNR for frames 15 – 59 is plotted against target bit rate in Figure 8.

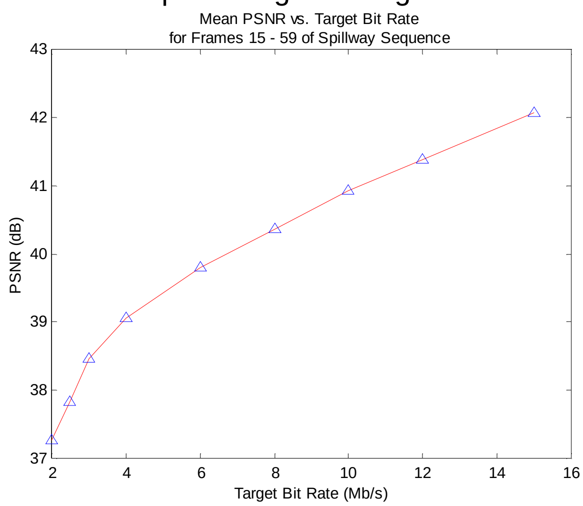

Figure 8 – Mean PSNR vs. target bit rate for frames 15 – 59 of spillway sequence.

2.1.3 *Contouring and Mosquito Noise Impairments* – Source sequences for these impairments may be produced via computer graphics construction or captured on video tape. C. Fenimore at NIST is currently developing a synthetic image sequence to induce mosquito noise (or edge business) when subjected to MPEG-2 encoding. This sequence may be a candidate for tests of sensitivity to this impairment.

2.1.4 *Production of Graded Impairments* – Given source images either designed or selected to favor the production of one type of compression artifact over others satisfies only one requirement of the experiments proposed by the UCTF. It is also essential that the artifacts under study can be varied finely in amplitude and selected under the direct control of the investigators. Psychophysical measurements of sensitivity to variation of some stimulus generally require that the stimulus under examination be continuously variable along some measurable scale under the control of the experimenter. Moreover, the product of the proposed study a set of video sequences exhibiting each target video impairment at levels representative of increments of JND. Several methods were examined relative to this requirement.

2.1.4.1 Artificial impairments – A method was described in ITU-T Recommendation P930, Principals of a Reference Impairment System for Video, describes procedures for synthesis of signal patterns designed to be similar in appearance to actual compression-induced artifacts that can be inserted into unprocessed video sequences to simulate impaired video for use in subjective quality measurements. A software implementation of P930 was developed by Bellcore and assigned the acronym VIRIS (Video Reference Impairment System.) Attempt was made to locate principals at Bellcore involved in the development of this model, but all had either left the company or moved into different positions at Bellcore and the VIRIS product was essentially dropped by the company. One Bellcore engineer involved with VIRIS indicated that the code might be recoverable, but that it would be difficult to make it operational without assistance from one of the authors and that such assistance was unlikely to be authorized without funding. The P930 document was found to detail the algorithms use in simulating the impairments, but implementation would require fairly significant programming effort. In addition to these obstacles, several G-2.1.6 members who had seen the demonstrations of VIRIS indicated that while some of the impairments were reasonable visual representations of actual artifacts, others were rather questionable. In the proposed experiment, it is not only important that the impairments be visually comparable to encoder-induced defects, but that they be objectively similar as well otherwise they would be of limited use in calibrating objective metrics. Hence, it is suggested here that the VIRIS method would not be adequate to support the threshold measurements.

2.1.4.2 **Weighted Summation (Mixing)-** Recognizing that impairments being evaluated both subjectively and objectively be as realistic as possible, some investigators have recognized that encoded and reference video sequences may be mixed via weighted summation (or averaging) to produce impairment sequences that vary as finely as desired between the two end-member sequences. In addition to the fine control afforded by this method, it can also be performed in real-time or nearly so, thus presenting the opportunity for adaptive adjustment of stimulus levels for more efficient convergence on visual threshold levels.

Of course, the principal advantage of this method is that the impairment patterns, having originated from actual compression encoding, tend to closely resemble the "natural" impairments visually and to a great extent objectively as well. Facilities were not available for planned quantitative subjective measurements of visual similarity of the impairments produced by mixing to those induced by encoding. However, during repeated viewing of alternating "real" and artificially- impaired sequences of several scenes, it was not possible to distinguish between them.

Some objective quantitative measures were applied to the mixed sequences in an effort to assess the degree of similarity to encoder-induced impairments. To the extent that the bit rate reduction is due mainly to quantization, one would expect the impairment pattern to change in amplitude, but not significantly in structure.

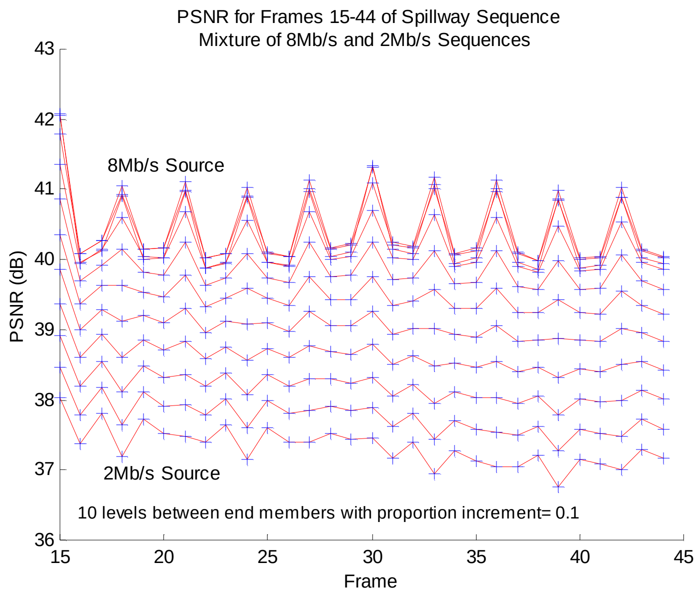

Figure 9 – PSNR for frames 15 – 44 of Spillway sequence constructed via weighted summation of 8- and 2-Mb/s sequences, proportion step = 0.1 between end-members.

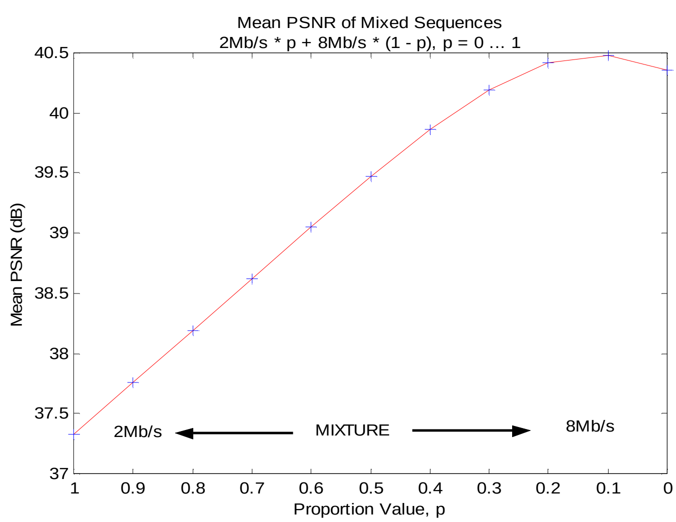

Figure 10 – Mean PSNR for end members (8 & 2 Mb/s) and 9 intermediate mixed sequences.

Figure 9 shows the pattern of PSNR values for two GOPs between frames 15 and 44 of Spillway. The proportion of 8Mb/s and 2Mb/s end members was varied linearly using an increment of 0.1. Though showing more gradual change at the high-bit rate end, the change in PSNR is approximately linear over much of the range. The pattern of the PSNR variation among I-, B-, and P-frames, however, shows some degree of parallelism as the basic structure of the traces remain similar, though damped from 8 to 2 Mb/s. Some peak reversals occur as well, however.

The mean PSNR values over the 30 frames are plotted in Figure 10 against the value of the proportion, *p*, of the 2 Mb/s sequence, the proportion of the 8 Mb/s sequence being, of course, *1 – p*. The linearity of much of the sequence is apparent, though the curvature at one end of the plot suggests that one

Fit-Directed Compression - While in general impairment level tends to increase with decrease in the target bit rate specified for encoding schemes, the precise response of an image sequence to a change in target bit rate is affected significantly by spatio-temporal characteristics of the source video sequence. Hence, it is difficult to predict with reasonable certainty what level of impairment might result from encoding at a particular bit rate.

However, if one is interested in the impairment behavior of a particular video sequence, it is possible to make reasonable predictions about the effects of bit rate changes for the particular sequence. Linear[3] or non-linear regression can be used to determine a function that can predict rather accurately the bit rate to be input to an encoder to produce an output sequence that will exhibit, for example, a desired PSNR value.

Regression of target bit rate (input to the TM5 codec) against mean PSNR of the output sequences was performed for Polynomials of the 3rd and 4th order were found to fit this relationship rather well for the Spillway sequence as well as several other sequences, including Mobile & Calendar and several other sequences.

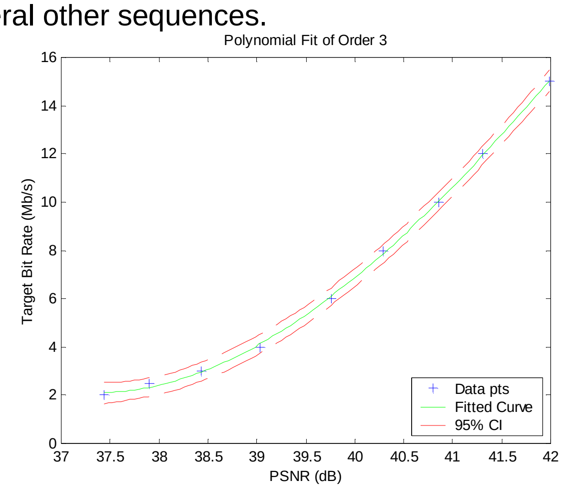

Figure 11 – Target Bit Rate vs. PSNR of MPEG-2 sequences of Spillway.

[3] Here functions are considered "linear" if linear in their regression parameters even if not in their variables. Hence $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ is linear in its parameters, even if not in x.

Figure 11 shows the fit of a 3rd order polynomial having intercept and coefficient values:

Regression Summary

Parameter Values:

B1 = -0.037669
B2 = 4.995117
B3 = -215.476711
B4 = 3044.552998

The regression procedure above was applied to sets of MPEG-2 sequences having source video titled "Duck", "Water", and "Mobile and Calendar." Still frames from each clip are shown in figures 12, 13, and 14.

Figure 13 - DUCK

Figure 12 - WATER

Figure 14 – MOBILE and CALENDAR

Mean PSNR was computed for each of 7 video sequences produced by MPEG2 encoding and decoding using the TM-5 software codec. Target bit rates of 15,12,10,8,6,4, and 2 MB/s were input to the codec. Regression analysis was performed to define for each a fitting function by which Target Bit Rate could be predicted from selected mean PSNR values. Good fits were obtained with 3rd and 4th order polynomials. The regression equations for each were then used to predict target bit rates that when input to TM-5 would produce sequences having PSNRs selected intermediate to values computed for the original set of MPEG-2 sequences. Mean PSNR values were computed for the new set of encoded sequences and the observed PSNR values compared via correlation with the predicted values.

For each source sequence, average PSNR values selected from original regression (in column 1) are compared with average PSNRs of new MPEG2-encoded sequences. Correlation matrices follow:

```
Duck =
   34.2060   33.9422
   35.8989   35.9058
   37.1768   37.2078
   38.0399   38.0204
   38.6540   38.6405
   39.7493   39.7298

Correlation Duck =
    1.0000    0.9992
    0.9992    1.0000

Water =
   33.0300   32.8835
   34.4600   34.4761
   35.6700   35.6823
   36.5100   36.4993
   37.1000   37.0809
   38.2200   38.2685

Correlation Water =
    1.0000    0.9997
    0.9997    1.0000

Mobile and Calendar =
   27.7300   27.6827
   29.5600   29.5377
   30.7400   30.7215
   31.5100   31.5408
   32.0900   32.1072
   33.2200   33.1980

Correlation M & C =
    1.0000    0.9999
    0.9999    1.0000
```

Correlation values indicate good agreement between PSNR values of sequences derived using the bit rates predicted via regression with sampled PSNR values. Thus, it is quite reasonable to use this fitting procedure to determine compression bit rate parameter required to generate an MPEG2 sequence exhibiting any desired average PSNR value. Given that a hardware encoder is available, it should be possible to compute graded impairment sequences using this procedure. This procedure can be used to generate a coarse sampling of "end-member" sequences over the range of JND testing to be fine-tuned via the mixing approach to converge on the jnd thresholds.